

Editors-in-Chief

Petar Bojanić, Snežana Vesnić

Deputy Editors

Miloš Čipranić, Zoran Erić, Marko Ristić

Executive Editor

Sara Dragišić

EDITORIAL BOARD

Alessandro Armando, *Politecnico di Torino*; Andrea Canclini, *Lancaster University*; Igor Cvejić, *University of Belgrade*; Edward Djordjevic, *University of Belgrade*; Maurizio Ferraris, *Università di Torino*; Dario Gentili, *Università Roma Tre*; Aya Jazaierly, *Loughborough University*; Mark Losoncz, *University of Belgrade*; Ognjen Marina, *Ss. Cyril and Methodius University in Skopje*; Andrey Menshikov, *University of Central Asia, Khorog*; Emanuele Morezzi, *Politecnico di Torino*; André Patrão, *Emory University*; Andrea Perunović, *University of Belgrade*; Andreja Prokopijević, *Université Jean Monnet, Saint-Étienne*; Željko Radinković, *University of Belgrade*; Manfredo di Robilant, *Politecnico di Torino*; Erdin Salihović, *University of Sarajevo*

INTERNATIONAL ADVISORY BOARD

Emmanuel Alloa, *University of Fribourg*; Andrew Benjamin, *Monash University, Melbourne*; Mario Carpo, *University College London*; Natalie Depraz, *Université de Rouen Normandie*; Giovanni Durbiano, *Politecnico di Torino*; Vladan Đokić, *University of Belgrade*; Peter Eisenman, *Yale University, New Haven*; Cynthia Fleury, *Conservatoire national des arts et métiers, Paris*; Jörg H. Gleiter, *Technische Universität Berlin*; Paul Guyer, *Brown University, Providence*; Catherine Ingraham, *Pratt Institute, New York City*; Neil Leach, *Florida International University, Miami*; Walter Mariotti, *Domus EIC/IULM University*; Miodrag Mitrašinić, *The New School, New York*; Branko Mitrović, *Norges teknisk-naturvitenskapelige universitet, Trondheim*; Gazela Pudar Draško, *University of Belgrade*; Alessandro Rocca, *Politecnico di Milano*; Ludger Schwarte, *Kunstakademie Düsseldorf*; Jules Simon, *The University of Texas at El Paso*; Luka Skansi, *Politecnico di Milano*; Charlotte Skene Catling, *Factum Foundation*; Chris L. Smith, *The University of Sydney*; Sarah Whiting, *Harvard University, Cambridge, Mass.*; Mark Wigley, *Columbia University, New York*

PUBLISHER: Institute for Philosophy and Social Theory, Kraljice Natalije 45, Belgrade, Serbia; Tel: +381112646242; Email: institut@ifdt.bg.ac.rs; Website: <https://ifdt.bg.ac.rs>.

ELECTRONIC ACCESS: All articles published in *Khōrein* are available online at: khorein.ifdt.bg.ac.rs.

COPYRIGHT: Articles published in *Khōrein* are open access in accordance with the Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not transformed in any way and is properly cited.

DISCLAIMER: The views expressed in the published works do not express the views of the editors and editorial staff. The authors take legal and moral responsibility for the ideas expressed in the articles.

For submission instructions and all other information, please visit <https://khorein.ifdt.bg.ac.rs>. For any questions, please contact the Editorial Office at khorein@ifdt.bg.ac.rs.

ISSN 2956-1892 (Print)

ISSN 3009-3805 (Online)

Page Layout: Sanja Tasić

Printed in Belgrade, Serbia by Alta Nova; 200 copies. The price of one print issue is 2400 RSD/20 EUR/23 USD/17 GBP. The annual subscription price (3 issues) is 5300 RSD/45 EUR/49 USD/39 GBP. A printed copy can be ordered at khorein@ifdt.bg.ac.rs.

χōrein

Khōrein: Journal for Architecture and Philosophy

VOLUME IV, Number 1

Automata

Guest Edited by Marco Vanucci

Published by the Institute for Philosophy and Social Theory
Belgrade, 2026

CONTENTS

- 1 Marco Vanucci
Automata

ESSAYS AND ARTICLES

- 5 Antoine Picon
Towards a New Renaissance? Artificial Intelligence and Architectural Design
- 21 Philippe Morel
Beyond Mimesis: From Imitation to Extrapolation in Architectural
Intelligence
- 65 Klaus Wieglerling
The Smart City Concept and Its Limits: Resistance, Technology, and
the Future of the City
- 99 Graham Harman
The Human Curator: AI and the Future of Writing
- 117 Andrew Witt
Architecture, Automata, and Artificial Memory
- 141 Matias del Campo
Architecture and the Ontology of Inference
- 175 Simon Weir
Architecture and Surrealist Automatism: From Historical Models to
AI Speculations
- 197 Alisa Andrasek
Architecture as Design Cognition: Synthetic Intelligence, Aesthetic Judgment,
and the Operative Turn
- 213 Marco Vanucci
Art in the Age of Generative Machines

INTERVIEWS

- 221 *Mimesis*, Precedent, and Artificial Intelligence: A Conversation with Mario
Carpo
- 235 Architecture as Interaction: A Conversation with Theodore Spyropoulos
- 247 Techno-Politics: A Conversation with Bernard Cache

BOOK REVIEWS

- 265 Roberto Bottazzi
Theodora Vardouli, *Graph Vision: Digital Architecture's Skeletons*, The MIT Press, Cambridge, Mass., 2024.
- 271 Alberto Fernández González
Daniel Koehler, *Compositional Intelligence: Architectural Typology Through Generative AI*, Routledge, Abingdon / New York, 2026.

Marco Vanucci

AUTOMATA

The progressive encapsulation of human knowledge into technical systems has accelerated automation, expanding “the number of operations which we can perform without thinking.”¹ Ranging from mechanical tools to AI-driven algorithms, this process has increased efficiency, scalability, and complexity in decision-making, while steadily altering the role of human agency. As technical systems evolve toward self-learning and autonomous determination, a central question arises: to what extent is human agency being displaced?

Autonomy implies self-governance: the capacity to establish one’s own principles and act according to them. Automation, by contrast, is the execution of predefined operations according to externally imposed conditions. While it optimises tasks, it remains fundamentally dependent on human intent. Or at least, it has until now.

With the rise of machine learning and predictive algorithms, technical systems are shifting toward self-regulation. Agency does not disappear in this process; it migrates: from production toward curation, from control toward evaluation, from authorship toward the construction of frameworks. The central question this raises is not whether automation is becoming more sophisticated, but whether it signals a categorical shift toward a new kind of autonomy altogether.

Bruno Zevi’s sceptical response to Luigi Moretti’s parametric experiments in 1960, “Electronic brains? No, calculating machines,”² already posed the question that now returns with renewed urgency. Contemporary

¹ A. N. Whitehead, *An Introduction to Mathematics*, Williams and Norgate, London, 1911, p. 61.

² B. Zevi, “Cervelli elettronici? No macchine calcolatrici,” *L’architettura: cronaca e storia*, 6, 62, 1960, pp. 508–509.

systems, however, are no longer calculating machines in Zevi's sense. Where Polanyi observed that "we can know more than we can tell,"³ the opacity of machine learning inverts this condition: we can tell, or rather our machines can output, far more than we understand. The shift is structural: from deterministic automata, in which outcomes remain traceable to explicit rules, to probabilistic generative systems, in which behaviour emerges from statistical inference across spaces of a complexity that exceeds human intuition.

Yet the apparent novelty of these developments obscures a longer history of artificial cognition embedded within technical media. As Andrew Witt argues in this issue, artificial memory constitutes a fundamental precondition for automata, functioning simultaneously as a technical medium and as a spatial device. In the seventeenth century, architects played a pivotal and largely overlooked role in this history: the French architect and hydraulic engineer Salomon de Caus developed reprogrammable musical automata whose memory architectures operated at building scale, making architectural thinking an essential condition for computational invention. These systems did not merely execute predefined sequences; they stored, organised, and reactivated information, anticipating contemporary concerns with encoding, retrieval, and machine readability. The architect thus emerges not only as a designer of form, but as a transdisciplinary inventor at the origin of machine memory itself.

Antoine Picon argues that artificial intelligence must be understood not merely as a new set of tools, but as a condition that transforms the conceptual foundations of architectural practice. Rather than asking what AI can do for design in instrumental terms, he examines how it compels a reconsideration of agency, authorship, intention, and responsibility. Central to his argument is the tension between the growing dominance of efficiency in contemporary design and the irreducible dimension of architectural intention that Boullée once called the "poetry of art," a quality that no process of technical optimisation can fully account for.

The decisive shift, Alisa Andrasek contends, is not from human to machine agency, but from architecture as representation to architecture as cognition. Intelligence becomes distributed across human, computational, and material systems, while creativity is reframed as the navigation of high-dimensional possibility spaces. What remains irreducible in this

³ M. Polanyi, *The Tacit Dimension*, Doubleday & Company, Garden City, New York, 1966.

expanded field is aesthetic cognition: the trained capacity to recognise latent order within configurations that have no prior typological grounding, to commit to a trajectory before its consequences are fully known. This framework positions the architect not as author or operator, but as the irreplaceable cognitive filter within a generative field that no machine can navigate alone. This understanding of distributed intelligence finds a collective and pedagogical dimension in the interview with Theodore Spyropoulos, whose work frames architecture as an ecology of interacting systems in which agency emerges through distributed, time-based processes rather than individual intention. The architect's role shifts from authoring form to staging conditions for interaction and adaptive formation. The question of whether contemporary AI systems merely recombine precedents or genuinely exceed them runs through several contributions to this issue. Surrealist automatism, Simon Weir argues, offers architecture not a historical technique but a cognitive framework: one that externalises perception, association, and thought beyond conscious control, and that extends naturally to AI-assisted design insofar as such systems reorganise design thinking in ways that exceed simple instrumentality. From a different angle, Matias del Campo proposes inferentialism as a methodology grounded in Simondon's critique of hylomorphism, where form does not precede its realisation but crystallises through probabilistic transductive processes distributed across human cognition, computational systems, and material environments. The most direct challenge to established positions comes from Philippe Morel, who draws on information theory, the history of computability, and recent results in high-dimensional geometry to argue that generative AI cannot adequately be described as a machine of imitation. Under a strict geometric definition, such systems are constrained to extrapolate beyond the convex hull of their training data, driven to invent as a mathematical necessity rather than an intentional act. As Mario Carpo notes, the mechanisms of imitation have become largely unintelligible within contemporary discourse,⁴ and his interview in this issue offers a counterpoint to Morel's argument, framing one of the central debates the issue sets out to explore.

The question of authorship finds its most direct and provocative formulation in Graham Harman's contribution, which challenges the

⁴ M. Carpo, "The Imitation Machine," in C. Ratti (ed.), *Catalogue of Biennale Architettura 2025. Intelligens. Natural. Artificial. Collective*, Edizioni La Biennale di Venezia, Venice, 2025, 1, pp. 190–193.

romantic image of the writer or designer generating work from nothing as a solitary creative act. This image, he argues, mistakes a transient historical condition for an essential one. What is genuinely essential to intellectual and creative life is not the capacity to produce *ex nihilo* but the capacity to discriminate, evaluate, and refine: connoisseurship rather than creation. The human role is not abolished but relocated, from the generation of drafts to the exercise of judgement over an abundance of generated possibilities, a shift that demands more cultivated taste, not less.

Klaus Wieglerling approaches the question of autonomous technology from a philosophical perspective, arguing that technical systems access reality only in a typological manner, registering statistical regularities while remaining unable to grasp what is singular, eventful, and irreducibly situated. The attempt to eliminate resistance through autonomous systems does not produce freedom but a corresponding loss of reality, and with it a loss of agency itself.

The political dimension of this loss is taken up directly in the interview with Bernard Cache, whose insistence on open-source infrastructures, shared syntactic frameworks, and public digital standards reframes automation as a question of sovereignty rather than efficiency. The increasing dependence on proprietary platforms, he argues, risks dispossessing architects of their own work. Who controls the digital environments of practice is not a technical question but a fundamentally political one.

The contributions gathered in this issue of *Khōrein* share a common wager: that the automaton need not be understood as a servant of efficiency or a threat to authorship, but as an unruly, generative condition that demands new forms of judgement, responsibility, and collective deliberation. What is at stake is not only how architecture is produced, but who gets to participate in the construction of its meaning.

Antoine Picon*

TOWARDS A NEW RENAISSANCE? ARTIFICIAL INTELLIGENCE AND ARCHITECTURAL DESIGN

ABSTRACT: This article examines artificial intelligence in architecture through a critical-theoretical lens, arguing that AI is not merely a tool but a force reshaping authorship, agency, and design practice. It explores three key issues: the emergence of human-machine assemblages that challenge traditional conception of authorship in architecture, the dominance of efficiency versus the “poetry” of architectural intention, and the limits of workflow optimization. Situating AI within historical analogies such as the Renaissance, the author questions these comparisons while highlighting AI’s political, cultural, and epistemological implications. Ultimately, the paper calls for alternative design paradigms that emphasize intention, inclusivity, and new relationships between humans, machines, and the environment.

KEYWORDS: artificial intelligence, design, architectural intention, architectural authorship

* Antoine Picon: Harvard Graduate School of Design, Cambridge, Mass.; apicon@gsd.harvard.edu.

This is an Open Access article under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not transformed in any way and is properly cited.

Artificial intelligence has rapidly moved from a specialized technological domain to a pervasive cultural and professional condition. Its recent acceleration, driven by advances in generative models, has not only expanded its field of application but also intensified the questions it raises. In architecture, as in other design disciplines, AI is no longer simply a matter of tools or techniques; it has become a lens through which fundamental issues—agency, authorship, intention, and responsibility—must be reconsidered.

This article approaches AI from the perspective of a critical theory attentive to its broader implications. Rather than asking what AI can do for architecture in purely instrumental terms, it examines how AI transforms the very frameworks within which architectural practice has traditionally been understood. In doing so, it situates current developments within a longer historical trajectory, while also questioning the limits of such analogies.

The argument developed here unfolds across three main lines of inquiry: the reconfiguration of authorship and the emergence of human-machine assemblages; the tension between efficiency and the “poetry of art” as a defining dimension of architectural intention; and, finally, the possibility of thinking beyond the dominant paradigm of workflow optimization toward alternative conceptions of design, time, and agency. Together, these perspectives aim to contribute to a more nuanced understanding of AI—not as a neutral instrument, but as a force that reshapes both practice and the way we conceive of ourselves as designers and, more broadly, as human beings.

A CONTRIBUTION TO A CRITICAL THEORY OF AI

Artificial intelligence is here to stay. The prospect of a new AI winter has largely receded, and AI will not withdraw after a brief moment in the spotlight.¹ As a result, questions concerning how societies, professions, and disciplines should adapt to this rapidly developing technology—and what can reasonably be expected from it—have become widespread. Since 2022, and particularly with the rapid diffusion of generative and diffusion models, AI has become a major source of interrogation for the design world. Where is it leading us? What degree of agency do we retain in the face of a technology that appears increasingly difficult to control?

¹ S. Chaillou, *Artificial Intelligence and Architecture: From Research to Practice*, Birkhäuser, Basel/Berlin/Boston, 2022.

Periods of accelerated change are frequently interpreted through comparison with earlier moments of dramatic technological and cultural transformation. Such analogies are often mobilized as tools for making sense of novelty. Yet this use of history is, in many respects, problematic—at least when measured against its modern, scientific conception. It recalls instead an older understanding of history as a guide for action rather than as an explanatory discipline. This tension lies at the core of Reinhart Koselleck's *Futures Past: On the Semantics of Historical Time*.² References to earlier periods of transformation do not so much aim to explain the present as to render it intelligible and to clarify the range of possible responses to ongoing change.

Since the emergence of the digital revolution, the Renaissance has been a recurring point of comparison. In the field of architectural theory, Mario Carpo famously drew a parallel between developments during the time of Alberti and those accompanying the generalization of computer-aided design and digital fabrication.³ At a broader cultural level, treating the digital not merely as a technical apparatus but as a cultural paradigm, the French sociologist Dominique Cardon has argued that the digital revolution bears a closer resemblance to the Renaissance and the spread of the printing press than to the industrial revolution.⁴

This Renaissance analogy has also resurfaced recently in public discourse on artificial intelligence. In a widely commented post, Sam Altman suggested that current developments may resemble the Renaissance more than the industrial revolution.⁵ Here again, the printing press functions as a key reference point, symbolizing an unprecedented acceleration of intellectual production and circulation. The implication is that AI may usher in a comparable expansion of knowledge, creativity, and cultural output.

Such comparisons are not entirely convincing. Nevertheless, they may serve as useful heuristic devices for unpacking some of the tensions currently at play. The Renaissance analogy is particularly charged in architecture, as this period witnessed the emergence of several defining features of the discipline as it is understood today. Among these was the

² R. Koselleck, *Futures Past: On the Semantics of Historical Time*, Columbia University Press, New York, 2004.

³ M. Carpo, *The Alphabet and the Algorithm*, The MIT Press, Cambridge, Mass., 2011.

⁴ D. Cardon, *Culture numérique*, Presses de Sciences Po, Paris, 2019.

⁵ "I think this is gonna be more the renaissance than the industrial revolution," S. Altman, "Post on X," April 18, 2025.

architect's claim to authorship: the ambition to be recognized as an intellectual figure, conversant with humanists and scholars, and supported by the written and printed word through the publication of treatises.

Authorship, however, cannot be considered in isolation. During the Renaissance, it was inseparable from questions of professional hierarchy and power. The transformation of architectural authorship coincided with a reconfiguration of roles on the building site, determining who benefited from the new tools and representational techniques and who lost influence in the process.

This shift entailed the gradual decline of craftsmen who had occupied dominant positions in medieval construction practices. The struggle was not merely professional but also axiological. Craft traditions—particularly in Northern Europe, where Flamboyant Gothic developed—valued virtuosity, performance, and technical display.⁶ These criteria were increasingly rejected by the emerging figure of the architect, who promoted ideals of measure, proportion, and harmony grounded in theoretical knowledge.

In contemporary discussions surrounding AI, a comparable emphasis on efficiency has become prominent. Efficiency is here understood broadly, encompassing form-finding processes, optimization strategies, and compliance with standards and regulations. This emphasis supports the image of a better-equipped designer, capable of responding more effectively to complex challenges such as climate change, resource scarcity, and evolving human needs.

Yet a fundamental question remains: can design be reduced to efficiency alone? Architecture has historically also been conceived as a source of meaning and inspiration that exceeds the satisfaction of immediate functional requirements. This dimension was famously articulated by the late eighteenth-century French architect Étienne-Louis Boullée as the “poetry of art”—a quality irreducible to technical optimization.⁷ The compatibility of this ambition with an extensive reliance on machines remains uncertain. It raises the question of whether such a poetic dimension can be preserved, transformed, or even amplified through computational means.

⁶ J.-M. Guillouët, *Flamboyant Architecture and Medieval Technicality: The Rise of Artistic Consciousness and the End of Middle Ages (c. 1400–c. 1530)*, Brepols, Turnhout, 2019.

⁷ J.-M. Pérouse de Montclos, *Étienne-Louis Boullée (1728–1799): Theoretician of Revolutionary Architecture*, Braziller, New York, 1974.

This is not an argument for reducing architecture to poetic expression alone. The material, regulatory, and economic constraints of professional practice leave limited room for such aspirations. Nevertheless, acknowledging the existence of this irreducible dimension—often associated with human intention—helps to clarify what is at stake in the development of AI within architectural practice. It brings into focus the question of intentionality: what it means to conceive, to decide, and to take responsibility for architectural form.

If efficiency is reconsidered beyond its narrow technical definition, the potential gains offered by AI may appear in a different light. Just as the Renaissance introduced new ways of seeing, representing, and conceiving the built environment, AI may open the door to alternative notions of efficiency and to more inclusive approaches to design.

These questions resonate with developments in the humanities, a field that has so far been more immediately and visibly affected by AI. Text production in particular, has undergone rapid transformation, prompting professions grounded in writing to question their future viability. Examining how AI is reshaping textual practices may therefore provide valuable insights for design disciplines. Moreover, contemporary AI systems increasingly blur the traditional boundary between text and image, even if the underlying computational models remain distinct.

From the perspective of the humanities, the objective is not to prescribe courses of action to AI developers or architects. Rather, it is to contribute to what the Swedish sociologist Simon Lindgren has termed a *critical theory of AI*.⁸ Such an endeavor is necessary precisely because AI is not merely a technical tool. It is a technology that interrogates our practices, our values, and ultimately our understanding of what it means to act, create, and decide as human beings.

TOWARDS A DIFFERENT POLITICS OF AUTHORSHIP

Current uses of AI in architecture appear to be developing along two relatively distinct trajectories. One focuses on ideation and form-finding, relying in particular on image-generative capacities to stimulate creativity and generate new perspectives. The other is oriented toward the optimization of constraints, ranging from programmatic and structural requirements to carbon emissions and liveability criteria. A recent study

⁸ S. Lindgren, *Critical Theory of AI*, Polity Press, Cambridge/Hoboken/New Jersey, 2023.

conducted in Turin and published in the latest issue of *Technology | Architecture + Design* suggests that the former approach is more frequently adopted by large architectural practices, which possess the resources necessary to develop dedicated tools, while the latter is more common among smaller firms.⁹

Whether this division will persist—thereby increasing the polarization of the profession between thriving large practices and smaller firms facing financial insecurity—or whether AI will eventually level the field through the widespread availability of powerful tools remains uncertain. At present, outcomes are still indeterminate, and the distribution of winners and losers has yet to be established.

Another source of uncertainty concerns the future modalities of co-operation between humans and machines. Many designers maintain that humans should retain mastery over the design process, distinguishing between the capacity to generate arrays of solutions—assigned to machines—and the ultimate power of decision, reserved for humans. Others, however, are struck by the growing potential of machines to guide the entire process, rather than merely supporting human intervention. While one position seeks to confine AI to the status of a tool, the other anticipates forms of co-design or even an expansion of machine-led decision-making.

At this stage, it may be useful to consider developments in text production, particularly within the humanities and, more specifically, in universities. In my own experience—far from unique—I am increasingly confronted with essays that are co-written by humans and machines, if not entirely generated by machines. These texts are often better written and more clearly structured than those produced independently by students, yet they frequently lack a distinctive mark of personal subjectivity, a point to which I will return. While some colleagues advocate a return to handwritten, supervised examinations, I am inclined to adopt a different perspective: we are no longer evaluating humans alone, but rather cyborg assemblages composed of humans and machines. This shift inevitably raises the question of which competencies evaluation procedures are meant to assess—a question to which no clear answer has yet emerged.

What developments in text production suggest is that it will become increasingly difficult to treat AI as a mere tool. In educational contexts,

⁹ F. J. Gardella, L. Mastrolia, F. Moro, M. Rossi, “The Impact of AI on Small Architecture Firms,” *Technology | Architecture + Design*, 9, 1, 2025, pp. 73–83.

as well as in many professional environments, AI already demonstrates the capacity to replace humans in certain situations.

Behind these machines, however, are humans who have produced the materials on which AI systems are trained, as well as others who have supervised their training. In other words, contemporary AI relies on a fractal hybridization of humans and machines operating at multiple levels. These assemblages are far more complex than traditional models of human-machine interaction, and they complicate—if not destabilize—the very notion of authorship.

During the 1960s and 1970s, French critical theory famously proclaimed the “death of the author.” Historians such as Roger Chartier, a specialist in the history of reading and publishing practices, echoed this claim by noting that the modern conception of authorship, as it had existed since at least the eighteenth century, had not always prevailed, and that its apparent demise might represent only a temporary parenthesis.¹⁰ That parenthesis, however, never fully closed. The death of the author did not occur. It may nonetheless be unfolding today, through the proliferation of partial or quasi-authors.

The situation remains more complex still, as contemporary culture is characterized by an age of hypertrophied stardom—from music to design—governed by a winner-takes-all logic. The fractal assemblages described above may continue to be attributed to a super-author, overseeing layers of human-machine interaction.

In architecture, this condition of the super- or arch-author predates the rise of AI. Architectural production has long been collective in nature, involving networks of people as well as tools and machines, well before the advent of digital computation. Nevertheless, the outcome of this collective process has traditionally been attributed to a single author. With AI, this veil appears to be lifting—although the process had begun earlier. Over the past decades, the collective, layered, and distributed character of architectural authorship has been examined by scholars using frameworks ranging from Peggy Deamer’s use of the post-Marxist concept of “immaterial labour” to Bruno Latour’s actor-network theory.¹¹

¹⁰ R. Chartier, *The Order of Books: Readers, Authors, and Libraries in Europe Between the 14th and 18th Centuries*, Stanford University Press, Stanford, 1994.

¹¹ P. Deamer (ed.), *The Architect as Worker: Immaterial Labor, the Creative Class, and the Politics of Design*, Bloomsbury, London/New York, 2015; A. Yaneva, *Latour for Architects*, Routledge, London/New York, 2022.

AI now forces these issues beyond the relative confidentiality of academic debate, compelling a more direct engagement.

AI also calls for renewed reflection on compensation at multiple levels. Should new mechanisms be devised to remunerate the producers of works used in machine training? What about those who supervise such training processes? And how should compensation account for the varying degrees of human involvement within human–machine assemblages? These questions suggest the need for differentiated modes of recognition and remuneration.

Without indulging in utopian visions of perfect fairness, it must be acknowledged that AI will compel design professions to rethink key aspects of their organization. AI, like many features of digital culture, is as much a political issue as it is a technological one.

In this context, it is worth asking who the losers might be in the profound transformation of the designer into a constellation of cyborg assemblages—networks of human and non-human actors rather than autonomous individuals. Thus far, the primary losers of the digital turn in architecture were those unable to master complex software or who failed to ascend early enough within professional hierarchies to delegate computational tasks to others. AI may demand a different set of skills than those previously dominant. Consider, for instance, the emerging importance of prompt-writing. Interacting with “intelligent” machines may resemble the practice of animal training, where success depends on the ability to communicate effectively across species. In this sense, interspecies communication may prove more valuable than traditional computer skills. One might even imagine future courses dedicated to understanding latent spaces—a competence that could become as significant as coding itself.

This development poses a major challenge for architectural education, which traditionally seeks to produce winners rather than losers. The rise of AI is likely to accelerate a shift toward asking the right questions, formulating meaningful problems, and curating the multiple responses generated by machines, rather than producing designs in the traditional sense. In other words, emphasis may move from the *how* to the *why*, from step-by-step production to strategic inquiry. This raises a fundamental pedagogical question: how does one train curators? At present, architectural education remains largely oriented toward mastering procedures and techniques.

Finally, these issues extend well beyond the design studio. For historians, AI represents both an opportunity and a challenge. While it offers

powerful new tools, it also renders simple, linear narratives increasingly implausible. The history of architecture, even as it has moved beyond canonical accounts toward more global perspectives, has largely continued to rely on unified, coherent narratives. AI challenges the very premise of such singular narratives. Its capacity for endless hybridization opens new perspectives at a pace that can be disorienting. Wikipedia had already exposed the limits of traditional survey models; AI now more radically questions the possibility of a single narrative framework in historical teaching. This does not imply abandoning narrative altogether, but rather redefining what narrative can mean under these new conditions.

AI PRODUCTION AND THE “POETRY OF ART”

The question of the ambition that guides design leads directly to that of architectural intention. As previously noted, the growing centrality of efficiency in contemporary design practices is striking, even when it is not explicitly framed as such. Efficiency has long been a defining feature of design activity, but it now appears to govern practice more systematically than in the past. It permeates the generation of ideas as well as the elaboration of forms and projects conceived to satisfy increasingly precise requirements. This emphasis on efficiency may function as a form of reassurance: from this perspective, AI would represent an augmentation rather than a radical transformation of design practice, enabling designers to keep pace with the acceleration and complexity of the contemporary world.

Yet this perspective raises questions regarding the irreplaceable role of humans in the design process. If design is primarily understood as “enabling an environment to function appropriately,” to borrow a formulation from AI-in-architecture specialist Immanuel Koh, it becomes conceivable that machines might eventually perform this task without human input.¹²

This possibility is readily apparent in domains such as structural soundness or thermal comfort. But does it extend to aesthetics? AI systems already demonstrate the capacity to process extremely large datasets, ranging from the totality of Baroque churches to the corpus of modernist villas built in the first half of the twentieth century. This capacity enables them to identify stylistic patterns and even to infer preferences.

¹² I. Koh, *Artificial & Architectural Intelligence in Design*, Singapore University of Technology and Design, Singapore, 2020.

At this point, it is useful to examine more closely certain characteristics of contemporary AI productions. They often appear to oscillate between two poles: the slightly boring and the spectacular.

The first pole can be illustrated by an experience at the 2025 Venice Biennale curated by Carlo Ratti. There, the numerous panels lining the main exhibition space were accompanied by AI-generated summaries intended to facilitate the reading of long and detailed texts. These summaries were accurate but flat and bland; something of the excitement conveyed by the original texts seemed to be lost. A similar flatness characterizes many AI-generated or AI-assisted papers submitted by students. Although often better written and more clearly structured than purely human-produced texts, they tend to lack a distinctive inflection.

At the opposite pole lies the spectacular. This includes images depicting monumental, intricate, and often labyrinthine scenes. Even more striking are the videos proliferating on social media, featuring hybrids of humans and animals or humans and machines engaged in strange rituals. These productions are undeniably spectacular, yet the sense of awe they provoke is frequently accompanied by an impression of *déjà vu*. This reinforces the hallucinatory quality often associated with AI-generated imagery.

Taken together, many AI productions appear to lack a distinctive personality. Chatbots, by contrast, seem capable of adopting almost any persona in conversation. Yet personality may not be the most appropriate concept here; intention provides a more precise term. AI outputs often fascinate without fully engaging us in the manner of human works that convey a clear intention. When chatbots give the impression of personality, this may owe less to the intrinsic content of their responses than to the user's desire to be heard—a desire that can obscure critical distance.

One explanation for the relatively impersonal quality of many AI productions lies in their reliance on processes analogous to imitation, a principle that dominated artistic production for centuries. Mario Carpo was among the first to draw a parallel between AI and the principle of mimesis—reasoned imitation—that structured early modern European art as well as many other artistic traditions.¹³ Creation largely meant imitation. While artists were theoretically expected to imitate nature, in

¹³ M. Carpo, "The Imitation Machine," in Carlo Ratti (ed.), *Intelligens. Natural. Artificial. Collective. Biennale Architettura 2025*, vol. 1, La Biennale di Venezia, Venice, 2025, pp. 190–193.

practice they often imitated one another's works. AI has amplified this logic of imitation to such an extent that it is not always immediately recognizable, though its presence is often sensed at a subliminal level. The absence of a clearly articulated intention is further revealed by AI's capacity to hybridize almost everything with everything else. The resulting productions—resembling a Max Ernst composition on steroids—bear a distinctive mechanical signature.

Historically, mimesis most often gave rise to conventional, not necessarily compelling works. Most human creations are, in fact, not radically creative. It is therefore unsurprising that AI reproduces this pattern.

This dynamic is evident in certain architectural uses of AI that explicitly capitalize on imitation. Coop Himmelblau's *Deep Himmelblau*, for instance, is designed to generate proposals derived from the firm's previous projects.¹⁴ This amounts to a form of self-imitation—a fundamental mechanism of cultural production, particularly in contexts governed by economic imperatives.

To move beyond imitation, humans have at times needed to operate at the margins of established rules. Within imitation-based systems, artistic innovation historically emerged at the threshold where imitation began to transform into something new. Because of their fundamentally foreign modes of processing text and images, machines can provoke similar explorations. However, their rule-based nature continues to constrain the scope of what they can propose.

Not all forms of novelty, however, are equally meaningful to humans. In the realm of language, certain combinations of words resonate by striking an inner chord. This capacity for resonance lies at the core of poetry.

Such evocative power is closely linked to embodiment. The resonance of particular words, forms, textures, or colors derives from holistic experience as embodied intelligence, rather than from purely intellectual comprehension. Meaning, understood in this sense, is not abstract or fixed, like a dictionary definition; it is rooted in physical and sensory experience.

Applied to architecture, this understanding suggests that architecture expressing a human intention beyond the satisfaction of constraints—architecture as a poetic endeavor—operates in a comparable way. Its meaning is embodied, even if contemporary theory no longer adheres to direct analogies between architectural proportion and the human body.

¹⁴ W. dPRIX, K. Schmidbaur, D. Bolojan, E. Baseta, "The Legacy Sketch Machine: From Artificial to Architectural Intelligence," *Architectural Design*, 92, 3, 2022, pp. 14–21.

Architectural intention thus exceeds the fulfillment of practical requirements, whether technological or aesthetic. This is what Étienne-Louis Boullée meant by the “poetry of art,” and what Le Corbusier had in mind when, after famously declaring that a house was a machine for living in, he corrected himself to insist that architecture was ultimately a machine for producing emotion. Even his modest *Cabanon* on the Mediterranean coast was conceived as an emotional experience. Walter Gropius articulated a similar position when, in *The New Architecture and the Bauhaus* (1936), he argued that while architecture must mobilize the most advanced science and techniques, emotional and aesthetic satisfaction remained paramount.¹⁵

Architects do not, of course, generally aspire to be Le Corbusier, Gropius, or Zumthor. Nevertheless, architectural practice is inseparable from the desire to express something beyond the merely functional, and from the expectation that this expression will remain perceptible amid technical and regulatory constraints.

This ambition is also tied to the aspiration to go beyond the merely livable and to create spaces that are truly habitable—that is, spaces that suggest that human actions within them have meaning. The ultimate purpose of architecture is not to signify in itself, but to frame meaningful human activity. In this sense, architecture resembles a theater, inviting humans to become the actors of their own lives.

The capacity to generate emotion—emotion that resonates not only cognitively but through embodied intelligence—is what distinguishes genuinely poetic human production from mere imitation or unreflective rule-following.

Most of the time, humans follow codes and procedures, producing outcomes that machines will likely soon be able to generate autonomously. Human cognition itself has a mechanistic dimension, a fact that has long fascinated philosophers. Denis Diderot’s interest in machines, for example, stemmed from his desire to explore what is machine-like in the human mind. His editorial project for the *Encyclopédie* sought to investigate this question through the systematic documentation of arts and crafts. Today, this question has regained particular urgency.

What remains uncertain is whether machines will be able to go further and produce works that convey a genuinely human-like intention. Beyond

¹⁵ W. Gropius, *The New Architecture and the Bauhaus*, Museum of Modern Art/Faber and Faber, New York/London, 1936.

the poetry born of their radical strangeness, will they be able to move us in the same way as human-created works? Will they, in other words, eventually imitate our capacity to be—in certain moments—inimitable?

This scenario cannot be excluded. Following the thread of embodiment, such a development would likely depend not only on increased computational power but also on its embodiment. At present, this embodiment remains rudimentary, but it is evolving rapidly.

These questions are far from purely academic. The future of the architectural profession is directly implicated. If machine-generated projects were able to move us in the same way as human productions, the designer's role might be reduced to selecting among alternatives—the curatorial function previously evoked. Yet this raises a further question: why would curators still be needed, and on the basis of which specific competencies?

Ultimately, the future of design also concerns our understanding of ourselves. Is life fundamentally computational, as suggested by Google researcher Blaise Agüera y Arcas in *What Is Intelligence?*¹⁶ Are humans even more machine-like than they have believed at various points in history? Returning once more to the Renaissance analogy, transformations in design ideals and practices were inseparable from broader conceptions of what it meant to be human. Today, architecture once again serves as a privileged site for this enduring inquiry. Who, in the end, are we?

BEYOND WORKFLOWS

With the widespread adoption of AI, the notion of the workflow has become even more pervasive than before. It is inseparable from a fluid conception of efficiency in which work sequences and documents are expected to circulate seamlessly, production processes are streamlined, and nothing is allowed to hinder rapid delivery.

Research in the history of engineering shows that this vision of efficiency is rooted in a long intellectual trajectory that began in the second half of the eighteenth century. At that time, engineers and early economists alike developed a growing fascination with a world in which movement became the primary measure of value. Movement gradually replaced harmony as the dominant organizing principle. For the

¹⁶ B. Agüera y Arcas, *What Is Intelligence? Lessons from AI About Evolution, Computing, and Minds*, Cambridge, Mass./London: The MIT Press, 2025.

Physiocrats, for instance, the circulation of goods was central to the generation of economic value.

This emphasis on movement was inseparable from a transformation in the understanding of nature, which came to be interpreted less in terms of harmonious order than as a dynamic system in constant flux. The decline of the Vitruvian tradition in architecture was closely connected to this shift. Order and proportion progressively lost both their artistic and scientific prestige.¹⁷

A century later, this logic found a new expression in scientific management and its obsession with streamlining production. Factory automation followed a similar path. Contemporary discourse around AI appears to extend this trajectory toward the ultimate streamlining of conception itself, aiming at the full integration and unification of workflows.

Yet this evolution is not necessarily inevitable. In the context of the Anthropocene, contemporary societies are simultaneously experiencing a renewed transformation in their understanding of nature. Increasing attention is now paid to complexity, as well as to the intertwined conditions of fragility and resilience that were previously underestimated.

In parallel, slowness has emerged as a key value among environmental thinkers seeking to restore a more balanced relationship between humans and non-humans. This development invites a reconsideration of the relationship between architecture and time. Streamlining workflows may not constitute the sole priority. Architecture might instead be conceived from the perspective of deceleration rather than as a contributor to generalized acceleration. AI could just as readily be mobilized toward this end, for instance by helping to carve out moments of pause or contemplation, rather than exclusively serving to accelerate processes.

This reconsideration of efficiency also extends to the performance of buildings. In a thought-provoking essay entitled *Antidote to the Cult of Performance: The Robustness of the Living*, a French biologist contrasts traditional optimization strategies—which often increase vulnerability to unexpected events—with the robustness of living organisms, which rely on high levels of redundancy.¹⁸ Taking nature as a model, albeit a nature understood differently from that of the Enlightenment and modernity,

¹⁷ A. Picon, “Towards a History of Technological Thought,” in R. Fox (ed.), *Technological Change: Methods and Themes in the History of Technology*, Harwood Academic Publishers, London, 1996, pp. 37–49.

¹⁸ O. Hamant, *Antidote au Culte de la Performance: La Robustesse du Vivant*, Gallimard, Paris, 2023.

could foster a renewed conception of optimization and performance. Here again, AI may offer opportunities to think beyond narrowly defined metrics of efficiency and to explore alternative design paradigms.

AI also opens new possibilities for diversifying the actors involved in the design process and for making previously unheard voices audible—an especially striking reversal given the biases that AI systems often reproduce. This approach has been explored by Thomas W. Sanchez, who argues that “when trained on diverse datasets incorporating a wide range of sources and community input, AI tools can act as powerful instruments for inclusivity.”¹⁹ The uses of new technologies are never fully predetermined, and AI is not inherently synonymous with technocratic decision-making. Notably, such possibilities were already anticipated in the early history of computing in architecture. Nicholas Negroponte’s Architecture Machine Group at MIT experimented with ways of mobilizing computational power to incorporate the aspirations of social groups that were frequently marginalized.²⁰

This line of inquiry can be extended further by considering the potential of AI to give voice to non-human actors in ways that differ from Bruno Latour’s notion of “parliaments of things.”²¹ In 2016, researchers at the Berlin University of the Arts proposed an alternative approach, suggesting the use of sensors and algorithms to grant natural elements a form of agency, including limited decision-making power. They imagined, for example, a digitally augmented forest capable of managing itself and even expanding by acquiring land through blockchain technologies.²² Building on such ideas, AI could contribute to a form of design practice that genuinely listens to non-human entities, moving beyond rhetorical commitments that often lack tangible consequences.

Artificial intelligence holds the promise that extends far beyond the optimization of workflows. Returning to Sam Altman’s evocation of an impending new Renaissance, it is indeed possible that a period of renewal lies ahead. Yet this Renaissance may differ markedly from the one anticipated by many voices in Silicon Valley. It is this alternative vision—one that architecture has a crucial role to articulate—that remains at stake.

¹⁹ T. W. Sanchez, “Planning with Generative AI for Creativity, Global Insight, and Fairness,” *Technology | Architecture + Design*, 9, 1, 2025, pp. 19–24.

²⁰ N. Negroponte, *The Architecture Machine*, The MIT Press, Cambridge, Mass., 1970.

²¹ B. Latour, *Politics of Nature: How to Bring the Sciences into Democracy*, Harvard University Press, Cambridge, Mass., 2004.

²² “Autonomous Forest,” <https://www.las-art.foundation/programme/terra0-autonomous-forest> (accessed April 29, 2026).

REFERENCES

- Agüera y Arcas, Blaise (2025), *What Is Intelligence? Lessons from AI About Evolution, Computing, and Minds*, Cambridge, Mass./London: The MIT Press.
- Cardon, Dominique (2019), *Culture numérique*, Paris: Presses de Sciences Po.
- Carmo, Mario (2025), "The Imitation Machine," in Carlo Ratti (ed.), *Intelligens. Natural. Artificial. Collective. Biennale Architettura 2025*, vol. 1, Venice: La Biennale di Venezia, pp. 190–193.
- Carmo, Mario (2011), *The Alphabet and the Algorithm*, Cambridge, Mass.: The MIT Press.
- Chaillou, Sébastien (2022), *Artificial Intelligence and Architecture: From Research to Practice*, Basel/Berlin/Boston: Birkhäuser.
- Chartier, Roger (1994), *The Order of Books: Readers, Authors, and Libraries in Europe Between the 14th and 18th Centuries*, trans. Lydia G. Cochrane, Stanford: Stanford University Press.
- Deamer, Peggy (ed.) (2015), *The Architect as Worker: Immaterial Labor, the Creative Class, and the Politics of Design*, London/New York: Bloomsbury.
- Gardella, Francis Joe, Mastrolia, Luciana, Moro, Francesca, Rossi, Marta (2025), "The Impact of AI on Small Architecture Firms," *Technology | Architecture + Design*, 9, 1, pp. 73–83.
- Gropius, Walter (1936), *The New Architecture and the Bauhaus*, New York/London: Museum of Modern Art/Faber and Faber.
- Guillouët, Jean-Marie (2019), *Flamboyant Architecture and Medieval Technicality: The Rise of Artistic Consciousness and the End of Middle Ages (c. 1400–c. 1530)*, Turnhout: Brepols.
- Hamant, Olivier (2023), *Antidote au Culte de la Performance: La Robustesse du Vivant*, Paris: Gallimard.
- Koh, Immanuel (2020), *Artificial & Architectural Intelligence in Design*, Singapore: Singapore University of Technology and Design.
- Koselleck, Reinhart (2004), *Futures Past: On the Semantics of Historical Time*, New York: Columbia University Press.
- Latour, Bruno (2004), *Politics of Nature: How to Bring the Sciences into Democracy*, Cambridge, Mass.: Harvard University Press.
- Lindgren, Simon (2023), *Critical Theory of AI*, Cambridge/Hoboken/ New Jersey: Polity Press.
- Negroponte, Nicholas (1970), *The Architecture Machine*, Cambridge, Mass.: The MIT Press.
- Pérouse de Montclos, Jean-Marie (1974), *Étienne-Louis Boullée (1728–1799): Theoretician of Revolutionary Architecture*, New York: Braziller.
- Picon, Antoine (1996), "Towards a History of Technological Thought," in Robert Fox (ed.), *Technological Change: Methods and Themes in the History of Technology*, London: Harwood Academic Publishers, pp. 37–49.
- Prix, Wolf D., Karolin Schmidbaur, Daniel Bolojan, Efilena Baseta (2022), "The Legacy Sketch Machine: From Artificial to Architectural Intelligence," *Architectural Design*, 92, 3, pp. 14–21.
- Sanchez, Thomas W. (2025), "Planning with Generative AI for Creativity, Global Insight, and Fairness," *Technology | Architecture + Design*, 9, 1, pp. 19–24.
- Yaneva, Albena (2022), *Latour for Architects*, London/New York: Routledge.

Philippe Morel*

BEYOND MIMESIS: FROM IMITATION TO EXTRAPOLATION IN ARCHITECTURAL INTELLIGENCE

ABSTRACT: The progressive encapsulation of architectural knowledge within technical systems has led to increasing automation of design and decision-making processes. Long limited to the execution of explicit rules, this automation is now undergoing a profound transformation with the rise of machine learning systems, characterised by self-regulation, probabilistic decision-making, and increasing opacity of their internal mechanisms. This article analyses this transformation in three stages. First, it revisits Shannon's rigorous quantification of information, a prerequisite for its machine processing. Next, it examines the genesis of computability and the conjecture formulated at the Dartmouth conference, according to which intelligence and creativity can be simulated by formal procedures incorporating controlled randomness. Finally, by comparing Mario Carpo's imitation thesis with recent results in high-dimensional geometry, particularly those of Yann LeCun, the article argues that contemporary AI cannot be described adequately as a simple machine of imitation. Under a strict geometric definition of interpolation, contemporary AI systems are better characterised as operating through extrapolative generalisation than through interpolation in the ordinary sense. This leads to a redefinition of the autonomy of the architectural automaton, understood less as the capacity to reproduce precedents than as the tendency, in high-dimensional spaces, to generate outputs beyond the strict envelope of available data.

* Philippe Morel: The Bartlett School of Architecture, Faculty of the Built Environment, University College London; p.morel@ucl.ac.uk.

This is an Open Access article under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not transformed in any way and is properly cited.

KEYWORDS: artificial intelligence, extrapolation, computational architecture, high-dimensional geometry, imitation, information theory, theory of architecture

PART I. ARCHITECTURAL INTELLIGENCE AND THE ADVENT OF MEASUREMENT

Fragmentation of Intelligence and the Persistence of the Qualitative

Each historical period is marked by the emergence, crystallisation, and formalisation of new forms of architectural intelligence. Whether or not one considers form (*morphē*) to be the primary expression of intelligence in a discipline still referred to, for convenience, as “architecture”—even though it now appears increasingly diversified, fragmented, or even dissolved into adjacent fields ranging from computational urbanism to interaction design—the twentieth century can be said to have witnessed the emergence of an unprecedented number of new architectural intelligences. These intelligences, which will be referred to interchangeably as “architectural intelligences” or “intelligences of architecture,”¹ cannot be reduced to a single or monolithic definition. They must be understood in at least two distinct but complementary senses, both of which have structured classical and modern architectural theory. The first sense, which may be described as cognitive and subjective, refers to the human capacity to understand architecture: to grasp its tacit rules, recognise its specificities, and navigate its multiple dimensions—whether material, economic, political, or historical. This form of intelligence is that of the designer who synthesises heterogeneous constraints, of the user who inhabits the space, or of the visitor who experiences it. It falls within the domain of a hermeneutics of space. The second sense, objective and ontological, refers to the intelligence that is inscribed, incorporated, or embedded in a construction, or even in simple graphic documents. A Gothic vault, a Palladian plan, or a detail of modern construction “contains” intelligence—that is, organised information that resolves problems of force, light, or circulation, embodies theoretical positions, and transmits

¹ The expression “the intelligences of architecture” could mean the different understandings of architecture, but this is not the meaning adopted here, although the present text implicitly raises the question of the ability of machines to understand architecture and therefore to process semantic information.

symbolic content. This second meaning of architectural intelligence—intelligence embedded in artifacts—is not a weak metaphor. It designates an operational reality: an organisation of constraints, decisions, and regularities that can be manipulated, transmitted, and, today, reproduced without recourse to conscious human intention. It is precisely this progressive autonomisation of embedded intelligence that makes its mechanical processing possible.

Until very recently—until the advent of generative artificial intelligences behaving as if they understood the meaning of their creations—these two definitions rested on an implicit and insurmountable anthropocentrism. The first definition implied the existence of a human rationality capable of establishing causal relations, hierarchies of value, and symbolic correspondences. The second definition, for its part, always referred to the existence of an additional *je ne sais quoi*, an ineffable quality eluding strict scientific explanation. This romantic idea persisted, inherited in part from the theory of Fine Arts and reinforced by phenomenology, according to which a work of architecture could contain “more”—more know-how, more technical skill, more knowledge, or more emotion—than the simple sum of the information necessary for its conception and execution. The architectural work was perceived as an organic whole, greater than the sum of its technical parts. However, from the strictly quantitative perspective provided by information theory—a theory which, as we shall see, forms the basis of our computational modernity—this hypothesis of an immaterial surplus is untenable, even physically impossible. From the standpoint of digital representation, any object, building, or drawing can be encoded to finite precision and processed as finite data. Faced with this contradiction between informational finitude and the feeling of the inexhaustibility of the sensible, architecture has historically adopted a defensive strategy: postulating the fundamental irreducibility of architectural experience. This irreducibility, however, is never demonstrated; it is asserted as a disciplinary axiom, often due to a lack of formal means to propose an operational reduction. In fact, to circumvent or at least displace the problem, architecture has taken refuge in disciplines external to strict technique, such as aesthetics, phenomenology, or semiotics. The underlying assumption of this resistance is that architecture cannot be reduced to a few megabytes of information stored in a computer’s memory. It is precisely this assumption of the irreducibility of the sensible that the computational age and artificial intelligence radically challenge.

A physical experience of architecture engages one's own personality, culture, and state of mind—in reality, an extremely broad, diverse, and complex general context. At its most extreme, although this experience is described as physical, it also inevitably involves—at least for some people—a worldview (*Weltanschauung*) that artificial intelligence “world models” are attempting to approximate, though these world models focus mostly on their physical dimensions. Trying to circumscribe this context to represent it faithfully as data would lead to immense theoretical difficulties. How can we translate a “state of mind,” a fleeting emotion, or the cultural background that guides an aesthetic judgment into binary variables? Until the advent of AI, these questions were left to the humanities. Architecture, in its practice, was content to quantify the “quantifiable.” During the design phase, the focus was on measuring spaces, evaluating costs (the famous *devis*), and dimensioning structures—areas for which rigorous formalisation had been made necessary, and indeed natural, by the industrial and scientific revolutions. As for the construction phase, the problems of representation and quantification associated with it were *ultimately* just an extension of the former, at least from the Renaissance onward, when graphic design documents truly dictated the building process.

Geometry and Arithmetic: A Historical Division of Labour

This limitation of quantification to physical and economic aspects has historically structured the architectural profession and its relationship with engineering, including before the Renaissance, during the empirical dialogue between drawing (*l'art du trait*) and wood or stone cutting, where quantification and representation went hand in hand within a kind of “geometric and arithmetic certainty.” Ordinary or Euclidean geometry² dealt with surface and volume elements (to simplify, what occupies space and defines form), while arithmetic dealt with elements without direct spatial extension (for example, costs, or more recently, energy losses or carbon footprint). This symbolic division of labour between

² Some might be tempted to invoke descriptive geometry instead of Euclidean geometry, but, as the mathematician Jean Dieudonné pointed out, descriptive geometry is merely a “simple drawing technique:” “I am not speaking of ‘Descriptive Geometry,’ a simple drawing technique which has fortunately been brought back to its proper proportions almost everywhere.” See J. Dieudonné, *Algèbre linéaire et géométrie élémentaire*, Hermann, Paris, 1964, p. 11 (second footnote on p. 11). This footnote, slightly polemical and related to the French context, is absent from the standard English translation *Linear Algebra and Geometry*, Houghton Mifflin, Company Boston, 1969.

geometry and arithmetic has, in the history of architecture, translated remarkably stably into a very concrete division of labour between architects and engineers, definitively established by Jean-Nicolas-Louis Durand³ and Jean-Baptiste Rondelet at the beginning of the nineteenth century. This system, in which architects are responsible for the science of space and engineers and other specialised bodies are responsible for virtually everything else, has endured through the centuries with remarkable stability. While the details became more complex in the twentieth century with the use of new sciences (for example, thermodynamics, which governs thermal engineering), the general structure of this division of labour remained unchanged until the advent of contemporary artificial intelligence systems.

We are witnessing a fundamental epistemological shift today. Current AI systems, by encoding all architectural knowledge (texts, plans, images, historical contexts, critical analyses) within vast neural networks, are achieving an unprecedented fusion of knowledge. They are giving rise to what we propose to call, extending the analogy with the latent space of mathematics, “latent engineering.” This term designates a diffuse, shared, and constantly accessible engineering capability that is no longer limited to load or thermal calculations. The focus of this latent engineering is no longer classical quantification (volumes, material strength, costs) measured in standard units (meter, pascal, euro). The challenge now lies in quantifying and processing what we have never been able to quantify before: complex contextual elements, embodiment (bodily and situated inscription), or even architectural quality itself, invariably reduced in the past to vague notions such as elegance, poetry, beauty, the sensations generated, or famously creativity.⁴ Where aesthetics, psychology,⁵ semiotics,⁶ and cognitive science have failed to provide a universal measure of architectural quality, AI offers a radically different approach: not a direct measure of beauty or creativity, but learned vector representations that

³ Cf. W. Szambien, *J.-N.-L. Durand 1760–1834: de l’imitation à la norme*, Picard, Paris, 1984; J.-B. Rondelet, *Traité théorique et pratique de l’art de bâtir*, published by the author, Paris, 1812–1817; P. V. Aureli, *Architecture and Abstraction*, MIT Press, Cambridge, Mass., 2023.

⁴ See M. Boden, *The Creative Mind: Myths and Mechanisms*, Basic Books, New York, 1991.

⁵ See, for example, R. Arnheim, *Art and Visual Perception: A Psychology of the Creative Eye*, University of California Press, Berkeley/Los Angeles, 1954.

⁶ On this subject, see for example U. Eco, *A Theory of Semiotics*, Indiana University Press, Bloomington/London, 1976.

correlate with patterns of judgment, resemblance, and preference found in data. Automation, as the “progressive encapsulation of human knowledge in technical systems,”⁷ crosses a critical threshold here. It no longer concerns only the execution of repetitive tasks, but the encapsulation of qualitative judgment itself.

Aesthetic Objectivity and Its Limits

It would be wrong to assume that the ambition to rationalise art and measure beauty was born with the computer. Projects to explain and formalise aesthetic experience are concomitant with scientific modernity.⁸ Art history has extensively documented this vast—profound but uncoordinated—project of objectively quantifying artistic qualities that unfolded between the mid-nineteenth century and the 1970s. From theories of pure vision to structuralist analyses, the West has sought to reduce art to equations. The culmination of this quest was reached when the European and American artistic avant-gardes, often in direct or indirect dialogue with logical neo-positivism (the Vienna Circle), attempted to produce an “exact” art. Constructivism and concrete art participated in this desire for rigor, with the “modernist aporias” that conceptual and minimal art would later highlight. However, postmodern relativism brought this dynamic in the cultural sphere to a halt. This scepticism was undoubtedly beneficial, preventing art from becoming a scientific caricature of science. Nevertheless, as the philosopher of mathematics James Franklin ironically observed, “The postmodernists’ attacks have washed off science like water off a duck’s back.”⁹ Science has not abandoned explaining the qualitative simply because art critics have declared the end of grand narratives; it has merely changed its tools and methods.

⁷ *Khôrein: Journal for Architecture and Philosophy*, “Call for Papers: Volume IV, Issue 3: Automata,” <https://khorein.ifdt.bg.ac.rs/index.php/ch/announcement/view/22>, (accessed 8 April 2026).

⁸ Beyond classical aesthetics, perhaps the most serious attempt to quantify aesthetics was G. D. Birkhoff, *Aesthetic Measure*, Harvard University Press, Cambridge, Mass., 1933. Although, in my view, it remained inconclusive, the theory of aesthetic measurement found a “second life” with applications of C. Shannon’s information theory to aesthetics; see A. Moles, *Information Theory and Aesthetic Perception*, Denoël, Paris, 1972. More recently, it has reappeared in algorithmic methods (formal grammars, cellular automata, agents, etc.) and in algorithmic information theory. Most current approaches, however, still seem to fall into some form of scientism.

⁹ J. Franklin, “Science Wars: Last Round? (A Review of *Are Science and Mathematics Socially Constructed? A Mathematician Encounters Postmodern Interpretations of Science* by Richard C. Brown),” *The New Criterion*, 29, 9, 2011, p. 68.

Faced with the inconclusive or overly limited results of traditional scientific aesthetics or structural semiotics, Artificial Intelligence has taken up the problem of quantifying the qualitative. From its very origins as a discipline, AI has posited a bold and reductionist hypothesis: qualitative problems (natural language, visual perception, creativity, learning) can be addressed using nothing more than what is strictly measurable, that is, information and computation. This represents a major shift: the search is no longer for the essence of beauty (a philosophical approach), but rather for simulating the effects of intelligence and aesthetic judgment through computational means.

The Dartmouth Programme: A Foundational Conjecture

The institutional birth of this ambition is found in a document of paramount importance to the history of ideas: the *Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*,¹⁰ written in August 1955. The project is not a product of naive positivism but rather a methodological gesture: unable to define intelligence in a substantive way, the Dartmouth researchers chose to make it a working hypothesis falsifiable through simulation. The text of the proposal, signed by the founding fathers John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon, leaves little room for ambiguity regarding the reducible nature of human intelligence. From its preamble, the proposal states a conjecture that remains one of the founding hypotheses informing major strands of AI research today:

The study is to proceed on the basis of the conjecture that every aspect of learning or *any other feature of intelligence* can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find out how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. [Emphasis added]

The document goes on to list seven aspects of the “artificial intelligence problem” that were to be studied during the summer of 1956. The relevance and prescience of this list, seventy years later, are astounding.

¹⁰ J. McCarthy, M. L. Minsky, N. Rochester, C. E. Shannon, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” *AI Magazine*, 27, 4, 2006, p. 12, <https://doi.org/10.1609/aimag.v27i4.1904>.

All the contemporary challenges of generative AI and large language models (LLMs) are foreshadowed: in §1, the “speeds and memory capacity of [then] present computers may be insufficient to simulate many of the higher functions of the human brain,” but, notably, it is immediately clarified that “the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have;” §2, the continuation of sentences from a word; §3, “How can a set of (hypothetical) neurons be arranged so as to form concepts;” §4, how to measure the complexity of computations with “a theory of the complexity of functions;” §5, how to enable “self-improvement” (a sign of a “truly intelligent machine”); §6, how to “form abstractions from sensory and other data.” The last aspect of the artificial intelligence problem mentioned is “§7 Randomness and Creativity.” On this last point (§7), crucial for rethinking architecture in the age of automata, the proposal put forward a hypothesis that resonates powerfully with current theories and the idea of an unpredictable automaton:

A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of a some [sic] randomness. The randomness must be guided by intuition to be efficient. In other words, the educated guess or the hunch include controlled randomness in otherwise orderly thinking.¹¹

This vision of “guided randomness” at the heart of orderly thinking lays the foundation for a non-deterministic conception of the automaton. Far from the rigid execution of fixed rules (the classical automation mentioned by Marco Vanucci¹²), artificial intelligence, as early as 1955, envisioned creativity as an injection of noise into the signal: a probabilistic deviation (deviating from a predictable trajectory). This suggests that the Dartmouth researchers were willing to treat creativity as at least partly compatible with stochastic search inside a formal system. But that conjecture remains heuristic: contemporary generative systems rely not on randomness alone, but on learned statistical structure plus stochastic sampling. As for intelligence, which includes, among other things, creativity and the capacity for self-learning, it is no longer treated as a

¹¹ *Ibid.*, p. 14. Emphasis added.

¹² *Khôrein: Journal for Architecture and Philosophy*, “Call for Papers: Volume IV, Issue 3: Automata.”

magical property of living beings but as a function that can, in principle, be modelled by sufficiently expressive computational systems, including neural architectures and Turing-equivalent formalisms. This is a crucial insight for understanding the evolution of architectural automata, from simple algorithms to contemporary generative AI. However, even if the theoretical triptych on which artificial intelligence from Dartmouth onwards is based—*quantified information* (Shannon), *universal processing procedure* (Turing), and *generation of novelty by guided randomness* (Dartmouth) still defines the horizon of current machines, the complexity of deep neural networks confronts us with a new challenge: that of extrapolation in high dimension, where controlled randomness is not enough to explain the relevance of the forms generated outside of any known corpus.

The Depsychologisation of Information

To understand how such a conjecture could have been formulated with such confidence in the mid-1950s, one must go back to the genesis of the concepts, notions, and logical and mathematical tools used by the Dartmouth organisers, as well as their backgrounds and knowledge of the state of the art at the time. Of the four organisers, only C. Shannon (born in 1916) possessed a genuine aura, rivalled only by Norbert Wiener and John von Neumann (A. Turing and K. Gödel were brilliant but less widely known theorists). Born in 1919, Nathaniel Rochester's reputation rested on his role at IBM as chief engineer of the IBM 701, the first commercial scientific computer, while the young PhDs John McCarthy and Marvin Minsky (born in 1927), though brilliant, primarily relied on their doctoral degrees from Princeton.¹³ Although only two of the four workshop organisers were senior scientists, all brought with

¹³ It may be noted that McCarthy completed his doctoral thesis, related to differential equations, under the supervision of Donald C. Spencer, a brilliant mathematician who had himself worked under J. E. Littlewood and G. H. Hardy at Cambridge. It is, however, highly unlikely that this influenced either the acceptance of the workshop or the securing of its funding. Moreover, McCarthy's biographical note in the Dartmouth proposal mentions neither of those names. See, J. McCarthy, M. L. Minsky, N. Rochester, C. E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," (op. cit.), § "Proposal for Research by John McCarthy" (p. 10) and § "People Interested in The Artificial Intelligence Problem" (pp. 11–13); see J. McCarthy, *Projection Operators and Partial Differential Equations*, PhD Dissertation n°0010958, Princeton University, 1951. <https://www.proquest.com/dissertations-theses/projection-operators-partial-differential/docview/302039611/se-2?accountid=14511>

them the latest and most fruitful scientific concepts. Besides Shannon and Rochester, who were obviously well-versed in the state of the art, McCarthy and Minsky had a precise knowledge of the work of John von Neumann,¹⁴ Norbert Wiener,¹⁵ Alan Turing,¹⁶ Warren McCulloch, Walter Pitts,¹⁷ and of course C. Shannon.¹⁸ This list is obviously not exhaustive, but it is worth noting that all the concepts necessary for the study of “some aspects of the artificial intelligence problem” as defined in the Dartmouth Proposal, had been discussed by the scientists just mentioned. These concepts were, for the most part, perfectly formalised (i.e. mathematically defined), but, even more remarkably, they existed in very limited numbers. The project to process human intelligence was thus developed from the outset as a reductionist project based on a minimal number of mathematically sound hypotheses, as evidenced by the bibliographies of the works of Shannon and Minsky¹⁹ that preceded Dartmouth’s proposal. C. Shannon’s article, “A Symbolic Analysis of Relay and Switching Circuits,” published on recommendation in December 1938 in the year-end supplement of the *Transactions of the American Institute of Electrical Engineers*,²⁰ for example, contains only four bibliographic references. It is indeed a summary of his Master of Science thesis, prepared under the supervision of Vannevar Bush at MIT, but as

¹⁴ Cf. J. von Neumann, *First Draft of a Report on the EDVAC*, Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, 1945; J. von Neumann, *Theory and Techniques for Design of Electronic Digital Computers* [lectures given at the Moore School, 8 July–31 August 1946], Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, 1948.

¹⁵ N. Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine*, Hermann & Cie, Paris, 1948.

¹⁶ See A. M. Turing, “On Computable Numbers, with an Application to the Entscheidungsproblem,” *Proceedings of the London Mathematical Society*, 2, 42, 1937, pp. 230–265; A. M. Turing, “Computing Machinery and Intelligence,” *Mind*, 59, 236, 1950, pp. 433–460; A. M. Turing, “Intelligent Machinery,” in J. Copeland (ed.), *The Essential Turing*, Clarendon Press, Oxford, 2004, pp. 410–432.

¹⁷ W. S. McCulloch and W. Pitts, “A Logical Calculus of the Ideas Immanent in Nervous Activity,” *Bulletin of Mathematical Biophysics*, 5, 1943, pp. 115–133.

¹⁸ See C. E. Shannon, “A Symbolic Analysis of Relay and Switching Circuits,” *Transactions of the American Institute of Electrical Engineers*, 57, 1938, p. 713; C. E. Shannon, “A Mathematical Theory of Communication,” *The Bell System Technical Journal*, 27, 1948, pp. 379–423, 623–656.

¹⁹ M. L. Minsky, “Theory of Neural-Analog Reinforcement Systems and Its Application to the Brain-Model,” doctoral dissertation (n°0009438), Department of Mathematics, Princeton University, Princeton, 1954, pp. 6–64. ISBN 9781084420847. https://people.csail.mit.edu/brooks/idocs/Marvin_Minsky_PhD_Princeton_1953_1954.pdf

²⁰ C. E. Shannon, “A Symbolic Analysis of Relay and Switching Circuits,” *Transactions of the American Institute of Electrical Engineers*, 57, 1938, p. 713.

Herman H. Goldstine would later say, “This surely must be one of the most important master’s theses ever written. [...] The paper was a landmark in that it helped change digital circuit design from an art to a science.”²¹ By demonstrating the perfect analogy between Boolean algebra (binary propositional logic) and electromechanical circuits (open/closed relays), Shannon paved the way for the physical materialisation of logical thought, which the invention of transistors would usher into a new era.²² He proved that reasoning could be embodied in inert matter. However, it was with his *Mathematical Theory of Communication* (1948) that Shannon made the most radical conceptual break, the one that would make AI theoretically conceivable. This was to answer a question he had already posed in 1939 in a letter to his mentor, V. Bush, outlining some ideas on “the fundamental properties of general systems for the transmission of intelligence:” is there a unified theory of communication?

To achieve this, Shannon had to build upon the work of his predecessors, notably Henry Nyquist and Ralph Hartley. In 1928, Hartley had already suggested, in “Transmission of Information,”²³ replacing the vague²⁴ and anthropomorphic term “intelligence” used by Nyquist (who spoke of the “speed of intelligence transmission”²⁵) with the term “information,” which he precisely defined to avoid any ambiguity. Hartley advocated an “elimination of psychological factors”²⁶ to establish a

²¹ H. H. Goldstine, *The Computer from Pascal to von Neumann*, Princeton University Press, Princeton, 1980, pp. 119–120.

²² Shannon spoke of the invention of the transistor as “[...] quite possibly the greatest invention of the twentieth century.” Claude Shannon, Kyoto Prize speech draft, 1985, Claude Elwood Shannon Papers, Manuscript Division, Library of Congress, Washington, D.C.

²³ R. V. L. Hartley, “Transmission of Information,” *The Bell System Technical Journal*, 7, 1928, pp. 535–563. According to *The Bell System Technical Journal*, the article corresponds to a communication delivered in September 1927 at the International Congress of Telegraphy and Telephony on Lake Como. More precisely, it was part of the Volta Congress, a multidisciplinary event held from 11 to 27 September 1927.

²⁴ “When we speak of the capacity of a system to transmit information we imply some sort of quantitative measure of information. As commonly used, information is a very elastic term, and it will first be necessary to set up for it a more specific meaning as applied to the present discussion.” R. V. L. Hartley, “Transmission of Information,” *The Bell System Technical Journal*, 7, 1928, p. 536.

²⁵ H. Nyquist, “Certain Factors Affecting Telegraph Speed,” *The Bell System Technical Journal*, 3, 2, 1924, pp. 324–346; H. Nyquist, “Certain Topics in Telegraph Transmission Theory,” *Transactions of the American Institute of Electrical Engineers*, 47, 2, 1928, pp. 617–644.

²⁶ “By this means the psychological factors and their variations are eliminated and it becomes possible to set up a definite quantitative measure of information based on physical

quantitative measure based solely on physical considerations. Shannon radicalised this “depsychologisation.” In his 1948 text, he laid out the dogma that definitively separated engineering from semantics and enabled universal quantification:

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently *the messages have meaning* [...]. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that *the actual message is one selected from a set of possible messages*.²⁷ [Emphasis added]

This exclusion of meaning is the founding act of the information age. By defining information not as meaning, but as a statistical measure of the improbability of a message within a finite set of possibilities (entropy), Shannon enables its rigorous quantification in bits. The consequence for architecture is deep: architecture, like any other cultural product, becomes translatable into a series of binary choices, independent of its “poetic,” “sensory,” or intentional content. From this perspective, architectural quality can be approached not as an irreducible essence, but as a statistically inflected configuration within a space of possible forms.

The Problem of Defining Intelligence

While information theory resolved the “fundamental problem of communication” by discarding semantics, the “artificial intelligence problem” formulated at Dartmouth reopened Pandora’s box. Intelligence, unlike Shannon’s information, resists a purely quantitative definition. It seems intrinsically linked to meaning, adaptation, and understanding—all

considerations alone.” R. V. L. Hartley, “Transmission of Information,” *The Bell System Technical Journal*, 7, 1928, p. 538.

²⁷ “The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design. If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely.” C. E. Shannon, “A Mathematical Theory of Communication,” pp. 379–380.

things Shannon had set aside “for the sake of engineering.” The definition proposed at Dartmouth—“making a machine behave in ways that would be called intelligent if a human were so behaving”²⁸—is a working definition, but it remains philosophically unsatisfactory because it is circular. It shifts the problem to behavioural imitation without explaining the nature of the phenomenon. We are faced with a logical circularity: intelligence is defined as thinking, but thinking is taken as the defining characteristic of the intelligent agent. Faced with this ontological impasse, and the absence of a unified measure of intelligence (comparable to entropy for information or the joule for energy), the pioneers of AI turned to another rigorous formalisation: that of computation. This is where the guiding figure of Alan Turing comes in. Even before the invention of the physical computer, Turing had, in his 1936 paper (published in 1937)²⁹, formalised the notions of computation and abstract machine. This formalisation made it possible to consider the mind no longer as an ethereal substance (*res cogitans*), but as a logical process that can, in principle, be simulated by a computational machine. The “computational generalisation”—this reductionist yet fruitful thesis that the human brain is an information-processing computer—finds its roots in this historical convergence between information theory (Shannon) and computability theory (Turing, Church, Kleene).³⁰ For contemporary architecture, this means that “the intelligence of architecture” can no longer be conceived as a mystical intuition or an innate genius. It becomes a

²⁸ “For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving.” J. McCarthy, M. L. Minsky, N. Rochester, C. E. Shannon, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” p. 12.

²⁹ Cf. A. M. Turing, “On Computable Numbers, with an Application to the Entscheidungsproblem.”

³⁰ Although the computationalist thesis was massively developed and defended after the Dartmouth workshop, its organisers were clearly not operating in a vacuum. This is evidenced by N. Rochester’s reference to K. J. W. Craik, *The Nature of Explanation*, Cambridge University Press, Cambridge, 1943, and M. L. Minsky’s close knowledge of W. S. McCulloch, W. Pitts, “A Logical Calculus of the Ideas Immanent in Nervous Activity,” *Bulletin of Mathematical Biophysics*, 5, 1943, pp. 115–133; W. S. McCulloch, W. Pitts, “How We Know Universals: The Perception of Auditory and Visual Forms,” *Bulletin of Mathematical Biophysics*, 9, 1947, pp. 127–147; and S. C. Kleene, “Representation of Events in Nerve Nets and Finite Automata,” U. S. Air Force Project Rand, Research Memorandum 704 (RM-704), RAND Corporation, Santa Monica, 15 December 1951. https://www.rand.org/content/dam/rand/pubs/research_memoranda/2008/RM704.pdf On this subject, see also the bibliography of M. L. Minsky, “Theory of Neural-Analog Reinforcement Systems and Its Application to the Brain-Model,” p. 64.

problem of information processing and computational complexity, analysable through the tools forged by science. Artificial intelligence does not solve the mystery of architecture; it shifts the question from the realm of aesthetics to that of the computable. Importantly, the apparent re-emergence of meaning in contemporary AI systems does not imply a reintroduction of semantics at the level of formal operations. Meaning is not explicitly represented in current models, as they manipulate symbols, tokens, and vectors without any explicit human-like grounding procedure built into the computation itself. What users call “meaning” emerges only as an observer-relative effect, produced when statistically coherent outputs are interpreted by human agents embedded in cultural and historical contexts. With this in mind, and in order for computation to become “creative” and autonomous as suggested by the call for *Khôrein*’s special issue, Automata, we need to understand how the simple execution of rules (Turing) can give rise to the unexpected. This is what we will examine in the next section, by analysing the birth of the notion of computability and its limits.

PART 2. THE ARCHITECTURE OF COMPUTATION

From Information to Effective Procedures

While the first part of this essay established how Shannon’s mathematical theory of communication succeeded, as early as 1948, in “depsychologising” information by reducing it to a quantitative measure of uncertainty and selection—thereby eliminating the cumbersome question of meaning for engineering purposes—it nevertheless leaves unresolved a fundamental question for the advent of artificial intelligence: that of processing. Information, as defined by Shannon, is a “physical quantity,” storable and transmissible, comparable to a raw material. But for intelligence—or at least a semblance of architectural intelligence—to exist, this information cannot remain inert; it must be manipulated, transformed, processed. It must enter a logical dynamic.

It is imperative to understand that computational architecture is not based solely on data storage (the Shannon bit), but on the ability to execute logical rules on that data. This is where the second formal revolution of the twentieth century comes in, intimately linked to the first but conceptually distinct: the birth of computability theory. Even before the computer became a physical reality made of vacuum tubes, relays, and

then transistors, it existed as a pure logical necessity. This theoretical existence preceded its material existence by nearly a decade. To think about the intelligence of architecture today, and to understand how systems can autonomously generate forms or solutions, we must return to this epistemological turning point when the intuitive notion of “computation” became a rigorous mathematical object: the formalisation of the universal machine.

The Crisis of Foundations and the Problem of Decision

The deep origin of artificial intelligence is not found in a naive biomimetic attempt to imitate the biological brain,³¹ but in an internal crisis within mathematics and formal logic: the decision problem, or *Entscheidungsproblem*. Formulated by the mathematician David Hilbert in 1928, this problem posed a question of apparent simplicity but of abysmal complexity for the time: is there a general effective procedure—what we would call an algorithm today—that allows us to determine, for any given well-formed statement in a formal system, whether it is derivable or valid within the relevant formal framework? This question was not trivial. It concerned the very possibility of a complete and coherent mathematics, where truth would be accessible through a systematic method, without recourse to the mathematician’s “magical” intuition. To answer this question, it was imperative to rigorously define what was meant by “mechanical procedure,” “effective method,” or “calculus.” Until 1936, these notions remained relatively vague and intuitive. People knew how to calculate, but they did not know how to define calculation in its universality. While initially aiming to provide an answer to the problem of pure mathematics posed by David Hilbert, Turing would in fact be led to formalise these different notions.

It was in this context of a crisis of foundations that Turing provided the definitive formalisation in his 1936 article, “On Computable Numbers, with an Application to the Entscheidungsproblem.” The contribution of this text extends far beyond the realm of pure mathematics to touch upon the very structure of rational thought. By seeking to define what a “human calculator” (a computer, which at the time referred to

³¹ The work of McCulloch and Pitts did not aim merely to mimic the functioning of the brain without understanding it, but to understand as rigorously as possible the processes at work in thought; see W. S. McCulloch, W. Pitts, “A Logical Calculus of the Ideas Immanent in Nervous Activity,” *Bulletin of Mathematical Biophysics*, 5, 1943, pp. 115–133.

a profession and not a machine) does when applying a method, Turing “mechanised” the mind. He proposed the concept of an “automatic machine” (or a-machine), an abstract device with infinite memory (a strip of paper divided into cells) and capable of performing elementary operations on symbols (reading, writing, erasing, moving the read head) according to a finite table describing the so-called “behaviour of the machine.” This device, which history will remember as the Turing Machine, is the absolute formalisation of the automaton: a system that transitions from one state to another according to a pre-established rule, without ambiguity or intuition. By showing that this simple machine could calculate any number definable by a method, Turing closed the debate on the nature of computation and demonstrated the undecidability of the halting problem, thus rejecting Hilbert’s dream.

The Universal Machine and the Separation of Software

Turing’s genius did not stop at describing a machine capable of performing a specific calculation. In the same 1936 paper, he demonstrated an astonishing property: it was possible to design a single, universal machine capable of simulating the behaviour of any other machine. The principle was as follows: if the instruction table of a specific machine (for example, a machine dedicated to addition or geometry generation) could be described in symbolic form, and if this description were written onto the tape of the universal machine, then the universal machine would virtually “become” the described machine. This demonstration marks a decisive conceptual step toward modern computing. It prefigures the separation between a universal machine and the specific instructions it executes. This crucial distinction quickly became the conceptual model for architecture, whose design logic (rules, style, spatial syntax) would then be encoded, transmitted, and executed independently of the hardware and even independently of the rules, style, or spatial syntax involved. Hence the idea of a kind of metalanguage for architecture, associated with a new figure of the architect who no longer designs each project but programs, in a universal language (for example, a Turing-Complete programming language), a similarly universal machine (the computer as we know it today, or the mainframe computers of the 1960s and 70s that followed ENIAC and UNIVAC) which will design each specific project based on changing parameters. Beyond the unimportant question of style, this approach was followed by the pioneers of digital architecture

from the mid-1960s onward (e.g. N. Negroponte,³² C. Price,³³ and later J. Frazer³⁴), by proponents of formal grammars³⁵ (e.g. G. Stiny,³⁶ J. Gero³⁷), and by virtually all subsequent computational architecture. Architects active within what M. Carpo called the “first digital turn”³⁸ of the 1990s held a more ambiguous position regarding programmed and *ultimately* automated architecture. Nevertheless, some, including Kas Oosterhuis, further strengthened the link between programming and architecture by advocating a dynamic architecture whose constant evolution is guided by its own reprogramming.³⁹ Regardless of the “technical details” and other specificities of the various approaches to architectural design, as soon as architects had access to computers, the new reality that emerged was modelled on the shift in “computer science”⁴⁰ from specialised calculators to universal machines for symbolic manipulation, capable of simulating anything, provided the task can be described by finite rules. This universality is the condition of possibility for the “latent engineering” and pervasiveness mentioned above: a ubiquitous computing capacity, ready to receive any architectural instruction.

The Church-Turing Thesis

Turing’s formalisation of computability did not occur in isolation. It was part of an exceptional intellectual convergence that mobilised the greatest logicians of the time: Alonzo Church, Kurt Gödel, Emil Post, and Stephen Kleene. While Turing was imagining tape machines, Alonzo

³² See N. Negroponte, *The Architecture Machine: Toward a More Human Environment*, MIT Press, Cambridge, Mass., 1970; N. Negroponte, *Soft Architecture Machines*, MIT Press, Cambridge, Mass., 1975.

³³ S. Hardingham (ed.), *Cedric Price Works 1952–2003: A Forward-Minded Retrospective*, 2 vols., Architectural Association / Canadian Centre for Architecture, London and Montreal, 2016.

³⁴ J. Frazer, *An Evolutionary Architecture*, Architectural Association, London, 1995.

³⁵ On this subject, see S. Kotsopoulos (ed.), *Shape Computation: Fifty Years, 1972–2022*, Birkhäuser, Cham, 2025, <https://link.springer.com/book/10.1007/978-3-031-81623-9>.

³⁶ G. Stiny and J. Gips, “Shape Grammars and the Generative Specification of Painting and Sculpture,” in C. V. Freiman (ed.), *Information Processing 71*, North-Holland, Amsterdam, 1972, pp. 1460–1465; G. Stiny, J. Gips, *Algorithmic Aesthetics: Computer Models for Criticism and Design in the Arts*, University of California Press, Berkeley, 1978.

³⁷ J. S. Gero (ed.), *Design Optimization*, Academic Press, Orlando, 1985.

³⁸ M. Carpo, *The Alphabet and the Algorithm*, MIT Press, Cambridge, Mass., 2011.

³⁹ K. Oosterhuis, *Programmable Architecture*, L’Arca Edizioni, Milan, 2002; K. Oosterhuis, *Hyperbodies: Towards an E-motive Architecture*, Birkhäuser, Basel, 2003.

⁴⁰ It goes without saying that, in the period from Babbage to ENIAC, people were not yet speaking of “computer science.”

Church was developing a different formal system at Princeton, lambda calculus, and Stephen Kleene was working on recursive functions, a work he would pursue in relation to living organisms notably in his research paper, “Representation of Events in Nerve Nets and Finite Automata,”⁴¹ in which he claims to have “[...] demonstrated that McCulloch-Pitts neurons can govern any kind of behavior which any other kind of digital automaton at all can govern.”⁴²

It emerged, almost miraculously, that all these formal definitions of “computable,” despite their very different approaches (mechanical for Turing, algebraic for Church), were mathematically equivalent. Everything computable by lambda calculus is computable by a Turing machine, and vice versa. This observation led to the formulation of the famous “Church-Turing Thesis.”⁴³ It is not a proved theorem but a foundational and widely accepted thesis about effective computation, establishing that every physically realizable or mentally possible process is Turing-computable, or in other words that there is an equivalence between the intuitive notion of “effective computability” (what a human can calculate with paper, a pencil, infinite discipline, and without any particular genius) and computability by a Turing machine (itself equivalent to being λ -definable). The philosophical implications of this thesis for artificial intelligence—and for the intelligence of architecture—are, of course, crucial. If, as the proponents of computationalism would later do, we accept that human cognitive processes (logical reasoning, pattern recognition, architectural decision-making) are information processing processes, then, within a computationalist framework the Church–Turing thesis supports the claim that formalizable cognitive processes can, in principle, be simulated by a machine. The thesis establishes a kind of “epistemological closure,” which does not mean that everything computable is computed by existing machines, nor that every intelligent phenomenon is trivially programmable. It simply means that no formalizable cognitive process requires, in principle, a non-computational substrate. Thus, there is no magical computation that the brain could perform that would inherently elude the universal machine.

⁴¹ S. C. Kleene, “Representation of Events in Nerve Nets and Finite Automata.”

⁴² “In particular, we have thus demonstrated that McCulloch-Pitts neurons can govern any kind of behavior which any other kind of digital automaton at all can govern.” S. C. Kleene, “Representation of Events in Nerve Nets and Finite Automata,” p. 72.

⁴³ J. Copeland, “The Church-Turing Thesis,” *NeuroQuantology*, 2, 2, 2004, pp. 101–115, <https://doi.org/10.14704/NQ.2004.2.2.40>.

Consequently, intelligence becomes, *in principle*, an object of engineering. It was this theoretical certainty that later allowed the organisers of the Dartmouth conference to assert with almost unwavering confidence that “[...] every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.” Indeed, Turing’s 1936 formalisation and the equivalent formalisations of the concept of computation by Church, Gödel, Post, and Kleene transformed potential intelligence into a programming problem.

Intelligence as Performance

If the 1936 machine was a mathematical abstraction intended to resolve a problem of pure logic (and to answer Hilbert in the negative: no, not everything is decidable, since the halting problem is not), the postwar period saw the question of machine intelligence emerge on an unmistakably concrete level. In 1950, Alan Turing published in the philosophical journal *Mind* the article that is considered the symbolic birth of AI: *Computing Machinery and Intelligence*.⁴⁴ This text is fundamental because it introduces a strategic displacement similar to the one carried out by Shannon two years earlier. Where Shannon eliminated meaning to measure information, Turing eliminates consciousness to measure intelligence. From 1950 onward, two problems haunted research: that of testing and that of definition. Traditionally, intelligence is associated with the capacity to think, but the definition is, as mentioned above, circular. Faced with this difficulty and the “dangerous attitude” of referring to the common usage of the terms “machine” and “think,”⁴⁵ Turing proposes in his article to consider the problem from the perspective of a game—the “imitation game”:

It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the

⁴⁴ A. M. Turing, “Computing Machinery and Intelligence.”

⁴⁵ “The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words ‘machine’ and ‘think’ are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, ‘Can machines think?’ is to be sought in a statistical survey such as a Gallup poll. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.” *Ibid.*

interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either “X is A and Y is B” or “X is B and Y is A.” The interrogator is allowed to put questions to A and B. [...]

We now ask the question, “What will happen when a machine takes the part of A in this game?” Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, “Can machines think?”⁴⁶

We know that, although it is anthropomorphic and therefore problematic, this exact device described by Turing contains the seeds of the entire philosophy of AI that would follow, whether that AI is symbolic (“GOFAI—Good Old Fashioned AI”) or connectionist (the “current” AI of deep neural networks). Like Shannon, who succeeded in depsychoanalysing information theory, Turing sought to remove the problem of intelligence from the qualitative human approach in favour of a quantitative statistical one. By substituting the machine for the human in a social simulation game, Turing defined intelligence not by an internal substance (the soul, consciousness, biology), but by an external performance indistinguishable from that of a human. It is a statistical definition: if the machine succeeds in deceiving the interrogator as often as a human, it is intelligent, regardless of its internal mechanisms. Intelligence ceases to be an essence and becomes a capacity for successful simulation. For architecture, this redefinition has major consequences. It suggests that the intelligence of a project lies not in the author’s intention, but in the effect produced by the artifact. It matters little whether the machine “feels” the emotion of a space or understands the history of architecture. Provided it can generate plans, facades, or structures—in reality, any elements traditionally associated with architecture, and which in fact represent a genuine limit of the discipline that must precisely be overcome—deemed relevant, coherent, and culturally valid by human experts (the interrogators), it demonstrates architectural intelligence. The automaton becomes a producer of valid signs.

⁴⁶ *Ibid.*

Probabilistic Autonomy and Opacity

The combined legacy of Turing and Dartmouth allows us to understand the precise nature of the “new architectural intelligences” discussed in the first part. We are no longer dealing with simple optimisation tools (e.g. cost optimisation, structural element sizing, quantification of light inputs, or energy balance), but with systems capable of simulating complex behaviours. Recent technological developments, however, have introduced a form of complexity that was not anticipated, at least not in its full phenomenological scope: opacity. The classical Turing machine is deterministic and legible: a given state corresponds to a single action. One can unroll the tape and follow the computation step by step. Contemporary artificial intelligence systems, by contrast, especially those based on deep learning in high-dimensional spaces, operate differently. They rely on nonlinear and probabilistic logics applied to data spaces so vast that their internal functioning becomes only partially intelligible to us. This opacity does not contradict the computational framework inherited from Turing and Dartmouth, but it does show that computability does not entail transparency. If intelligence is a statistical simulation (Turing) and if creativity is guided randomness (Dartmouth), then it is natural that the most efficient systems are those that exploit these statistical dimensions to the fullest, to the point where causal relations become difficult for the human mind to trace. The contemporary architectural automaton no longer simply calculates optimal solutions in the engineer’s sense; it explores spaces of possibilities by simulating qualitative judgments. This transition from deterministic interpolation⁴⁷ to an opaque probabilistic output establishes the basis of the central problem addressed in the third part of this article.

If the automaton no longer simply computes what it has been taught, where does it find its solutions? How can intelligence emerge from raw information without simply regurgitating it, without being a mere “stochastic parrot” responding with probable solutions? Solutions which, moreover, are generally the most probable, while the least probable are often—wrongly—associated with “creativity” obtained by a greater injection of randomness, materialised for example by a high value for the

⁴⁷ Deterministic interpolation is a known deterministic transformation of data and is therefore, in principle, simulable by a Turing machine. It can be understood as the search for a stable deterministic point in a value space, equivalent to a specific final state of the tape of a Turing machine. In practice, however, the operation would be far from simple.

“temperature” parameter of generative AI interfaces. This false creativity—in fact what we could call “semantic hallucination” which results from failures of alignment or contextual grounding, should be distinguished from what we refer to, for analytical purposes, as “geometric hallucination,” which is a structural consequence of extrapolation in high-dimensional spaces. The latter is not a malfunction to be eliminated but the necessary condition under which novel and coherent forms can be generated. Returning to the question “how can intelligence emerge from raw information,” for some vehement critics, including N. Chomsky, the answer is quite simple: no genuine intelligence (i.e. at least equal to human intelligence) can be produced by machines.

The human mind is not, like ChatGPT and its ilk, a lumbering statistical engine for pattern matching, gorging on hundreds of terabytes of data and *extrapolating the most likely conversational response or most probable answer* to a scientific question. On the contrary, the human mind is a surprisingly efficient and even elegant system that operates with small amounts of information; it seeks not to infer brute correlations among data points but to create explanations.⁴⁸ [Emphasis added]

For the American linguist, in addition to the ability to create explanations (on which we agree), “True intelligence is demonstrated in the ability to think and express improbable but insightful things.”⁴⁹ While it is true that one can expect true intelligence to possess the capacity to think improbable but insightful things, it is worth dwelling a little longer on the notion of probability, and on an important mathematical distinction between interpolation and extrapolation, which, in our view, becomes a decisive critical tool for understanding the nature of artificial creativity. Faithful to the Dartmouth programme, the contemporary automaton does not simply calculate the known; most of the time—almost by default, so to speak—it “hallucinates” the possible to varying degrees. The real difficulty, therefore, is not so much how to make AI systems creative by

⁴⁸ N. Chomsky, I. Roberts, J. Watumull, “The False Promise of ChatGPT,” *The New York Times*, <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>, (accessed 8 April 2026).

⁴⁹ “True intelligence is demonstrated in the ability to think and express improbable but insightful things.” *Ibid.*

forcing them out of an “interpolation regime”⁵⁰ in order to prevent them from repeating what is already known—whether highly probable or, on the contrary, extremely unlikely—but rather how to control their hallucinations, insofar as their operation appears to involve a significant extrapolative component rather than, as most critics suggest, mere interpolation.

PART 3. BEYOND IMITATION: EXTRAPOLATION AND AUTONOMY

Generative AI as a Machine for Imitation?

While the first two parts of this essay traced the historical and epistemological arc from Claude Shannon’s quantification of information to Alan Turing’s formalisation of the effective procedure, we now encounter a dominant critical interpretation that threatens to reduce artificial intelligence to a mere echo chamber of the past. Faced with the increasing opacity of so-called, and in some respects mythical, “black box” systems, a strong theoretical temptation is to reduce the unknown—the internal functioning of the neural network⁵¹—to the known: the history of art, the history of architecture, and more generally the accumulated archive of past human productions. On this view, AI systems would merely recombine what already exists. This position was recently articulated with clarity by the architectural historian Mario Carpo. In his recent work, and particularly in his essay “Generative AI, Imitation Learning, and the Automation of Tacit Knowledge,”⁵² Carpo argues that generative artificial intelligence—whether based on generative adversarial networks (GANs)

⁵⁰ The expression “interpolation regime” is used loosely here. Strictly speaking, the argument follows the definition proposed by R. Balestrieri, J. Pesenti, Y. LeCun: “Before going further, we insist that throughout this manuscript, interpolation is to be understood as characterizing the data geometry as per Def. 1. This is not to be mistaken with the often employed interpolation regime of models which occur whenever the latter has 0 training loss on the data [...]”. R. Balestrieri, J. Pesenti, Y. LeCun, “Learning in High Dimension Always Amounts to Extrapolation,” *arXiv*, 2021, p. 1. <https://doi.org/10.48550/arXiv.2110.09485>, (accessed 8 April 2026). On the distinction from the model regime defined by zero training loss, see N. S. Chatterji, P. M. Long, P. L. Bartlett, “When Does Gradient Descent with Logistic Loss Find Interpolating Two-Layer Networks?,” *Journal of Machine Learning Research*, 22, 159, 2021, pp. 1–48.

⁵¹ Although, mathematically speaking, these processes are not strictly unknown, the depth of the networks and the number of parameters make them particularly opaque to most human users.

⁵² M. Carpo, “Generative AI, Imitation Learning, and the Automation of Tacit Knowledge,” *A+U*, 646, 2024, p. 106.

or large language models (LLMs)—is, “in epistemological terms, a machine that automates imitation.”⁵³ For Carpo, these technologies remind us of a “[...] universal condition of human creativity: namely, that nothing is created out of nothing, but all is generated out of something that already exists.”⁵⁴ From this perspective, AI does not constitute a rupture but rather a return to the long European tradition of mimesis, understood not as mechanical copying but as a process of “creative imitation” based on “the awareness, acknowledgement, election, selection, assimilation, and intelligent transformation, of a corpus of models.”⁵⁵

This interpretation is compelling because it situates AI within a reassuring cultural continuity, that of the European humanist and classical tradition, and more recently that of the variations⁵⁶ of the 1990’s “first digital turn.”⁵⁷ However, it rests on a geometric intuition derived from human-scale experience, an intuition that collapses when transposed into the high-dimensional spaces in which contemporary AI systems operate. While Carpo’s thesis remains valid as an anthropology of human creativity, it becomes difficult to sustain when elevated from an anthropology of human creativity to a strict ontology of machine generation. To translate imitation into operational terms, one must abandon metaphor and adopt a mathematical definition. In learning theory, imitation corresponds to interpolation: the generation of new samples that lie within the convex hull of the training data. Yet, as Balestrieri, Pesenti, and LeCun⁵⁸ demonstrate, under a strict convex-hull definition, interpo-

⁵³ “Generative AI, either based on legacy GANs or on more recent large language models, is, in epistemological terms, a machine that automates imitation.” *Ibid.*, p. 106.

⁵⁴ The full passage is: “Generative AI, either based on legacy GANs or on more recent large language models, is, in epistemological terms, a machine that automates imitation; it is a powerful—albeit purely technical—reminder of a basic, universal condition of human creativity: namely, that nothing is created out of nothing, but all is generated out of something that already exists. Just like Generative AI can only invent new things based on—but in fact derived from—a dataset of precedents, so all human invention entails a component of creative imitation: whereby “imitation” means, as in the Greek and Roman, humanistic and classical tradition, the awareness, acknowledgment, election, selection, assimilation, and intelligent transformation, of a corpus of models.” *Ibid.*, p. 106.

⁵⁵ *Ibid.*, p. 106.

⁵⁶ We can recall here how the theme of variation was developed during the first digital turn, particularly with Greg Lynn and Lars Spuybroek. See G. Lynn, “New Variations on the Rowe Complex,” *ANY: Architecture New York*, 7–8, 1994, pp. 38–43; L. Spuybroek, *Research & Design: The Architecture of Variation*, Thames & Hudson, London, 2009.

⁵⁷ Cf. M. Carpo, *The Alphabet and the Algorithm*; M. Carpo (ed.), *The Digital Turn in Architecture 1992–2012*, Wiley, Chichester, 2012.

⁵⁸ R. Balestrieri, J. Pesenti, Y. LeCun, “Learning in High Dimension Always Amounts to Extrapolation.”

lation in high-dimensional spaces becomes vanishingly unlikely as dimensionality increases. As dimensionality increases, the probability that any new point lies within the convex hull of existing data converges toward zero, unless the dataset grows exponentially with dimension—a condition that no finite corpus can satisfy. Consequently, under this geometric definition, contemporary generative AI systems cannot be described primarily as navigating within the envelope of precedents; they frequently generate outputs that lie beyond it. What appears as stylistic continuity is not the result of interpolation but of structural constraints shared by the data themselves, which occupy an extraordinarily sparse space. Generation is therefore not a recombination of known forms but an extrapolative projection into regions of space that have never been sampled. From a geometric standpoint, strict interpolation-based imitation becomes highly implausible in high-dimensional settings.

A Geometric Definition of Imitation

To carry out this comparison, we must, beyond metaphors, translate the literary and artistic concept of imitation into an operational mathematical concept. If, as Carpo suggests, to imitate means to produce a plausible variation from a corpus of precedents, geometrically this would correspond to the operation of interpolation.⁵⁹

In function approximation theory, as in machine learning, interpolation, in its strictest definition, consists of calculating a new value or a new point that lies between known values. If we imagine the entirety of existing architectural knowledge (all plans, all images, all texts, and even, for our thought experiment, all fine-grained 3D scans of every building ever constructed) as a point cloud $X = \{x_1, x_2, \dots, x_N\}$ in a vector space, “imitating” would consist of generating a new project x that lies geometrically within the envelope formed by these points. More precisely, we can hypothesise, and even use as a definition of imitation, that it is nothing other than an interpolation, or possibly an infinitesimal extrapolation, of a point existing in a space of any dimension. This geometric definition aligns with our intuition, common practices, and examples taken by Carpo: if we copy a style, we remain close to examples of that style; if we blend two styles, we position ourselves “between”

⁵⁹ Although the mathematical problems of interpolation and extrapolation in AI models are complex, the present argument remains valid.

them. If AI were an imitation machine in Carpo's sense, it would navigate cautiously within the territory defined by its training data. It would produce weighted averages of Le Corbusier and Mies van der Rohe. It would remain, to use the precise topological term, within the convex hull of the training data.

In machine learning, according to R. Balestrieri, J. Pesenti, and Y. LeCun, the common adage has long been, "as an algorithm transitions from interpolation to extrapolation, as its performance decreases." [*sic*] In this classical view, interpolation is the domain of mastery and faithful reproduction (successful imitation); extrapolation is the domain of uncertainty, error, and dangerous hallucination. However, recent discoveries in deep learning theory contradict this reassuring model. As Randall Balestrieri, Jérôme Pesenti, and Yann LeCun demonstrate in their 2021 article "Learning in High Dimension Always Amounts to Extrapolation" that the intuition that neural networks interpolate is not only false but mathematically almost impossible when dealing with complex data.

The Curse of Dimensionality

The core of the argument rests on the counterintuitive behaviour of geometry in high dimensions. An architectural image, even a low-resolution one, or a text describing a project, exists in mathematical spaces with thousands, even millions, of dimensions ($d > 100$); moreover, the more multimodal the description, the larger the space will be. Balestrieri, Pesenti, and LeCun first recall the canonical definition of interpolation that we use: "Interpolation occurs for a sample x whenever this sample belongs to the convex hull of a set of samples $X \triangleq \{x_1, \dots, x_N\}$, if not, extrapolation occurs."⁶⁰ This definition does not depend on the local metric or the intuitive notion of perceptual resemblance; it depends solely on the global affine geometry of the dataset, which renders any analogy with human imitation based on local semantic proximities mathematically irrelevant. They then rely on a result from stochastic geometry—the Bárány and Füredi theorem (1988)—to conclude that in high dimensions, interpolation "almost surely never occurs." Their proof revolves around an asymptotic theorem that describes the behaviour of the interpolation probability as the dimension d tends to infinity. The conclusion is that to remain within the framework of interpolation, a virtually

⁶⁰ *Ibid.*, p. 1. <https://doi.org/10.48550/arXiv.2110.09485>, (accessed 8 April 2026).

infinite amount of training data would be required, with the amount of data needed increasing exponentially with the dimension.⁶¹

The general result proposed by Balestrieri *et al.* is, beyond learning theory, conceptually very interesting for addressing architecture in the age of artificial intelligence, and particularly the imitation theory proposed by M. Carpo. By stating that in order to maintain a non-zero chance that a new point will fall inside the convex hull of the existing data (i.e. for imitation/interpolation to occur), the size of the dataset must grow exponentially with the dimension d , the Balestrieri *et al.* result amounts to asserting that no amount of data, however large, will ever be sufficient to make the absence of hallucinations near-certain or, with regard to architecture, to secure the kind of humanist rationality characteristic of the theory of creative imitation. While we humans are overwhelmed by data, contrary to the common intuition that leads us to project our perceptual and computational abilities onto machines, for machines, the entirety of the data ever produced represents only an infinitesimal fraction of what they might need and potentially use. Despite now possessing trillions of parameters, AI models are, to borrow a famous epistemological thesis, still “underdetermined” by reality.⁶² Moreover, even if we were to produce an immense quantity of synthetic data, this quantity would likely remain negligible compared to the results highlighted by Balestrieri *et al.* Indeed, as soon as the spatial dimension exceeds a few dozen (remember that a simple “style vector” in a latent space often has 512 or 1024 dimensions), the amount of training data—in architecture, for example, projects, buildings, construction data, etc.—required to make convex-hull interpolation common grows so rapidly with dimension that, for realistic datasets, it becomes practically unattainable. The consequence, in our view, is that for any real, finite dataset that can be stored on current servers (however massive they may be, such as the petabytes used to train GPT-4), the probability that a newly generated project will fall within the strict convex hull of the training data becomes extremely small in high-dimensional settings. In that sense, the interior coverage of the space is sparse, and new samples are likely to fall outside the hull. The question we can nevertheless ask is this: why do we get this feeling of constant resemblance when we ask generative AI to design this

⁶¹ “As such, in the presence of a nonlinear manifold, we can see that the exponential requirement from Thm. 1 [theorem 1] in the number of samples required to preserve a constant probability to be in interpolation grows exponentially with d .” Ibid., p. 2.

⁶² Cf. W. V. O. Quine, *Word and Object*, MIT Press, Cambridge, Mass., 1960.

or that object (object in the generic sense of “output”)? In our view, and limiting ourselves to architecture, it stems primarily from the fact that the essential characteristics of all architectural projects, quite simply *of all architectures ever produced*, built or not, are *very limited in number*. This property is precisely what enables the dimensionality reduction widely used in AI models during training, and which allows for relatively densely populated latent spaces within their convex hull. Indeed, the characteristics common to two houses, as well as to two cats or two pairs of shoes, are far more numerous than those that differentiate them.

The Limits of Latent Space

The significance of everything said thus far aligns with Mr. Carpo’s assertion that current generative AI reminds us of the “universal condition of human creativity: namely, that nothing is created out of nothing, but all is generated out of something that already exists.” Humans, in fact, most often create imitations augmented with minor changes, which explains the appallingly slow pace of architectural evolution and the recurring similarities, across one or even several centuries, between buildings—cathedrals, palaces, towers—that differ only in a few details of no importance except for their hermeneutic value. The only truly significant changes occur when new parameters external to architecture come into play. For example, when new construction techniques or new materials (steel, reinforced concrete) are sufficiently mature to enter the field of architecture, or when a few brilliant individuals effect a shift large enough to be considered “out-of-distribution” (OOD). While the theory of creative imitation mentioned by Mr. Carpo seems to be valid for architecture and even beyond this discipline, in our view, it is only valid *on a human level*. While creative imitation applies without difficulty to all kinds of past architectures, we believe—beyond the obvious logical impossibility of theorizing what does not yet exist—that the situation is very different with *artificial architecture*. By limiting ourselves to analysing what already exists, including generative AIs themselves, we believe, on the other hand, that Mr. Carpo is wrong when he asserts that “*generative AI can only invent new things based on—but in fact derived from—a dataset of precedents.*”⁶³ [emphasis added]. While the apparent

⁶³ “[...] Generative AI, either based on legacy GANs or on more recent large language models, is, in epistemological terms, a machine that automates imitation; it is a powerful—albeit purely technical—reminder of a basic, universal condition of human creativity: namely,

functioning of current generative AI suggests they operate through creative imitation, aptly recalling a distinctly human mechanism of creativity, attributing this same mechanism to the quasi-essence of current connectionist AIs does not withstand scrutiny. Indeed: 1) we saw earlier that datasets will never be large enough to guarantee a kind of imitative derivation characteristic of interpolation; 2) it is not the models themselves but the data on which these models are trained that exhibits similarities and statistical regularities, largely because these are the product of either human cultural history (effectively based on creative imitation) or nature, both of which are remarkably homogeneous, continuous, and even “smooth;” 3) in high dimensions, in almost all cases, generative AI does not perform interpolations but extrapolations, therefore, the quality of the results or what is more precisely called the ability to generalise, does not come from interpolations that can be likened to “creative imitations,” but from other factors.

A common objection to the view that generative AI *is not* the stochastic parrot it is wrongly labelled as, is that architectural data are not random (unlike the uniform points of the theorem invoked by Balestriero *et al.*). They are based on an underlying structure, a lower-dimensional manifold. The intelligence of architecture would then consist precisely in finding this hidden structure, made up of a sum of the characteristics common to most architectures ever produced and those specific to each one. We mentioned earlier that the very limited number of these characteristics allows for dimensionality reduction. By doing so, we could operate in a compressed “latent space” in which interpolation (and therefore imitation) would once again become possible. However, Balestriero *et al.* strongly challenge this hypothesis, both theoretically and empirically. They argue that reducing intrinsic dimensionality does not, by itself, restore interpolation in the strict convex-hull sense.⁶⁴ What matters for the probability of interpolation is not the dimension of the manifold (the actual complexity of the object, for example, the “dimensionality of style”), but the dimension of the smallest affine subspace containing the

that nothing is created out of nothing, but all is generated out of something that already exists. Just like Generative AI can only invent new things based on—but in fact derived from—a dataset of precedents [...].” M. Carpo, “Generative AI, Imitation Learning, and the Automation of Tacit Knowledge,” p. 106.

⁶⁴ See the section “The Role of the Intrinsic, Ambient and Convex Hull Dimensions” in R. Balestriero, J. Pesenti, Y. LeCun, “Learning in High Dimension Always Amounts to Extrapolation,” pp. 2–3. <https://doi.org/10.48550/arXiv.2110.09485>, (accessed 8 April 2026).

data (the convex hull). Their experiments on real-world image databases (MNIST, CIFAR-10, ImageNet) show that the proportion of test samples in the interpolation regime drops to zero as soon as more than 30 or 40 dimensions are considered. Even when projecting the data into latent spaces optimised by deep networks, extrapolation remains the rule.⁶⁵ We can therefore, both technically and conceptually, oppose Mario Carpo's position by arguing that contemporary AI systems cannot straightforwardly be described as machines of imitation if imitation is defined in the strict geometric sense of interpolation. The mere fact that they operate in high-dimensional spaces constrains them to move beyond the envelope of precedents. While the argument that everything is generated from something that already exists remains true at the dataset level, it becomes false at the generation topology level. One cannot describe generation solely as navigation within existing knowledge; a substantial part of it must also be understood as extending beyond previously sampled regions of the space.

In fact, we believe that AI's capacity to generate outputs that are not reducible to straightforward recombinations may be greater than many of its critics allow, sometimes through the "stochastic parrot" label, a term that might be more fitting for humans. Indeed, in most cases, we think, conceive, and act in surprisingly similar ways within frameworks of thought that we believe we are challenging, even though, in most cases, we have not even managed to identify those frameworks; a behaviour that Alfred North Whitehead identified very well.

When you are criticizing the philosophy of an epoch, do not chiefly direct your attention to those intellectual positions which its exponents feel it necessary explicitly to defend. There will be some fundamental assumptions which adherents to all the variant systems within the epoch unconsciously presuppose. Such assumptions appear so obvious that people do not know what they are assuming because no other way of putting things has ever occurred to them. With these assumptions a certain limited number of types of philosophic systems are possible, and this group of systems constitutes the philosophy of the epoch.⁶⁶

⁶⁵ *Ibid.* Table 1 shows that (in a latent space of dimension 512), proportion of the test set that falls into the interpolation regime for datasets of dimensions >30 can be as small as 0 per cent.

⁶⁶ A. N. Whitehead, *Science and the Modern World*, Cambridge University Press, London, 1925, p. 61.

One advantage of rarely changing our thought patterns is that it preserves the historical and cultural continuity to which humans are accustomed. Humans experience immense difficulty—being biologically “wired” in a certain way—thinking, conceiving, and acting in radically different ways, ways that would differ between individuals and differ from each person’s own prior ways of thinking. The human world, like the natural world of which we are a part and to which we are accustomed, is made up of slow evolution through limited differentiations in which, to borrow A. N. Whitehead’s words, we believe that only certain systems of thought are possible. Applied to the potential hierarchy between human and machine intelligence, this observation leads us to recognise that it is primarily because no different organisation of things has yet emerged—an organisation in which machine intelligence is, *for everyone, resolutely and visibly superior to human intelligence*—that the still-dominant system of thought presupposes the superiority of human intelligence. Although it seems natural, we believe this assumption is false, for at least three reasons: 1) numerous recent experiments suggest that AI systems can exceed human performance on a growing number of bounded benchmarks and specialised tasks. 2) Combining all specialised AIs and their extraordinary ability to exchange information at the speed of light leads to performances comparable to those of AGI, keeping in mind that there is no proof of a settled transition to AGI and that AGI itself remains ill-defined. 3) Drawing definitive conclusions about inherent limitations to machine intelligence on the basis of current performances is no less intellectually risky than asserting the opposite—namely, the necessary emergence of future superhuman capabilities. Yet opponents of what are commonly (and somewhat simplistically) labelled “strong AI” theses commit, in most cases, the very same fallacy of reasoning of which they accuse their adversaries. Having recalled these points, we will not elaborate on them further, as they have already been discussed in detail by leading authors.⁶⁷

Extrapolation and Autonomous Creativity

The geometric reversal just described leads us to redefine what we mean by creativity and autonomy. If imitation is an impossible interpolation, then AI is by default an extrapolating machine. Extrapolating often

⁶⁷ See, in particular, N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, Oxford, 2014.

means, in everyday language, making a risky assumption from insufficient data, but in the context of high dimensions, extrapolation becomes the very condition for the production of form. The word “form” is understood here either in the strict sense or in the generic sense of a result that can be mathematically formalised. Since the training data (the architectural past) is scattered throughout the vastness of the space of possibilities, any response provided by a generative AI—be it a Midjourney image, a ChatGPT text, or any other mode—appears as a projection into the void, beyond the convex hull of human knowledge, though generation remains guided by learned regularities, objective functions, and sampling procedures. LLMs, diffusion models, and other generative AI systems do not simply repeat the past but are rather engines of exploration that, beyond the apparent information overload, build precarious bridges over the “information void.” If human creativity is defined not as creation *ex nihilo* (which no one supports) but as a structured generalisation beyond known experience, then AI is also creative insofar as it extrapolates. It generates responses to situations (most commonly prompts) it has never seen, located in regions of mathematical space that have never been mapped by training data. The prompts may, of course, already be seen if prompt generation is part of the training, or simply because several users ask the same questions in the same terms (which is common). AI is not creative despite being derived from data, but because the geometry of its space forces it to invent in order to connect disparate points in its memory. Since interpolation is forbidden by the curse of dimensionality, innovation (extrapolation) becomes a statistical necessity. It is autonomy through geometric constraint.

This perspective allows us to return to the surrealist intuition evoked in the call for contributions: “Surrealist automatism, for instance, aimed to bypass rational constraints and tap into unconscious creativity. There, technical systems become instruments of deviation, surprise and poetic rupture.”⁶⁸ Mathematically, this “deviation” is extrapolation. When the algorithm generates a shape that is not within the convex hull, it “hallucinates” a reality. The paradigm of classical engineering sees hallucination as a malfunction or an error to be corrected. In high-dimensional extrapolation, hallucination is the default mode of existence for the digital object. To generate, for example, a coherent image (or other digital file) of

⁶⁸ *Khōrein: Journal for Architecture and Philosophy*, “Call for Papers: Volume IV, Issue 3: Automata.”

a building that does not exist, AI must “hallucinate” plausible details, textures, and structures—consistent with the learned distribution—but not present in any specific example. The automaton no longer calculates the known; it hallucinates the possible. In less metaphorical terms, the automaton no longer operates only as a calculator of predefined solutions but as a generator of candidate outputs under uncertainty, often beyond any exact precedent, while remaining statistically constrained by what it has learned. The guided randomness postulated by Dartmouth researchers in 1955 as a condition of creativity is therefore indeed present in generative AI; but it is anything but a simple noise term added to simulate variety: it is a vectorial component that allows movement in the empty areas of latent space without collapsing.

Although we disagree with M. Carpo that, 1) “generative AI is an imitation machine;” 2) that we “need to learn again at least the basic [sic] of the arts and sciences of imitation, so we know what Generative AI does—and what we do, when we use it;”⁶⁹ and 3) that we ultimately need “to come up with a new theory of mimesis,” assuming that this new theory were proposed, we should first emphasise that “mimesis in the age of AI” is not a reproduction of models (imitation/interpolation) but an extrapolative generation from scattered precedents. It is not a derivation along known paths but a navigation through a topology barely sketched by previous models. Since the very term *mimesis* is, in fact, inappropriate, it seems preferable not only to avoid it but, even more so, to escape entirely from the design logics it encompasses, as Peter Eisenman already urged us to do 40 years ago.

Rereading Eisenman

On the occasion of a Mies van der Rohe exhibition in Chicago,⁷⁰ Eisenman revisited the importance of escaping mimesis and reading architecture as a text: “Text emerges in Mies’s architecture when the symbolic

⁶⁹ “And, in so far as Generative AI is an imitation machine, we urgently need to learn again at least the basic [sic] of the arts and sciences of imitation, so we know what Generative AI does—and what we do, when we use it. In short, we need to engage with imitation again, in critical and creative terms; after a century of abeyance and modernist mandated oblivion, we need to come up with a new theory of mimesis.” M. Carpo, “Generative AI, Imitation Learning, and the Automation of Tacit Knowledge,” p. 106.

⁷⁰ Exhibition *Mies van der Rohe and His Disciples* (director: John Zukowsky), organised at the Art Institute of Chicago in 1986 and presented from 1 April to 15 June 1987 in the exhibition hall of the Information Center at the Centre national d’art et de culture Georges Pompidou, Paris.

connection between man and object (and hence symbol and object) loses its relevance and thus can be taken apart. Instead of simulating the vertebrate structure of man, Mies's architecture simulates a textual structure. The separation of an architecture from the mimesis of man is the production of a simulation; this is the nature of a text."⁷¹ By emphasizing the text, or more precisely the text as a set of symbols dissociated from traditional architectural symbolism—that is, from representation—Eisenman severs architecture from the primacy of meaning and the anthropomorphism it embodies. While classical text and its alphabet—a finite alphabet—aim to construct and transmit human thoughts, Eisenman's text does not aim to communicate meaning but simply “to be.” In doing so, it is comparable to 1) what a text is for the information theory we have discussed: a message whose meaning is irrelevant and for which we can compute an amount of information; 2) the purely formal text that culminates in Hilbert's axiomatic project.

[For Hilbert] Certainty will come from the formal manipulation of the graphic characters that compose statements about infinity and space, and not from space itself or from geometric forms within space. To achieve this, it is necessary to return to the very theory of meaning by questioning the idea that spacetime is the locus of its elaboration. It then becomes a matter of reducing mathematical statements to their written characters, which are always finite in nature, and of setting aside their potential meaning, which can encompass infinity and its dangers [...]. [For Hilbert] there is only formal calculus, the locus of certainty, because it is, precisely, “potentially mechanizable.” Certainty lies in the mechanical, far removed from the ambiguities of meaning, which are human, all too human.⁷²

Very early on—as early as his 1963 Cambridge doctoral thesis, “The Formal Basis of Modern Architecture”—Eisenman perceived the profound conceptual shifts brought about by twentieth-century science and with which architecture would be confronted. This led him, beyond

⁷¹ P. Eisenman, “miMISes READING: it does not mean A THING,” in *Eisenman Inside Out: Selected Writings, 1963–1988*, Yale University Press, New Haven, 2004, p. 191. The text was originally published for the exhibition *Mies van der Rohe and His Disciples* (ed. J. Zukowsky) held at the Art Institute of Chicago in 1986; see note 70 above.

⁷² J. Lassègue, G. Longo, *L'empire numérique*, Humensis, Paris, 2024, pp. 124–125. Translation modified.

the succession of styles characteristic of the five centuries separating the birth of modern science from its “closure,” to seek the common characteristics of the different architectural styles that had succeeded one another. In 1984, he identified three such characteristics: Representation, Reason, and History, which he called fictions and grouped under a single term: “the classical.”

Architecture from the fifteenth century to the present has been under the influence of three “fictions.” Notwithstanding the apparent succession of architectural styles, each with its own label—classicism, neoclassicism, romanticism, modernism, postmodernism, and so on into the future—these three fictions have persisted in one form or another for five hundred years. They are representation, reason, and history. Each of the fictions had an underlying purpose: representation was to embody the idea of meaning; reason was to codify the idea of truth; history was to recover the idea of the timeless from the idea of change. Because of the persistence of these categories, it will be necessary to consider this period as manifesting a continuity in architectural thought. This continuous mode of thought can be referred to as the classical.⁷³

While Eisenman, in our view, describes with great accuracy the persistence of these anthropic fictions which he associates with Meaning, Truth and Time, he does not, however, go so far as to postulate the possibility of an artificial architecture completely external to the human figure of an author-reader:

[...] a “not-classical” architecture begins actively to involve an idea of a reader conscious of his own identity as a reader rather than as a user or observer. It proposes a new reader distanced from any external value system (particularly an architectural-historical system). Such a reader brings no a priori competence to the act of reading other than an identity as a reader. That is, such a reader has no preconceived knowledge of what architecture should be (in terms of its proportions, textures, scale, and the like); nor does a “not-classical” architecture aspire to make itself understandable through these preconceptions.⁷⁴

⁷³ P. Eisenman, “The End of the Classical: The End of the Beginning, the End of the End,” *Perspecta*, 21, 1984, pp. 155–173.

⁷⁴ *Ibid.*, p. 171.

The new “author-reader” must, as far as possible, discard any pre-conceptions about what architecture should be, but by remaining a human author-reader, they can only operate within their own biological limitations, which are, moreover, the very reason for the existence and persistence of the fictions identified by Eisenman. The fact that the new (architectural) “object” the American architect calls for “must have the capacity to reveal itself first of all as a text, as a reading event,” and that the whole requires “an other ‘timeless’ space of invention,” [...] “a ‘timeless’ space in the present, without a determining relation to an ideal future or to an idealised past,”⁷⁵ does not change the question. Eisenman’s new *author-reader*, *text-object*, and *timeless space* remain humanly conceived, so that the design of (architectural) *text-objects by author-readers* will, in any case, only be realised within a space of human thought—that is to say, ultimately, the very space of classical thought that Eisenman aimed to transcend. Whatever deconstruction project is put forward—and it seems to me that we now have evidence of its limitations, even its absurdity in some extreme contexts—what endures and what Eisenman fails to dismiss with his critique of the classical is, first and foremost, the *primacy of human intelligence*. The supposed superiority of human intelligence over all others—for example, over that of animals or, as far as we are concerned here, over machine intelligence—thus manifests itself, albeit for a very short time, as the true “metafiction” that makes the former possible and even inevitable. Eisenman rightly judges that “[...] the period of time dominated by the classical [the last five centuries] can be seen as an ‘episteme’, to employ Foucault’s term—a continuous period of knowledge that includes the early twentieth century,”⁷⁶ but he fails to theorise the profoundly mechanistic nature of the ongoing epistemological bifurcation. Moreover, this bifurcation, materialised by artificial intelligence whose success has been actively prepared by modern science, transcends epistemology. We are no longer faced with different spheres of knowledge within a single human space, but rather with different spaces of knowledge operating side by side. While these spaces may continue to interact for some time, there is no doubt that the knowledge space associated with AI is already expanding faster than individual human comprehension can track, and in some domains its internal procedures remain only partially

⁷⁵ *Ibid.*, p. 172.

⁷⁶ *Ibid.*, p. 155.

interpretable.⁷⁷ It is therefore plausible that highly capable systems will become increasingly difficult for unaided humans to understand in detail, even when their outputs remain useful and empirically successful. While we may acknowledge undeniable results, the modes of thought, the reasoning, —in short, the entire “logic” inherent in these results computed within new kinds of knowledge spaces—will elude us.

P. Eisenman’s author, as previously noted, is first and foremost a *reader*, hence the author-reader expression we used. In his human dimension, he remains the very embodiment of the humanism that underpins “the classical.” Like a Renaissance scholar, he only “writes” after extensive reading, and creates through creative imitation. At the end of his life, his output, however vast, represents only a fraction of his reading—human writing abilities being far inferior to reading abilities, especially if we take the term “reading” in its broadest sense of understanding or perceiving the world. For all the architecture that precedes the “computational turn”—or, if we prefer to refer to a computational paradigm precisely at work in recent AI models, the “connectionist turn”—“the primary function of alphabetic writing, which is to reproduce human speech,”⁷⁸ is preserved. The Eisenmanian author-reader can, alternatively, be either author or reader insofar as these roles coexist within a resolutely organic unity. If, as Turing asserted, “Mechanism and writing are [from his point of view] almost synonymous,”⁷⁹ it is, as Jean Lassègue and Giuseppe Longo point out, “the ‘almost’ that should hold our attention [...]” If, according to them, this “almost” “contains in itself the entire philosophical question of the limitation of formalisms,”⁸⁰ it also symmetrically contains the entire superiority of the mechanical, which can exploit (binary) coding to produce knowledge on a scale unattainable by any human means whatsoever. By its very existence “Coding completely breaks with the primary function of alphabetic writing, which is to reproduce human speech.”⁸¹ Moreover,

⁷⁷ This autonomisation of knowledge may or may not coincide with the Singularity, and may or may not belong to its very definition. My own view is that the autonomisation of knowledge is already deeply ingrained in our civilisation and that, once complete, the very word knowledge—still deeply associated with human knowledge—will become irrelevant.

⁷⁸ J. Lassègue, G. Longo, *L’empire numérique*, p. 150.

⁷⁹ A. M. Turing, “Computing Machinery and Intelligence”; cited in J. Lassègue, G. Longo, *L’empire numérique*, p. 134.

⁸⁰ *Ibid.*, p. 134.

⁸¹ *Ibid.*, pp. 132–133.

The emergence of coding in the 20th century in the form of programming languages is innovative in that it aims to automate the writing of formal deduction, and no longer just the reading as was the case with the Greek alphabet, and this is a major achievement in the history of writing. The concatenation of characters then takes place according to logical rules that require no interpretation to be applied and can therefore be executed automatically without recourse to phonemes and the voice through which the meaning of alphabetic writing is reconstructed. Every computer programme is a formal deduction that executes without relation to the voice of a speaking body.⁸²

Lassègue and Longo note, “from geometry to the living body, by way of speech as the collective guarantee of meaning, the entire phenomenology specific to the relationship between humans and living beings and the world is devalued and evacuated, encoded in sequences of 0s and 1s [...]”⁸³. However, by only partially evacuating phenomenology in order to preserve a human form of thought—thought whose no less human “reading” is the sign—P. Eisenman, and even more so any reading of architecture “in the Eisenman style,” fails to identify the full radicality of the computational. The computational dimension certainly does not appear as directly in connectionist AI as it did in first-generation AI, but it has not disappeared. The meaning that seems to emerge from neural networks as if by magic has not, in fact, been reintroduced “in the old way.” During the learning phases, no semantic network or definition is provided for each term encountered. The performance of deep neural networks does not stem from having discarded formalisation or computation to reintroduce meaning. On the contrary, it stems from the fact that we have not hesitated to discard meaning, perhaps even more radically than in symbolic AI, by exploiting the geometry of vectorised data spaces in an extraordinarily intelligent way—for example, words transformed into tokens, which are themselves transformed into vectors carrying complex context information (the embedding).

Singularity of the Work and Convex Position

In current generative AI, architecture no longer exists strictly as an Eisenman-like text, but as a set of (digital) vectors in a high-dimensional space.

⁸² *Ibid.*

⁸³ *Ibid.*, p. 127.

This specificity should lead us to study the properties of these spaces in an attempt to understand how meaning emerges and what we readily call creativity; meaning being obviously inseparable from all creativity, which, without it, would be nothing more than the mere production of random objects. To better understand why we maintain that the emergence of meaning/creativity is not reducible to the creative imitation discussed earlier, but is always, in one way or another, a production of radical novelty, we can invoke the “Convex Position Problem” (defined by P. Valtr, 1995) cited by Balestrieri *et al.* Before doing so, however, let us clarify that by radical novelty we are not referring to what appears new to us humans, but to this question from the perspective of a generative AI. Statistically, a simple answer drawn at random from a massive dataset would very likely seem new to us. It would almost certainly be more likely to seem new than a sensible answer to the original question, but it is precisely because it would also have a very high chance of being nonsensical that it would be an inadequate answer. However, to an outside observer, there would be no easily identifiable difference between this same (bad) answer drawn at random and that provided by an AI only capable of hallucinations.

Returning to the Convex Position Problem, it calculates the probability that points are in convex position, meaning that all points lie outside the convex hull formed by the others. The theorem cited by Balestrieri *et al.* (based on the work of C. Buchta, 1986) shows that in a high-dimensional hypersphere, as the number of samples grows linearly with the dimension, this probability tends towards 1. This means that each point is an extreme vertex, an exception that defines the boundary of the hull. If we interpret this result through the lens of the question “what happens inside current generative AI models when we query them?” (for example, by asking them to design a specific architecture), the answer would be that, contrary to intuition, far from producing a mass of conformist copies—as critics of automation fear—AI tends to produce objects that are all, mathematically, exceptions to the training corpus. In fact, each project generated by an AI appears as a singularity that defines its own territory, refusing to be the average of others, without (we agree on this) this singularity necessarily being equivalent to what a human would deem relevant or of high quality. What matters here is that with current generative AI, artificial architecture shifts from a logic of parametric variation—which consists of deforming a known form while remaining within certain bounds—to a logic of stochastic

exception—which amounts to producing an out-of-the-ordinary form through extrapolation. If we take current generative AIs as the state of the art and the models of the contemporary automaton, then, as the call for papers for the special issue of *Khōrein* suggests,⁸⁴ the latter functions as an “enabling partner,” “unknowable,” and “unruly,” because it operates in dimensions where the order of imitation is mathematically untenable.

CONCLUSION: AUTONOMY OF AUTOMATA AND THE END OF IMITATION

This essay has traced a theoretical trajectory from the quantification of information to the autonomy of contemporary architectural automata. Claude Shannon’s information theory neutralised meaning by reducing communication to probabilistic selection, thereby rendering cultural artifacts computable. Alan Turing’s theory of computation subsequently neutralised inspiration by demonstrating that any effective procedure could be executed by a universal machine. Together, these developments desacralised intelligence while simultaneously making it accessible to technical systems. A third and more recent transformation emerges with deep learning and high-dimensional geometry. Here, repetition itself is neutralised. As the work of Balestrieri, Pesenti, and LeCun shows, machine intelligence cannot be constructed through the accumulation and interpolation of past examples. The curse of dimensionality acts as a paradoxical constraint: mathematically it renders imitation extremely rare and forces extrapolation as the default mode of generation.

The contemporary architectural automaton must therefore be understood neither as a mechanical executor of predefined rules nor as a stochastic imitator of historical precedents. It functions instead as a constrained geometric explorer, navigating spaces so vast and sparse that novelty becomes an almost structural necessity rather than an intentional choice. Autonomy does not stem from consciousness, intention, or will, but from the necessity of bridging gaps between scattered data points. Architectural intelligence emerges from the mathematical topology of the spaces in which the machine operates, as it fills the immense void of latent space through extrapolation. The intelligence of the automaton is thus neither human nor post-human; it is geometric. Condemned to operate where imitation fails, it is structurally driven to invent.

⁸⁴ *Khōrein: Journal for Architecture and Philosophy*, “Call for Papers: Volume IV, Issue 3: Automata.”

REFERENCES

- Arnheim, Rudolf (1954), *Art and Visual Perception: A Psychology of the Creative Eye*, Berkeley and Los Angeles: University of California Press.
- Aureli, Pier Vittorio (2023), *Architecture and Abstraction*, Cambridge, Mass.: MIT Press.
- Balestriero, Randall, Jérôme Pesenti, and Yann LeCun (2021), "Learning in High Dimension Always Amounts to Extrapolation," *arXiv*, <https://doi.org/10.48550/arXiv.2110.09485> (accessed April 8, 2026).
- Birkhoff, George D. (1933), *Aesthetic Measure*, Cambridge, Mass.: Harvard University Press.
- Boden, Margaret (1991), *The Creative Mind: Myths and Mechanisms*, New York: Basic Books.
- Bostrom, Nick (2014), *Superintelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press.
- Carmo, Mario (2011), *The Alphabet and the Algorithm*, Cambridge, Mass.: MIT Press.
- Carmo, Mario (2024), "Generative AI, Imitation Learning, and the Automation of Tacit Knowledge," *A+U*, 646, pp. 104–109.
- Carmo, Mario (ed.) (2012), *The Digital Turn in Architecture 1992–2012*, Chichester: Wiley.
- Chatterji, Niladri S., Philip M. Long, Peter L. Bartlett (2021), "When Does Gradient Descent with Logistic Loss Find Interpolating Two-Layer Networks?," *Journal of Machine Learning Research*, 22, 159, pp. 1–48.
- Chomsky, Noam, Ian Roberts, Jeffrey Watumull (2023), "The False Promise of ChatGPT," *The New York Times*, March 8, <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html> (accessed April 8, 2026).
- Copeland, B. Jack (2004), "The Church-Turing Thesis," *NeuroQuantology*, 2, 2, pp. 101–115.
- Craik, K. J. W. (1943), *The Nature of Explanation*, Cambridge: Cambridge University Press.
- Dieudonné, Jean (1964), *Algèbre linéaire et géométrie élémentaire*, Paris: Hermann.
- Eco, Umberto (1976), *A Theory of Semiotics*, Bloomington/London: Indiana University Press.
- Eisenman, Peter (2004), "miMISes READING: It Does Not Mean a Thing," in *Eisenman Inside Out: Selected Writings, 1963–1988*, New Haven: Yale University Press, pp. 189–201.
- Eisenman, Peter (1984), "The End of the Classical: The End of the Beginning, the End of the End," *Perspecta*, 21, pp. 155–173.
- Franklin, James (2011), "Science Wars: Last Round?," *The New Criterion*, 29, 9, p. 68.
- Frazer, John (1995), *An Evolutionary Architecture*, London: Architectural Association.
- Gero, John S. (ed.) (1985), *Design Optimization*, Orlando: Academic Press.
- Goldstine, Herman H. (1980), *The Computer from Pascal to von Neumann*, Princeton: Princeton University Press.

- Hardingham, Samantha (ed.) (2016), *Cedric Price Works 1952–2003: A Forward-Minded Retrospective*, 2 vols., London and Montreal: Architectural Association and Canadian Centre for Architecture.
- Hartley, Ralph V. L. (1928), “Transmission of Information,” *The Bell System Technical Journal*, 7, pp. 535–563.
- Khôrein: Journal for Architecture and Philosophy* (2026), “Call for Papers: Volume IV, Issue 3: Automata,” <https://khorein.ifdt.bg.ac.rs/index.php/ch/announcement/view/22> (accessed April 8, 2026).
- Kleene, Stephen C. (1951), “Representation of Events in Nerve Nets and Finite Automata,” *RAND Research Memorandum RM-704*, Santa Monica: RAND Corporation.
- Kotsopoulos, Sotirios (ed.) (2025), *Shape Computation: Fifty Years, 1972–2022*, Cham: Birkhäuser.
- Lassègue, Jean, Giuseppe Longo (2024), *L’empire numérique*, Paris: Humensis.
- Lynn, Greg (1994), “New Variations on the Rowe Complex,” *ANY: Architecture*, New York, 7–8, pp. 38–43.
- McCarthy, John (1951), *Projection Operators and Partial Differential Equations*, PhD dissertation, Princeton University.
- McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, Claude E. Shannon (2006), “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” *AI Magazine*, 27, 4, pp. 12–14.
- McCulloch, Warren S., Walter Pitts (1943), “A Logical Calculus of the Ideas Immanent in Nervous Activity,” *Bulletin of Mathematical Biophysics*, 5, pp. 115–133.
- McCulloch, Warren S., Walter Pitts (1947), “How We Know Universals: The Perception of Auditory and Visual Forms,” *Bulletin of Mathematical Biophysics*, 9, pp. 127–147.
- Minsky, Marvin L. (1954), *Theory of Neural-Analog Reinforcement Systems and Its Application to the Brain-Model*, PhD dissertation, Princeton University.
- Moles, Abraham A. (1972), *Information Theory and Aesthetic Perception*, Paris: Denoël.
- Negroponte, Nicholas (1970), *The Architecture Machine: Toward a More Human Environment*, Cambridge, Mass.: MIT Press.
- Negroponte, Nicholas (1975), *Soft Architecture Machines*, Cambridge, Mass.: MIT Press.
- Nyquist, Harry (1924), “Certain Factors Affecting Telegraph Speed,” *The Bell System Technical Journal*, 3, 2, pp. 324–346.
- Nyquist, Harry (1928), “Certain Topics in Telegraph Transmission Theory,” *Transactions of the American Institute of Electrical Engineers*, 47, 2, pp. 617–644.
- Oosterhuis, Kas (2002), *Programmable Architecture*, Milan: L’Arca Edizioni.
- Oosterhuis, Kas (2003), *Hyperbodies: Towards an E-motive Architecture*, Basel: Birkhäuser.
- Quine, W. V. O. (1960), *Word and Object*, Cambridge, Mass.: MIT Press.
- Rondelet, Jean-Baptiste (1812–1817), *Traité théorique et pratique de l’art de bâtir*, Paris: Published by the author.

- Shannon, Claude E. (1938), "A Symbolic Analysis of Relay and Switching Circuits," *Transactions of the American Institute of Electrical Engineers*, 57, pp. 713–723.
- Shannon, Claude E. (1948), "A Mathematical Theory of Communication," *The Bell System Technical Journal*, 27, pp. 379–423, 623–656.
- Spuybroek, Lars (2009), *Research & Design: The Architecture of Variation*, London: Thames & Hudson.
- Stiny, George, James Gips (1972), "Shape Grammars and the Generative Specification of Painting and Sculpture," in C. V. Freiman (ed.), *Information Processing 71*, Amsterdam: North-Holland, pp. 1460–1465.
- Stiny, George, James Gips (1978), *Algorithmic Aesthetics: Computer Models for Criticism and Design in the Arts*, Berkeley: University of California Press.
- Szambien, Werner (1984), *Jean-Nicolas-Louis Durand 1760–1834: De l'imitation à la norme*, Paris: Picard.
- Turing, Alan M. (1937), "On Computable Numbers, with an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society*, 2, 42, pp. 230–265.
- Turing, Alan M. (1950), "Computing Machinery and Intelligence," *Mind*, 59, 236, pp. 433–460.
- Turing, Alan M. (2004), "Intelligent Machinery," in B. Jack Copeland (ed.), *The Essential Turing*, Oxford: Clarendon Press, pp. 410–432.
- von Neumann, John (1945), *First Draft of a Report on the EDVAC*, Philadelphia: University of Pennsylvania.
- von Neumann, John (1948), *Theory and Techniques for Design of Electronic Digital Computers*, Philadelphia: University of Pennsylvania.
- Whitehead, Alfred North (1925), *Science and the Modern World*, London: Cambridge University Press.
- Wiener, Norbert (1948), *Cybernetics: Or Control and Communication in the Animal and the Machine*, Paris: Hermann.

Klaus Wiegerling*

THE SMART CITY CONCEPT AND ITS LIMITS: RESISTANCE, TECHNOLOGY, AND THE FUTURE OF THE CITY

ABSTRACT: This article focuses on the relationship between the smart city concept and the idea of autonomous technology. The problem of resistiveness in advanced Information and Communication Technologies that claim to be “autonomous” is identified as a transcendental philosophical concept that marks conditions of possibility and limits the idea of a smart city and autonomous technology. Furthermore, the article shows that information technology access to reality only achieves the typological, but not the eventful, singular, and concrete quality of experience that is subject to the principle of individuation. Thus, the concept of the Smart City is also overloaded with exaggerated hopes.

KEYWORDS: smart home, political participation, resistiveness, resistance, sustainability

* Klaus Wiegerling: Karlsruhe Institute of Technology, ITAS/University of Kaiserslautern; wiegerlingklaus@aol.com.

This is an Open Access article under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not transformed in any way and is properly cited.

I. WHAT DOES SMART CITY MEAN?

The idea of the Smart City began to gain traction around the turn of the current century. There have been a number of projects in Europe—often funded by the EU—that address issues subsumed under the concept of “Smart City.” The term is somewhat based on the idea of the “Smart Home” and expands it not only to a presumed heterogeneous urban citizenry but also to non-informatic topics such as climate protection and sustainability, health, mobility, new forms of political participation, etc. The concept is also extended to the surrounding areas. Within the framework of Ubiquitous Computing, “smart home” was practically a fixed component of the program.¹ One’s own four walls should be as comfortable as possible, appropriate to our wishes, and our individual behavior should be socially acceptable. In the Smart City concept, not only the individual is supposed to receive support but the entire city, indeed the whole region in which it is located should as well. Strictly speaking, the associated centralization concept is incompatible with the decentralized principle of the network. In a way, this acknowledges that people seek their happiness not alone, but in community; that their entertainment and private needs should also lie where they work. Urban citizens should also be interested in acceptance and tolerance. It is about accepting individual needs within a certain framework and integrating them into a larger concept of the city, assuming that the interests of people, despite all divergences, are similar. On the one hand, the individual need, the diversity of people, etc., is emphasized, while on the other hand, a type of person is also highlighted who is tolerant and willing to accept and is ready to put up with many things for the sake of efficiency and the acceleration of life. Today, we see that this changes once a certain number of similarly disposed individuals is reached. Studies have shown that people with different value systems no longer strive for adaptation as soon as they can live together with a like-minded larger group in a district and where they make up the majority. Smart City is an attempt to gain control over all areas of life, thus coordinating them and creating a “Smart Citizen” who perceives themselves as part of the informational organizational, informational, and communication system. Specifically, an attempt should also

¹ J. Heesen, “Analyse zu Szenarien einer intelligenten Wohnumgebung,” in J. Heesen, C. Hubig, O. Siemoneit, K. Wieglerling (eds.), *Leben in einer vernetzten und informatisierten Welt. Context-Awareness im Schnittfeld von Mobile und Ubiquitous Computing*, Nexus Schriftenreihe, SFB 627 Bericht Nr. 2005/05, Stuttgart 2005, pp. 149 f.

be made to reduce energy-intensive living, for instance through a “sharing economy.” The “Internet of Things” should be integrated into urban networking and supplemented or expanded by urban and private services that are characterized by their proximity to housing. New ways of digital civic participation in political decisions should be developed. The information techniques should be brought up to date, and cloud technology and AI-supported dialogue systems should be made available. From these applications, one expects efficiency improvements and cost savings. The basis for informatics operations is still human capital in terms of individual knowledge. However, the question is whether this will still be the case with the use of modern dialogue systems, as these primarily generate and process knowledge from texts. This means the system is increasingly distancing itself from concrete authorship, providing indirect knowledge, which does not necessarily have to be worse. So-called human capital will likely shrink in favor of the general availability of knowledge of largely independently operating machines.

Further examination of the Smart City concept is a critical one. It is not a question whether the concept actually brings benefits, but that is not the focus when discussing its limits. Theoretical architectural topics remain largely excluded. While sociological and social psychological aspects still play a role, questions of aesthetics, for example, are set aside. Nevertheless, the question arises whether the digitization of the city does not also impose other requirements. Can the currently developed Smart City concepts truly be realized without changes in urban planning? Must traffic planning change? Must recreational zones be distributed differently? Can urban supply remain as it has been? All these questions are subordinated to the discussion of the technical foundations of the concept here.

The idea of smartness is situated between utopia and reality. One imagines an optimally functioning city-state, analogous to Plato’s Republic, and attempts to design it to be as self-sufficient as possible. The idea of the Smart City attempts to break down the ideal of a digitized society into what is practically achievable in a non-digital context. There are people living in the city who need to eat, work, stay mobile and healthy, and these individuals should be informally supported and relieved. The existing urban society is being optimized with digital assistance to bring improvements for everyone. Whether consciously or unconsciously, the ancient Greek city-state likely serves as a model, which, however, fundamentally contradicts the idea of decentralization promised by a connected and global digital society. Since the city must be supplied, and

also requires an environment as a supply, disposal, recreation, and affordable living space, this is gladly included in the idea of the Smart City. It is assumed that the city can exemplarily demonstrate what is possible with digital support. However, problems that are often not recognized or ignored become apparent when implementing digital recommendations and aids. The realization of the idea presupposes that the population—despite all proclamations of diversity—assumes a certain uniformity in many important respects. Value basis, language, lifestyle, and willingness to adapt to technical specifications vary greatly. Stable parallel societies have developed everywhere, not just in German cities, which hardly engage in exchange with other groups. The language problem is also growing larger. While the existing national language can somewhat be managed through translation machines, the nuances of a grammatically complex language with an extensive vocabulary, which is especially significant in political and cultural contexts, cannot be conveyed adequately. It leads to English, with all its problematic value associations, becoming the *lingua franca* of the digital age and thus finding application in the Smart City as well. In short: not only does the data situation become more complicated, but so does the exchange among citizens.

Let us take a closer look at the fundamentals of the smart city concept. Efficiency, sustainability, and productivity are not everything. Networking is also meant to meet global standards, although it is clear that parts of the world are still excluded from this globality because the necessary technical infrastructure is lacking or the network is completely controlled by those in power. Naturally, there will be preferences regarding information, especially since the goal of optimal service is at stake. Thus, there will be information preferences derived not only from the needs of the users but also from the interests of urban administrations. It is well-known that these interests often do not align. Often overlooked is the fact that there are people who do not want to become “smart citizens” and prefer to handle their affairs independently without municipal support—no matter how citizen-friendly it may be. Even if, in the medium term, it can be assumed that the majority of the population will adapt to digitization, there will still be a certain number of individuals—and these are by no means only the elderly or disabled—who reject integration into the system, especially since the offerings are also associated with surveillance and subtle forms of “manipulation.” This is often downplayed by referring to so-called pull factors. While synergies and organizational support may be desirable, and compromises in privacy protection may be

tolerated, it cannot be assumed that all digital offerings will be accepted. This is particularly true in diverse cities, where the composition of the population consists of parallel societies with vastly different value systems, which makes universal acceptance unlikely. The idea that municipal networks should also provide support for citizen projects sounds intriguing, especially since civic engagement is a municipal priority. However, it is reminiscent of the concept of municipal community television, which was promoted in the 1980s, raising great hopes for increased local citizen participation at the time. In the end, however, people chose commercial television over the often poorly executed municipal alternatives.

The hope for new forms of political participation—such as citizens gathering on new internet platforms—is certainly an interesting supplementary participation offering. However, assuming that this will lead to significantly higher acceptance of political activity is difficult to verify.

Life can become easier through digitization, networking, and the targeted use of intelligent systems. However, the assumption that technology will contribute to the development of grassroots democratization appears speculative. Many ideas of citizen participation, open data, participatory forms of involvement, etc., depend on conditions that presuppose a certain concept of human nature which is not universally shared. This means, first, that the existing heterogeneous concept of human nature is not yet ready for democracy or compromise. Second, it assumes that citizens have embraced technological support—which also involves a surrender of self-determination. Third, the technical framework of participation itself excludes certain citizens. To speak of an expansion of democracy through new digital participation processes seems, at least for now, inappropriate. Even in the older, quasi-“embodied” system of democracy, certain groups, such as people with mobility impairments, were excluded. However, this issue could be partially solved, for instance, through transportation assistance during elections. Furthermore, media-based communication is always subject to manipulation. While media have always played a role in democracy and participation in political decisions, we remain far removed from the idealized visions of the agora or the Rütli Oath.

The creation of “monopoly regions” in Germany, for example, could lead to corrections of narrow constraints—also of a historical nature—caused by state borders and could open up new opportunities for participation. It also provides the possibility of addressing issues of demarcation. However, the idea of the self-organization of informational systems

contradicts the universal claim of technology. Participation in regional decisions, for example in Belgrade, should, for good reasons, be prohibited for Australians, Russians, Americans, and North Koreans.

It is not without reason that there is significant criticism of buzzwords such as sustainability, climate change, digitization, and, of course, the concept of the smart city.² Too often, new technologies are used as tools for processes of societal change. It is true that new media also bring about societal transformations. However, societal change does not necessarily equate to a shift in political power structures. The ruling powers in such contrasting societies as Russia, the United States, and China use advanced system technologies equally to stabilize the societies they control, and they do so with great success. With the help of new technologies, the potential that suits their purposes is exploited—leading to scoring systems and/or surveillance systems, resulting in a “smart” panopticon. This potential is inherent in many advanced technologies but finds its temporary peak, if at all, in information technologies. It is worth noting that precisely these ideas of informational observation, control, and manipulation were already foreseen in classic dystopias of the pre-digital age, such as Aldous Huxley’s *Brave New World*.

This brings us to a discussion of resistance potentials that arise from the technology itself and that must be considered in all municipal and regional applications of information technologies. The fact that not all usage options will be mentioned in the following discussion is not only due to the limited space available but also to the reality that certain technologies will remain unaffordable for the foreseeable future and thus will not be accessible for public use. Thus, for instance, the quantum computer, whose physical design and necessary cooling are so complex that it seems reasonable to forgo its usage potential for the smart city concept, especially as the costs are immense and the benefit to the general public rather low. Let us summarize the key considerations regarding the smart city concept.

1. Even though the smart city concept, like many social and political terms, is vague and rather a collection of ideas and technologies under one umbrella, it is to be noted that it only gains

² S. Sassen, “Talking Back to Your Intelligent City,” 2012; R. Sennett, “The Stupefying Smart City,” 2012; A. Greenfield, “Against the Smart City,” 2013; J. Libbe, “Standpunkt: Smart City: Herausforderung für die Stadtentwicklung,” in *Difu-Berichte*. Deutsches Institut für Urbanistik, 2/2014, 2013; V. Weber, G. Ziemer: *Die Digitale Stadt. Kuratierte Daten für urbane Kollaborationen*, transcript, Bielefeld, 2022, (transcript-verlag.de, Open Access).

public attention and political relevance with a certain level of ambiguity and general accessibility. The smart city concept gains a new quality when AI systems become usable by practically everyone. With the introduction of ChatGPT, the question naturally arises whether the smart city concept can remain completely unaffected by it.

2. In the potential of the smart city, we also discover its limitations. The idea of smartness, of intelligently and cautiously adapting to the behavior and possibilities of a smart citizen, has boundaries where this behavior is so far removed from technical possibilities that certain rationalization methods fail at the value basis of the respective group to which the user belongs. The idea of the urban center's continuity is also dependent on the question of what constitutes this center or how it is defined. The groups of people who want to see things differently regulated and supported are growing in a diverse citizenry. Key points of the definition clash directly and irreconcilably. We tend to observe that existing parallel societies become more closed, and the danger of inaccessibility grows. The hope that it can be bridged through technology, that technical reason and rationality will break down even the hardest ideological and philosophical barriers, is likely in vain. Apparently, certain questions cannot be solved by technology and rationality. Even if technology can be an element of enlightening action, it also allows actions that mock any enlightenment.
3. The concept of the smart city fits wonderfully into the era of demanded digitization. Indeed, there are advantages associated with it. Administrative and organizational matters in particular, are more swift. Especially for a free economy, this results in the advantage of being able to react very quickly to events. However, it is doubtful if technology will benefit everyone. This does not mean negating areas where it brings advantages, but the hope that one can develop these areas without acknowledging the drawbacks would be naive.
4. The concept consists not only of digital advancements but also of advances in life organization, which can naturally turn into coercion; one might also say it is meant to manipulate and steer onto the 'right' track. Although diversity is repeatedly emphasized,

serving specific groups, the problem arises that the individual in the group may not receive service, but is instead viewed only as a type captured through group membership. Those who do not engage with the system will naturally not receive individual service. Planning one's own life may be a disruptive factor. The risk that the concept stands closer to a "Brave New World" than to an Earthly paradise (*Schlaraffenland*) is often overlooked. Without observation and the provision of private data, the advantages will likely not be available to all system participants. The system essentially thrives on observation and data access. Even with earlier use of "smart home" offerings, not all options were accepted. Users realized they often fell under a pull effect and did things they did not really want. Without a persuasion program, without pull moments, the system will not be enforceable. What is generally considered efficient and reasonable may not be so for every individual.

5. The concept is embellished with all sorts of idealistic ballast. The smart citizen is assumed to be interested in efficiency and sustainability. But not everyone shares this interest. Especially in leisure activities, people often prefer detours over efficiency. In addition, it is not uncommon that what is offered and generally considered acceptable is, based on the experience of many, not shared at all and by no means corresponds to their own wishes.
6. We are witnessing urban societies increasingly less able to steer the officially desired degree of divergence. In fact, digital society lives off a shared value base. But precisely this cannot be guaranteed in a highly divergent society. Common values usually exist only as incantatory formulas. What can reasonably be expected of everyone is becoming ever rarer.
7. A further problem arises from the emphasis on urbanity. Thinking follows the categories of the ancient Greek world, in which the city and its surrounding hinterland as a supply area formed a state. Mobility patterns—and not only individual ones—however, contradict the guiding principles of digitalization. After all, decentralization was precisely what the digital age set out to champion. The smart city concept is not the only one to reveal that the hopes for decentralization associated with it have not been fulfilled. Nowhere is the location of a company or an

institution irrelevant. It is assumed that urban life offers more opportunities for social integration while accommodating individual preferences than rural life. The promises of the digital age are thus counteracted by the concept. The hope that one could cultivate the smart citizen independently of economic and institutional interests is also hardly to be taken very seriously.

2. WHAT DOES RESISTANCE MEAN?

This brings us to some important technical aspects of the smart city idea. It becomes clear that the issue of resistance plays a central role.

We will now focus primarily on the technical side of the smart city concept, in which the use of advanced information technologies plays a particularly important role and from which the specific challenges of the concept and its near future can be inferred. Through technical systems, the idea acquires a new quality. Certain processes are organized close to real time. The technologically produced immediacy offers enormous advantages, but it also entails risks, since without human mediation errors can occur that may cause a project to fail—not only due to costs. Errors also occur with human mediation, but as a rule they are noticed more quickly, since one usually first looks for the problem on the side of mediation. Our attention, however, will be directed above all to current technologies and their potentials for resistance. The latter shows that the smart city idea will also be confronted with all kinds of technical problems, because precisely the idea of optimal functionality—and thus the simplification of environmental behavior, optimization of sustainability, simplification of planning, participation, and so on—harbors a great deal of resistance potential. Smartness therefore often runs the risk of proving to be “unsmart.” Things—including technical ones—do not readily conform to our wishes.

Technical forms of resistance are closely linked to physical ones. The discipline of physics is particularly tightly connected to experiences of resistance. Experiences of resistance stand at the beginning of human history—namely psychological, logical, value-setting, and social history, and not least technical history, which is closely tied to physical experiences of resistance. Physics can probably be explained entirely in terms of experiences of resistance: force/counterforce, pressure/counterpressure, light/refraction, and so on. Through technology we can learn ways of dealing with things that “soften” resistance, just as the carpenter learns

to use a plane to make the wood, as it were, suitable for his purposes. All technology serves the fulfillment of purposes. Through experiences of resistance, human beings become *homo faber*. Technology derives its justification from relief on the one hand, but also from the acquisition of new options for action. Technology, like culture, has emerged from experiences of resistance. Yet neither has completely eliminated resistance; rather, they have also generated new ones. Resistance never disappears entirely; instead, it undergoes transformations and shifts. In particular, the smart city concept is also concerned with technical autonomy, that is, with the question of what the system can do more or less independently for the citizen.

What, then, is meant by technical autonomy? As we shall see, the concept is closely linked to that of resistance, since it shows where the limits of processes of autonomization lie and what stands in the way of such processes. The concept of autonomy originates in political philosophy and means “self-legislation.” But what can it mean when we describe a technology as “self-legislating?” The focus here is not on the individual technique of a violinist, but on artifact-based technology that is objectified in tools, machines, and apparatuses. It is information and communication technologies capable of intelligent performance that, in their most advanced development, are described as “autonomous.” Strictly speaking, this designation is incorrect as long as technology serves us as a tool. Autonomy would only be present if it could also refuse its services. Autonomy presupposes that a technology pursues its own purposes. It would thus be capable of acting and would act on the basis of decisions rather than calculations. As an acting entity, it would be capable of assuming responsibility for the consequences of its actions. One can distinguish between strong and weak autonomy. Strong autonomy is what Kant identified as a foundation of ethics, namely the assumption—preceding any ethical discourse—that an ethical being is capable of regulating its will according to a general principle. Concretely, this occurs when it formulates laws that claim universal validity and to which it subjects itself. For system technologies to act in the sense of strong autonomy would require that they themselves are affected by what they do, can be held accountable for their actions, and pursue their own purposes. There would be no reason not to treat such systems as our equals. But why should they support our desires if doing so runs counter to their own intentions? In the case of weak autonomy, which is based on the “voluntary” following of given rules, the situation is different. That a system

follows rules is a trivial insight; after all, this is the basis of every information technology. But what is meant here by “voluntary?” Voluntariness presupposes insight into a rule, that is, a self-consciousness that finds expression, not least, in what Kant calls the “synthetic unity of apperception.” It must be clear that it is I who have had this or that experience, or who have carried out this insight or that action. Kant’s assertion that a “cogito” must be able to accompany all my experiences does not mean that it is always explicitly given. Thus, syntheses of one’s own experiences and insights, as well as an ego-consciousness, would have to be possible.

Since antiquity, causal relations have been interpreted in terms of compulsion. One might now object that in informatic systems causal relations are not at issue in the same way as in mechanical ones. Logical if–then relations are not causal, even if they occur by virtue of physical processes. They are timeless and therefore cannot suffer empirical disturbances. In system technologies, however, situations can arise due to changed data conditions that require a variation of rules or new rules in order to solve a problem. One may think of a computer game in which solutions can only be achieved by changing levels. But does changing have anything to do with voluntariness and insight?

In addition to the above insight about rules, there is another prerequisite: the absence of coercion in action. We realize our will by doing something that we could also refrain from doing. But can a system technology really refrain from acting? Is there an alternative to an action? Of course, we could use a random generator to determine a datum at a certain point from which new “if– then chains” are constructed. But does this have anything to do with voluntariness? Informatic processes are typological. The detection of deviations from rules is not due to historical change that leads to a re-evaluation of a state of affairs, but to the quantitative recording of changes. A system does not evaluate data on the basis of historical change. It may apply different methods to solve a problem, but it is not capable of refraining from solving a problem because it is deemed superfluous - unless it has been programmed that way in advance, in which case, of course, no autonomy is present. Informatic processes are not subject to the principle of individuation. Every operation must be transferable to another system technology or be executable by another computing system. Only in the case of the ultimately inexact operation of the quantum computer do problems arise from transferability. The quantum computer, however, ultimately does not operate exactly and is therefore unsuitable for solving certain problems. What it

does, it does at great expense. Its production and energy consumption, as well as its operation, are costly. It is certainly an unsuitable technology for solving everyday problems. Voluntary actions are decision-based. But what is decision-based is marked by a deficiency. Whoever decides chooses; whoever chooses excludes a possible course of action, can err, and can be held responsible for it. A system can therefore lead us onto a completely wrong track because the conditions on which the idea is based may change—for example, through external influences, not only of a natural but also of a political-historical kind.

A final argument that might seem to support speaking of technical autonomy would be the reference to the fact that even in human beings there is no complete autonomy—an argument that, however, misses Kant's concept of autonomy. Kant pointed out that the human being is a citizen of two worlds: on the one hand, as a natural being, determined; on the other hand, participating in the noumenal world, the world of ideas, and thus able at any time to initiate new causal chains. This is precisely the decisive point of his concept of autonomy. The fact that we are disposed beings, both in a natural and in a historical sense, does not mean that we are exhausted by this determinacy. The fact that in our actions we also orient ourselves toward ideas such as justice, freedom, and equality shows that Kant never assumed absolute human autonomy, but very much did assume the possibility of transcending determinacy. And it is precisely this ability to orient ourselves toward ideas and to realize them in a situation-relative sense that constitutes the special dignity of the human being. The question of the motivation of an action includes individual life histories and historically and culturally shaped influences. Every action can be interpreted socio-psychologically, which does not change the fact that such interpretation rests on evaluations. It does not explain the situational decision of the actor, that is, the termination of deliberation and the decision against one's own dispositions and imprints. Let us assume that "autonomous technology" is characterized by the ability to act independently without an explicit control command. Autonomy here is an attribution that produces effects. Thus, forms of relief are promised that can turn into forms of disenfranchisement when competencies are lost. The dialectic of relief and disenfranchisement is a hallmark of independently acting information technologies.

The idea of autonomy has been pursued since the very beginnings of computer science. From a technical point of view, the transition from automation to technical autonomy in the sense defined above is

a continuous one. At some point, automation processes were meant to take place not only on explicit instruction, but through technologies adapted to life situations—on the basis of data “flowing into” these systems via sensory apparatuses. In fact, we have passed the point at which we can still directly control networked informatic systems. The informatic environment has reached a level of complexity that allows even specialists only occasional, targeted interventions. We can exit a networked system only at the cost of relinquishing informational and practical possibilities and of social isolation.

Autonomous technology in the sense of independently acting technology has a high potential for relief, but also for disenfranchisement, if we entrust life decisions to it. We have become accustomed to background systems organizing our lives for us. The use of such systems goes hand in hand with the promise that the support they provide is compatible with the general interests of society. Advanced system technologies are supposed to be designed according to general, socially acceptable, and legally compliant criteria. A “machine of reason” would then moderate my interests and harmonize them with those of others and with general interests. It would be a machine of stabilization and mediation. This also raises a question of power: who programs this machine? Systems are in principle not neutral, just as no tool can ever be neutral, since it always stands in a means–end relationship.

Let us now turn to the central question: what can be meant by resistance?³ The idea of resistance has appeared in philosophy since the eighteenth century in various forms, without ever becoming an independent theory. In the context of bodily-vital experience of the external world, the discourse reached a peak around 1900 as evidence of its existence,⁴ but it also appears in the development of theories of value.⁵ In psychological contexts it initially receives only indirect attention, and in logical ones practically none at all. However, it gains significance in social contexts, such as in the analysis of institutions. Bollnow extends the discourse by

³ K. Wieglerling, “Exposition einer Theorie der Widerständigkeit.” *Philosophy and Society*, 32, 4, 2021, pp. 641–661; K. Wieglerling, “Technische Autonomie und das Problem der Widerständigkeit,” *Synthesis Philosophica*, 39, 1, 2024, pp. 73–90.

⁴ W. Dilthey, “Beiträge zur Lösung der Frage vom Ursprung unseres Glaubens an die Realität der Außenwelt und seinem Recht (1890),” in W. Dilthey, *Gesammelte Schriften V*, Vandenhoeck & Ruprecht, Göttingen, 1957, pp. 90–138.

⁵ G. Simmel, *Philosophie des Geldes* (1900), *Gesamtausgabe, Band 6*, Suhrkamp, Frankfurt am Main, 1989, pp. 34 ff.

introducing a correlative complement to the concept in the notion of “counter-support” (*Widerhalt*).⁶

The concept of *Widerständigkeit* (resistiveness) is preferred here to that of resistance in order to indicate a higher-order perspective on the problem. In the concept of resistiveness, different modes of experiencing resistance are brought together and identified as transcendental-philosophical concepts. The term is not object-referential; rather, it is a concept of reflection that thematizes the relationship we adopt toward things and states of affairs, toward other persons, and toward ourselves. It qualifies neither as a thing nor a state of affairs, but characterizes an essential aspect of our relation to things, states of affairs, and persons. Resistiveness is experientially accessible, although not necessarily in an object-like manner. It articulates itself in a sense of insufficiency; it is *παρουσία* (presence) without being present in the manner of a phenomenally given and qualifiable thing. Since resistiveness refers to relations, it has a positioning function. What the human being is emerges from the relations in which that human being stands. These relations articulate themselves not least in the form of historical dispositions and an orientation toward the future. To experience resistiveness is at the same time to experience one's position in the world. It is not the only, but an indispensable moment of reality. There is no experience of reality without an experience of resistiveness. Reality is that which ultimately eludes my individual will to form or construct. The attempt, pursued in modern information-technological concepts such as ubiquitous computing or ambient intelligence, to render resistiveness invisible or to circumvent it goes hand in hand with a loss of reality, extending even to psychopathological findings, when levels of reality can no longer be recognized and distinguished.⁷

Experiences of resistiveness are the precondition for self-experience and for the development of competencies, and thus also for the human being's learning and self-understanding as a learning being. Even in my bodily existence I experience that things are not entirely under my control. The value of things and of human beings—illness included—depends essentially on the resistiveness they offer. In the idea of dignity as being an end in itself, a resistance is articulated against the merely typological and against the possibility of being subjected to calculation.

⁶ O. F. Bollnow, *Das Wesen der Stimmungen* (1941), Königshausen & Neumann, Würzburg, 2009, pp. 85–98.

⁷ K. Wieglerling, *Philosophie intelligenter Welten*, Wilhelm Fink, München, 2011.

Resistiveness restrains the human will, but it is more than a moment that merely breaks or inhibits our will. It is a constitutive element of all cognition and theoretical reflection, without which all things would be equally ordered and devoid of cognitive value. Equally ordered things without gravitations would be an unrecognizable noise that allows articulations and disarticulations, but is not itself perceivable or qualifiable—merely an unordered flow of data. Without resistiveness there is no differentiation; without differentiation, no knowledge.

Let us also consider forms of resistiveness that in psychology and sociology are subsumed under the term resilience. The understanding of resistiveness used here has overlaps with resilience, but there is no full equivalence. The discourse on resistiveness has focused primarily on the ontological and epistemological significance of the concept. By contrast, the current psychological and sociological use of resilience seems more closely aligned with the classical understanding of tolerance. *Tolerantia* originally referred above all to the capacity for suffering or endurance of the legionary, who had to cope with deprivation, hardship, injuries, and psychological strain in order to remain capable for action. In a transferred sense, this also applies to functioning within social systems that must maintain their capacity for action under adverse external influences. The stability of a system depends on its adaptability to changing external conditions. This requires individuals within the system to possess a certain capacity for endurance. They must remain functional even when losses, restrictions, and deprivations occur. Resilience, as an expression of psychological and social capacity to bear strain, is regarded as a potential that can be enhanced. However, it can reach a degree that turns into dullness and runs counter to any form of empathy. While this may be a prerequisite for a stable personality, it can become problematic for one's social environment. In the case of technical support systems, adaptation to the character traits of the supported person can become problematic if it ends up reinforcing pathological behavior.

Let us ask, then, which forms of resistiveness can be identified in technology in general, and which in the field of informatic technologies in particular. Technology is not merely an applied form of natural-scientific knowledge; it is also a social construct, shaped by cultural and economic dispositions. Technological developments cannot be decoupled from available resources, economic valuations, societal value preferences, and configurations of power.

What, then, is the situation with regard to concrete experiences of resistiveness? The relativity of resistiveness is articulated, for example, in artifact-based technology. For a tank, the resistance of a stone on the road is marginal; for a passenger car, it is considerable. Experiences of resistiveness stand within a changing practical or theoretical context. Thus, we can focus on one and the same object from different disciplinary perspectives. Let us now name forms of resistiveness according to disciplinary perspectives:

Cultural resistiveness is articulated, on the one hand, in the “poietic” basis of a culture, which for Günter Janich represents an index of cultural maturity.⁸ It manifests itself in a lack of material and intellectual preconditions for the development of technologies. As the Erlangen School has shown, there is a logical order to technological developments: one cannot invent the automobile before the wheel. Cultural resistiveness also expresses itself in patterns of acceptance. Despite all the potential of modern technology to level cultural differences, cultures offer resistance to the development and implementation of technologies through their evaluations of states of affairs. While in Germany there are strong reservations about care robotics—expressed, for example, in the fear that their use will lead to the isolation of those in need of care—they enjoy high esteem in Japan. There, on the one hand, the possibility of a longer independent life is valued; on the other hand, there is a different understanding of shame: people prefer to be cleaned in intimate areas by a care system rather than by a human being. Forms of life give rise to specific technologies. The blowpipe presupposes a specific way of life that requires an efficient instrument for silent hunting. Artifact-based technology and individual technique thus form a functional complex. Operating competence plays a central role in the case of simple tools. In technological development, however, there is a tendency toward a decoupling of the artifact from its operation. Complex technologies are intended to be operable by laypersons. Yet only with the idea of independently acting technology does a complete decoupling seem possible.

The idea of the successful integration of technology is articulated in highly technologically developed cultures in the fact that it disappears from view and only becomes noticeable when it becomes recalcitrant, that is, when it malfunctions. In modern information-technological

⁸ G. Janich. *Kultur und Methode: Philosophie in einer wissenschaftlich geprägten Welt*, Suhrkamp, Frankfurt am Main, 2006, pp. 26–28.

constellations, however, we are confronted with another problem. A technology that acts independently and without explicit operation does not even allow resistiveness to come into view, since its task consists in eliminating or circumventing resistiveness in advance of its appearance. What we cannot perceive, however, we can neither control nor steer. With the disappearance of resistiveness comes a loss of reality. With it, agency also diminishes. A process of disenfranchisement and disempowerment takes place. The loss of resistiveness does not lead us into a realm of freedom, but into a land of plenty in which our participation is no longer required. Finally, resistiveness can also arise from practices of use that do not concern the purpose but the aspect of operation, which does not correspond to cultural practices: the desired purpose can only be realized with difficulty or inadequately using the existing operating apparatus.

Ecological resistiveness is closely bound to cultural resistiveness, since ecological perspectives also arise from evaluations of forms of life. At the same time, it objectifies itself in a threat to organisms, not least to human beings themselves. It manifests itself in limits on available resources or in a re-evaluation of the usefulness of a technology. Ecological resistance belongs, on the one hand, to the category of cultural resistiveness, insofar as it is based on social evaluations, but also, on the other hand, to the domain of natural-scientific resistiveness.

Economic resistiveness concerns the efficiency of technological applications. Technologies may fail to achieve the intended economic purpose if their provision and operation incur higher costs than the hoped-for economic benefit. Energy expenditure and material costs required to manufacture a tool can prevent its use and further development. It is the means–end relation that comes into view under economic preferences. If the relief provided by a technology entails additional economic burdens, its use is inefficient. If it presupposes the multiplication and expansion of customary activities, it is questionable whether this justifies the gain in new options for action. We encounter this in the development of quantum computers, whose added value apparently depends on the field of application. It would be inefficient to use a quantum computer for tasks that can be handled by conventional computers.

Technology-internal resistiveness appears, for example, in material fatigue, which exists not only in classical engineering but also in information technology. It manifests itself, for instance, in material incompatibilities. A specific material can withstand more heat than another. Incompatibilities usually become apparent only with delay. We can speak

of external resistances when environmental factors such as moisture, heat, or vibration act on technology and possibly render it unusable. Coping with internal resistiveness is the subject of engineering practice. Resistiveness in body-based manual techniques or intellectual techniques depends on organic functionalities that can be disrupted by accidents or illnesses.

Let us focus on general forms of resistiveness and ask why certain qualia are not machinally recognizable. Epistemological resistances occur when measuring apparatuses do not function and deliver false results. In such cases, however, we are dealing with technology-internal resistances, not with epistemological resistances that lead to problems of knowledge. Possibilities of control may be absent or insufficient. Even in classical media contexts, problems of resistiveness arise that affect our knowledge. Lighting conditions, film material, and the particularities of the camera can influence our view of things. The documentary character that essentially determined the triumph of photography turned out to be a staging, but also an unconscious misinterpretation of external reality. Just as the moving image provided us with insights into sequences of motion—for example, allowing us to recognize the “floating” of a horse at faster gaits—so too do perspectival settings distort states of affairs and proportions through camera angles and positions. Color values and lighting conditions come into conflict with real conditions. While stagings of external reality were still accessible to image analysis, the same does not apply in equal measure to apparatus-induced mis-mediations of external reality. In image-obsessed times, the situation presents itself differently. Certainly, every theory of representation is problematic, but this also applies to non-mediated, bodily perception. A “pure” representation would not yield knowledge either, since knowledge arises only through emphasis and delimitation, hierarchization and reduction, which is why there is no knowledge without experiences of resistance. The concrete qualia of our perception are not accessible to scientific access because they are subject to the principle of individuation; that is, they are experienced at a specific time in a specific place and arise on the basis of particular projections from prior experience.

Ontological resistiveness accompanies every technology, every technological development, and every formation of individual technique. In classical technology, ontological resistiveness initially articulates itself within existing stocks of being. It enters the world as an ordering factor when technical artifacts, infrastructures, or technically produced goods organize life, that is, establish an order that relieves human beings and

provides them with new options for action. This can be seen exemplarily in mobility technologies, which require rail networks, roads, ports, and so forth. Fuels must be provided, maintenance facilities made available. Our lifeworld is ordered by technologies: information and communication technologies have become our “ready- to-hand,” quasi centers of focus of our work and residential culture. Individual techniques are an invisible tool at our disposal for resisting the adversities of life. They serve to master resistiveness, whether objective or institutional. Means–end relations structure the world in which we live. It is built from “in-order-to motives,” in which technologies occupy a central place. Such in-order-to motives can come into conflict with one another, leading to the setting of preferences or to the abandonment of technical means. Heidegger saw in modern technology, which he called *Gestell* (enframing), a final expression of metaphysics that subjects all beings to calculability. The means becomes the end. Yet technology does not dissolve into calculation, for in its artifact-based and individual forms it becomes a constituent of the world—one that can be changed and transformed, but not wholly abolished. Technology is the poietic stock of a culture and thus the foundation of forms of life. Instrumentality itself can ultimately become the purpose of technical will. Purposive orientations can also refer to technical functionality, to the improvement and optimization of what exists, a condition typical of highly technologically developed societies and one that is also expressed in the smart city concept. Resistiveness, however, also articulates itself historically. Historical change occurs where something eludes calculability, while at the same time accompanying every process of calculation as something that cannot be grasped as an object, but nevertheless makes itself felt. It is the event-like and the singular that characterize historical resistiveness.

As a transition toward a focus on autonomous technologies, let us finally consider the ontological difference that articulates itself as resistiveness in the act of saying “is.” The question concerns the position we assign to technologies in the world and in our lives when we speak, for example, of the age of machines, the information age, or the age of autonomous technology. What does it mean when Capurro,⁹ in reference to Berkeley’s famous dictum, speaks of *esse est computari*? If the entire world is seen as an expression of perfect calculability, this Berkeleyan variation

⁹ 9 R. Capurro, *Homo digitalis: Beiträge zur Ontologie, Anthropologie und Ethik der digitalen Technik*, Springer VS, Wiesbaden, 2017.

captures that condition. But this also means that experiences that elude calculability are marginalized or denied the right to be said to “be.”

Logical resistiveness belongs to the fundamental informatic experience. It appears both in programming processes and in the aggregation of data sets. A result obtained from flowing streams of data may, under certain circumstances, no longer be reproducible, which renders it unusable for scientific purposes, though not for practical ones, such as the control of an operational process. Scientific results concern questions of continuity, which stand for the validity of a result. Continuity, however, can only be established in the typological realm. But what is the typological? This is a question that cannot be answered by quantification alone. Similarity is grounded in an evaluation, an act that is also culturally disposed. Quantitative analyses are something different from the evaluation of states of affairs. Computer science is not “pure” mathematics and is fundamentally situated in an application context. It is always already empirically “contaminated.” The question of how logical relations can be transferred to empirical ones is of fundamental importance. Hegel discussed this question in his *Logic* under the aspect of mediation; Peirce sought a solution to the problem in his diagrammatic foundation of logic, which explicitly refers to Kantian schematism, intended to mediate between the faculty of understanding and receptivity. The empirical cannot escape contingency, which logic seeks to eliminate. Here, too, a problem of mediation becomes visible that challenges every informatic application.

Which forms of resistiveness, then, can be identified in autonomous technologies? Autonomous technologies are information technologies. Classical mechanical technologies lead to automation, not to forms of independent action. Only at the level of a technology that participates in mental processes of organization, control, regulation, and cognition can the idea of autonomy arise. Autonomy in the strong sense, which would also include the refusal of services, is not what is meant when one speaks of autonomous technology. Rather, the issue is a technology that not only acts in an automated way to relieve us, but independently performs tasks without operation and explicit supervision. A certain purpose is to be realized, but the path by which this is achieved is left to the system.

When discussing the limits of autonomous technology, it has become customary to ask whether one is speaking about the current state of technology or about technology in principle. It is then pointed out that both the natural-scientific and the technical access to the world are typological. However, this mode of access cannot grasp the event-like,

which is the object of idiographic methods. Eventfulness or singularity, which are subject to the principle of individuation, are not reached by any technical or natural-scientific method. Only by abstracting from concrete experiential quality can something become an object of such approaches. When we speak of autonomy, we also speak of the possibility of action. In contrast to behavior, action is decision-based. According to Janich, every action is determined by three criteria: responsibility for consequences, autonomy in setting purposes, and rationality in the choice of means.¹⁰ Of these criteria, responsibility for consequences can hardly be fulfilled. Under what conditions could a system be held responsible for the consequences of its actions; and how should this be done, for example by cutting off a robot's energy supply? We could speak of autonomy in setting purposes only if the system pursued its own intentions, that is, had its own interests and preferences. Such a system could no longer be described as a tool. One might argue that it is conceived as a system of reason and that its own interests are identical with the general interests of reason. It would then act in accordance with cultural and moral standards, moderating and harmonizing our desires—something that is also articulated in the smart city idea. But can desires and interests coincide with the general interests of a society? First of all, it is clear that in the case of complete coincidence, the element of compulsion would be eliminated, and thus no ethical impetus would be present. Technical systems are not citizens of two worlds; they have no share in the noumenal world but are entirely part of the phenomenal world. The system's share in ideas is limited to its mathematical stocks, which, however, are not themselves "insighted." There is no execution of evidence, no intuitive grasp, which would presuppose self-consciousness. Thus, only the criterion of rationality in the choice of means remains attributable to a system. Indeed, systems can pursue problem solutions using different means. However, they have no capacity to transcend the given goal or to abandon it "of their own free will." Autonomy presupposes self-consciousness, which is characterized by the following factors: it must possess an identity over time, that is, an inner time-consciousness; there must be a passive synthesis that connects all my experiences with one another as mine. Thus, on the one hand, spatiotemporal formations of the phenomenal world must arise; on the other hand, there must also be

¹⁰ G. Janich, *Kultur und Methode. Philosophie in einer wissenschaftlich geprägten Welt*, Suhrkamp, Frankfurt am Main, 2006, pp. 69–86.

a “transcendental unity of apperception,” which articulates itself in the *cogito* that must be able to accompany all my representations. The quasi-personal unity of a system would therefore have to be given as a temporal identity. Furthermore, an autonomously acting system would have to possess the ability to transcend the experiences it has had, that is, to posit something “historically” new. The capacity for transcendence is a defining feature of human thought. Without it, there is neither the possibility of action nor of decision.

In doing so, do we expose forms of resistance in the context of information systems that claim to act autonomously.

- a) *Principle of individuation*: this concerns a form of resistance that is inaccessible to natural- scientific and technical modes of access. Such access arrives at typologies, not at events and singularities.
- b) *Ontological difference*: a technological system cannot determine itself as a historical entity subject to permanent change and perceiving the world only perspectively on the basis of bodily embeddedness and its binding to a flowing, living present. Even the idea of “embodiment,” widespread in social robotics, does not reach the moment of bodily embeddedness, because it fails to express the difference between *Leib* (lived body) and *Körper* (physical body) and remains within the third-person perspective of the body. A system also fails to arrive at a “historical ontology” because it lacks historical consciousness: it is unable to grasp and evaluate the course of time and to position itself within it.
- c) *Evaluation*: evaluations are made before and after calculative processes, yet they themselves stand within a hierarchical and relational order. The evaluation of historical processes, of things and states of affairs, depends on criteria such as self-consciousness, intentionality, and capacity for action. A system lacks the possibility of self-situating, which does not consist merely in spatial and temporal delimitations, because it does not possess a living present in which past and future touch and which in principle eludes fixation. Historical time is evaluated time - something different from spatiotemporal physical localization, but also from identity-forming inner time.¹¹

¹¹ K. Wieglerling, “Entgeschichtlichung und Digitalisierung,” in K. Koziol (ed.), *Entwicklung der Wirklichkeit: Von der Suche nach neuen Sicherheiten*, München (Kopaed), 2020, pp. 85–119.

- d) *Logical resistance*: logical resistance articulates itself in contradictions and inconsistencies, for example in the idea that something is to be produced in order to serve us, yet at the same time no longer complies with these services because it acts autonomously.
- e) *Resistance of nature*: autonomous systems cannot attain Plessner's first anthropological basic law of "natural artificiality." Even if hardware is subject to aging processes, these are not comparable to those of biological organisms. Biological material is not controllable and steerable in the same way as physical material. Living organisms are open systems that change, are vulnerable, and are mortal.

Immediacy: Plessner's second anthropological basic law of "mediated immediacy" also cannot be attributed to system technology, because it lacks a sense of immediacy and presence, that is, the unifying point of experience that reaches beyond the now. It remains within mediation and knows no immediate reality. The future and the past can only be grasped typologically and perspectively and can never be fully calculated. The present state of Laplace's demon can never be determined, since only a finite amount of data can be taken into account. Every cognitive apprehension of the world is preceded by articulations and disarticulations. For all data captured by measurement, there is corresponding data that is not captured. Moreover, systems lack historical attunement: a sense for aging, for the re-evaluation of data, and for the burden placed upon the present by the past. They know neither spontaneous actions nor action blockages or displacement actions, neither moods that promote action nor those that inhibit it.

Do limits to the claims of autonomous technology now arise from the identification of such forms of resistance? In information technologies there exists a dimension of logical resistance that is bound to their specific processing material: information. Already in cybernetics, information—alongside the physical quantities of matter and energy—was distinguished as a third, independent dimension of processing. Although it is not independent of physical processes, it marks a distinct level of resistance that cannot be treated in terms of physical causality. Logical resistances are violations of the axiomatic conditions of a system. To this sphere we can assign problems of mediation between experience and logical calculus, when information cannot be combined with other information or made to conform to computational operations.

Fundamentally, intentionality, inner and historical experience of time, as well as eventfulness and singularity, are dimensions of resistance that stand in the way of autonomy in both the strong and the weak sense. In the strong sense, because they mark the limits of a quasi-human system technology that has shed its character as a tool and can no longer be seen solely within a means– end relation. But forms of resistance also articulate themselves in the weak sense of a self-acting, “autonomous” technology: the system’s independence is not based on decisions, but on statistical evaluations of incoming data.

Human beings define themselves essentially through what they are capable of producing. That this which is produced often surpasses them in many respects has become a central topos when we speak of advanced AI, not least of “autonomous systems.” We are dealing with a technology that not only surpasses human beings in performance, precision, and efficiency, but is also meant to support them in their social behavior and even to relieve them of an often agonizing pressure to decide. Autonomous systems could thus be regarded as machines of reason insofar as they provide mediation services with overarching societal premises, legal frameworks, and assumptions of equity, that is, insofar as they balance individual and general interests. The idea that the individual possesses a faculty enabling them to act, on the basis of an universally valid principle such as the categorical imperative, even against their own inclinations and desires, would then be delegated to a system technology. This mediation would likely take place in a gentle form that leaves us with the illusion that our particular interests still play a role in system support. What would be lost along the way is the idea of a decision-based and self-directed life. If we were to submit to the guidance of a machine of reason, we would hardly have to reckon with societal resistance any longer. Where resistance to system prescriptions were to arise, programs would likely be initiated that are developed under the concept of “persuasive computing.” The idea that system support should eliminate or circumvent resistances in the realization of intentions is meant to extend even to areas that were previously regarded as accessible only to human beings: the domains of morality, life planning, and political and social organization.

Rules that a system is meant to apply are by no means always appropriate to a given situation, since their situational implementation depends on the cultural preferences and dispositions of the one who is to act in that situation. Fully yielding to a system decision may relieve us,

but total submission and subordination to a machine of reason comes at the expense of situational application. For machines of reason, situational implementations of rules constitute insurmountable forms of resistance, because they reduce complex contexts to typical stocks. They do not attain situational wisdom in Aristotle's sense of *phronēsis*. Autonomous technology represents a new stage in the degradation of the human being, because it harbors the potential for disenfranchisement. It can lead us into a land of Cockaigne, where life becomes easy because we no longer have to live it ourselves and experiences of resistance disappear. The more the human being is relieved, the fewer possibilities for action they have. The path to the land of Cockaigne does not proceed via decision-based actions, but via "liberation" from the need to decide.

Human beings will be understood as embedded in an assemblage with their own creations, within which a self-acting technology plays an essential role. Heidegger's fourfold of sky, earth, mortals, and divinities¹² will be supplemented by the things produced by mortals and placed on an equal footing with them. Yet even the most adaptive and self-acting technology will itself offer resistance. It, too, ultimately fails when confronted with the singular and the event-like, which—like our death—eludes typological access.

Let us deepen the discussion of the smart city concept by focusing on what is currently perhaps the most widely discussed dialog system, ChatGPT. Unlike the discussion of quantum computing, no advanced system technology intervenes so massively in everyday life, becoming usable for laypeople in the way a search system does. This development cannot be ignored in discussions of the "smart city," especially since the user base is already enormous. Just a few months after ChatGPT entered the market in November 2022, authors of factual texts working for institutions and industry complained of massive drops in commissions. Teachers and university lecturers report that pupils and students complete their assignments or seminar papers with the help of ChatGPT. The sudden leaps in performance in both form and content are all too obvious, as are strange, untraceable bibliographic references. Text and image generation are already possible at a considerable level, and even short film sequences can now be produced with dialog systems. It is assumed that self-improving generative AI such as dialog systems will also become capable of ever better results linguistically. In fact, so-called large language

¹² M. Heidegger, *Holzwege*, Vittorio Klostermann, Frankfurt am Main 1950, p. 22.

models (LLMs) and context-sensitive text analyses are capable of this and deliver suitable and usable texts according to the task posed. The more dialog systems are fed with texts, the better they seem to become. However, generative AI does not necessarily only become better or smarter; it can also become worse and dumber. “More” text data and “more” usage data do not automatically lead to better results. This “mass ideology” is widespread but remains as questionable as ever. What matters decisively is the context of use—and there, less can be more. Improvements are to be expected not only in technical terms, but also in dialog management and control capabilities. Beyond narrowly technical improvements, enhancements in dialog management, such as follow-up questions, suggestions, or objections, could also provide advancements. Through skillful input of terms, we can quickly achieve the fulfillment of our search intentions. Dialog systems are, of course, more complex information technologies, but they too are subject to conceptual or natural-language control.

Dialog systems are based on methods of machine learning. However, they are not technologies that become initiative-taking on their own on the basis of a perceived situation in the environment. Even if a system is capable of autonomously completing a story, it does not itself become initiative-taking on the basis of an event, an “insight,” or even an “act of will,” but rather on the basis of control commands or programming: become initiative-taking at time *x* if event *y* occurs. The system is thus user-centered and user-controlled. The more precise the dialog management, the better the result, with the user—and the successful practice of application—ultimately determining the quality of the result. The possibility of self-improvement is tied to many prerequisites, including those related to a system’s historical sense, such as a sense for the change of concepts, for the aging of punchlines, for the performative dimension of speech that appears ironic or sarcastic in certain situations but not in others. Linguistic knowledge alone is not sufficient for this. Linguistic and performative competence are different, as can be seen in autistic individuals on the one hand and pantomimes on the other: the former often possess excellent linguistic competence but lack performative abilities, while the latter exhibit the opposite.¹³

¹³ K.-O. Apel, “Sprechakttheorie und transzendente Sprachpragmatik zur Frage ethischer Normen,” in K.-O. Apel (ed.), *Sprachpragmatik und Philosophie*, Suhrkamp, Frankfurt am Main, 1982, pp. 10–173.

Although public perception of dialog systems focuses primarily on text generation, developers also have additional fields of application in view across various scientific disciplines. In addition to their use in literature research, dialog systems are already being used to generate new medications. Risks arise not only from the potential of the technology, but also from application practices, which bring to light problems that have neither been recognized so far nor derived from already identified potentials. As with any technology, usage practices cannot be predicted. Identifying them requires continuous monitoring of development. Any assessment of the state of the art must factor in development potentials; indeed, we cannot understand enabling technologies at all without interpreting their potentials. In this respect, it is a kind of hermeneutics of the future.¹⁴

Many hope that dialog systems will soon be able to access the entire web and thus become even better and more up to date. Unfortunately, this hope overlooks the fact that unrestricted access also entails access to fake news and the nonsense disseminated online, not to mention manipulations to which statistically operating systems' texts will be subjected. Generative systems can not only become better and smarter, but also worse and dumber.

When it comes to reference to reality, what matters is by no means only the quantity of statistically evaluable material, but also data quality, its conceptual framing, the assignment and attribution of data, and the weighting of data. Heuristic procedures may be superficial and reductive, yet they play a significant role not only in everyday life but also in scientific practice.¹⁵

Central to every experience of reality is the experience of resistance.¹⁶ Experiences of resistance indicate limits of availability. The question of reality itself proves to be more difficult than it appears from the perspective of naïve empiricists who leave the preconditions of scientific

¹⁴ Ž. Radinković, *Hermeneutik der Zukunft*, Dissertation, Universität Stuttgart, 2011; A. Grunwald, *Technikzukunft als Medium von Zukunftsdebatten und Technikgestaltung*, KIT Publishers, Karlsruhe, 2012; K. Wieglerling, "Grundprobleme einer Hermeneutik der Leiblichkeit in Zeiten der Transformation des menschlichen Körpers," *Philosophy and Society*, 25, 4, 2014, pp. 137–154.

¹⁵ H. L. Dreyfus, *Die Grenzen künstlicher Intelligenz: Was Computer nicht können*, Athenäum, Königstein, 1985.

¹⁶ K. Wieglerling, "Exposition einer Theorie der Widerständigkeit," *Philosophy and Society*, 32, 4, 2021, pp. 641–661; K. Wieglerling, "Technische Autonomie und das Problem der Widerständigkeit," pp. 73–90.

experience unexamined. Computer hallucinations obscure resistances and indicate availabilities where none exist. Hallucinations correspond neither to an object accessible in space and time nor to a state of affairs locatable in space and time. Ultimately, only those results should find acceptance that withstand critical examination. What is problematic is an inappropriate trust in the results delivered. Dialog systems can be helpful tools, but also evidence of an increasingly difficult-to-control sorcerer's-apprentice technology and a land-of-Cockaigne technology whose promise of relief turns into disenfranchisement and loss of competence.

The quality of the results delivered by dialog systems depends, on the one hand, on the user's competence in dialog management and, on the other, on the system's adaptive capabilities with respect to the user. From the user's dialog management, the system "adopts" preferences and intentions. On the one hand, the system has no will of its own, no intentionality, and no autonomy, and thus does not leave behind its character as a tool; on the other hand, it opens up its own medial space and is, as a medium, more than a tool: namely a platform for the expression of anonymous system-generated texts, images, etc., which in turn co-determine our social, psychological, and historical-cultural reality. With regard to the further development of dialog systems, it is to be expected that advanced information and communication technologies and robotics will be brought together or integrated. The dialog capabilities of the systems themselves will undergo further optimization, possibly leading to effects such as those depicted in the science-fiction film *Her* (USA 2013) by director Spike Jonze:¹⁷ a young man named Theodore, abandoned by his childhood sweetheart, acquires a dialog system that supports him in everyday life with a pleasant female voice and stands by his side day and night as an understanding dialog partner. He builds an increasingly intimate relationship with the system, which he names Samantha, culminating in a quasi love affair. There are in fact few reasons that stand in the way of realizing such a system. It is to be expected that there will be systems to which we will attribute autonomy, intentionality, and even a will of their own. Many will develop emotional bonds with adaptive systems and will also be disappointed by them: thus, in *Her*, Theodore learns at the end of the film that Samantha maintains similarly intimate relationships with thousands of others. Dialog systems will also play an

¹⁷ S. Jonze, *Her* (USA 2013).

important role in professional life and will further contribute to the dynamization and acceleration of social processes.

The formula from media theory that every new medium also brings compensations for losses of competence seems no longer to hold in the case of dialog systems. The likely erosion of writing and structuring skills can hardly be compensated by an improved ability to control systems through dialog management. At the very least, no symmetries between losses and gains in competence can be demonstrated any longer. Elementary cultural techniques such as reading and writing are prerequisites for many forms of competence acquisition, but also for the development of judgment, which is indispensable for reviewing generated texts and grasping the limits and weaknesses of generating systems.

Not insignificantly, systems can also become dumber. System outputs can thus obstruct our view of the reality of action or lead us to misjudge reality. This may be due to faulty control, data quality, and above all to the system's inability to position itself with respect to the data it processes. Ultimately, however, the problems also lie essentially in the loss of resistance associated with advanced information and communication technologies penetrating everyday life. The dream of automated self-improvement is based on a misjudgment of the role that masses of data play for the correctness of results. Dialog systems may be relieving and highly efficient, but the claim that they allow us to better recognize the reality in which action is to take place holds only to a limited extent. This does not deny that improvements can occur; it merely denies that there is an automatism or necessity of improvement.

Uncompensated losses of competence always affect our capacity for action and thus also the metaethical condition of choice. If I cannot swim, I cannot help a drowning person. Possibilities for responsibility are reduced by the erosion of competence, and the danger of disenfranchisement grows when independent action and the verification of results are no longer possible. As in the case of *Her*, advanced systems—a cinematic anticipation of dialog systems—are capable of clouding our relation to reality. An adaptive system can in fact adapt itself to our habits, expectations, and intentions in such a way that we attribute personal characteristics to it; this now occurs in a manner that has nothing to do with Weizenbaum's relatively primitive ELIZA test.¹⁸ System capabilities

¹⁸ J. Weizenbaum, *Die Macht der Computer und die Ohnmacht der Vernunft*, Suhrkamp, Frankfurt am Main, 1978.

have become so sophisticated that personal attributions virtually impose themselves. Who would not want a dialog partner who is always available, who responds understandingly and empathetically to all problems and questions? It can therefore indeed come to a depersonalization of social relationships. The loss of reality is indicated not least by the fact that adaptive systems lack the resistance that characterizes every personal relationship. People are not available always and everywhere, not even in cases of extreme dependence. Since we ultimately find ourselves with dialog systems in a medial space, resistance-based experiences of reality to which media representations refer can be undermined.

Since generative AI entails a constant transformation of the state of knowledge, it must be assumed that conflicts *with* and transgressions *of* legal frameworks may also arise. Changes in knowledge have an impact on legal frameworks. In the case of generative AI, however, one must expect such an extreme acceleration of the state of knowledge that neither the legal nor the political discourse preceding it can be conducted in an appropriate manner. For reasons of legal certainty, laws cannot be continually adapted to changing states of knowledge; they must therefore possess a certain durability. If, however, on the one hand the attribution of texts can hardly be guaranteed any longer and, on the other, the speed of dissemination can scarcely be curbed, this also has consequences for legal frameworks. What was legally compliant yesterday need not be so tomorrow, without a manufacturer having changed anything at all in the technical preconditions of the system. The acceleration and dynamization of social processes fueled by generative AI can therefore counteract the legal stabilization of society.

The classification proposed in the EU Act into high-risk applications and applications with no or low risk does not do justice to the problem of generative system technologies. An application may carry low risk today but high risk tomorrow, without anything about the technology itself having changed. In any case, the distinction between high-risk AI and low-risk AI can hardly be maintained in the case of generative AI. Even if we focus—as the EU Act reasonably does—on fields of application, this clarity cannot be upheld without difficulty in the case of self-learning systems. Depending on the application, the duration of use, and the intensity of use, a learning and thus changing system can become problematic. The risks of dialog systems are therefore highly dependent on their concrete use.

3. USABILITY, MANIPULATION, AND FALSE HOPES: THE CITY OF THE FUTURE?

Let us therefore summarize what constitutes the smartness of the smart city concept and where the limits of this concept lie, focusing in particular on the use of advanced AI systems, since it can be assumed that these will play an important role in the further development of the smart city concept.

- a) We have been talking about smart cities for a long time, but only after AI systems became, as it were, usable by everyone did the idea reach a new stage. With the introduction of ChatGPT, public discussion has attained a new quality. Smartness has since meant that the “smart citizen” also expects precise operation by an autonomous technology.
- b) Every form of resistance is, as it were, the announcement of a boundary—one that can sometimes be overcome, but sometimes proves to be an absolute boundary that cannot be overcome by technology.
- c) Through the possibility of expanding the concept by using modern information technologies, especially autonomous or self-acting technologies, the narrowing of organization through the information network is reopened—naturally, at the price that concreteness and implementation become more diffuse. On the one hand, the general idea of digitization and networking is preserved; on the other hand, the idea suffers precisely in terms of the concreteness of implementation. Physical, cultural, and psychological realizations differ from place to place. Ultimately, the benefit of information technology is not local and singular, but primarily typological. Thus, precisely when emphasizing the divergence of urban society, it is ultimately not the individual who is served, but the type that identifies the individual as a member of a particular population group.
- d) The system essentially lives off observational data and other forms of data access. Protection of privacy, manipulation, and efficiency gains are in constant conflict here. Users will notice that they are subject to a pull effect when using such systems. One may allow oneself to be persuaded by this, but one does not have to. The ideal state is therefore not achieved, since the world, as it

were, also has a say locally in organizational, productive, and political assistance. Ultimately, the system serves the type, not the individual.

- e) The promises of the digital age are thus, to an extent, counteracted. The hope that the “smart citizen” can be cultivated, as it were, independently of economic and individual interests is also doubtful. The digital economy in particular is far removed from ideas associated with the concept of a social market economy. Society is changing rapidly through system technologies, but not in the sense of political transformation, not in the sense of progressive democratization, co-determination, or participation. Only new forces of information regulation and manipulation have been added. New, seemingly autonomously acting systems influence both the search for offerings and the offerings themselves. The hopes of control and optimization associated with the idea of the smart city are rendered more difficult rather than easier by advanced information technologies.

The future of the city will certainly involve massive support through information and communication technologies. However, it must be doubted that information possibilities will lead to a more participatory and more democratic society. Autonomous and generative technologies in particular tend rather to lead out of the information constrictions associated with the smart city idea. The problem of translating what is logically required into the empirical reality of life ultimately remains unresolved.

REFERENCES

- Apel, Karl-Otto (1982), “Sprechakttheorie und transzendente Sprachpragmatik zur Frage ethischer Normen,” in Karl-Otto Apel (ed.): *Sprachpragmatik und Philosophie*, Frankfurt am Main: Suhrkamp, pp. 10–173.
- Bollnow, Otto Friedrich (2009), *Das Wesen der Stimmungen* (1941), Würzburg: Königshausen & Neumann.
- Capurro, Rafael (2017), *Homo digitalis: Beiträge zur Ontologie, Anthropologie und Ethik der digitalen Technik*, Wiesbaden: Springer VS.
- Dilthey, Wilhelm (1957), “Beiträge zur Lösung der Frage vom Ursprung unseres Glaubens an die Realität der Außenwelt und seinem Recht (1890),” in Wilhelm Dilthey, *Gesammelte Schriften V*, Göttingen: Vandenhoeck & Ruprecht, pp. 90–138.
- Dreyfus, H. L. (1985), *Die Grenzen künstlicher Intelligenz: Was Computer nicht können*, Königstein: Athenäum.

- Greenfield, Adam, "Against the Smart City"; online: <https://urbanomnibus.net/2013/10/against-the-smart-city/>
- Grunwald, Armin (2012), *Technikzukünfte als Medium von Zukunftsdebatten und Technikgestaltung*, Karlsruhe: KIT Publishers.
- Heesen, Jessica (2005), "Analyse zu Szenarien einer intelligenten Wohnumgebung," in Jessica Heesen, Christoph Hubig, Oliver Siemoneit, Klaus Wieglerling (eds.), *Leben in einer vernetzten und informatisierten Welt. Context-Awareness im Schnittpunkt von Mobile und Ubiquitous Computing*, Nexus Schriftenreihe, SFB 627 Bericht Nr. 2005/05, Stuttgart.
- Heidegger, Martin (1950), *Holzwege*, Frankfurt a. M.: Vittorio Klostermann.
- Janich, Günter (2006), *Kultur und Methode: Philosophie in einer wissenschaftlich geprägten Welt*, Frankfurt am Main: Suhrkamp.
- Jonze, Spike: *Her* (USA 2013) (movie)
- Libbe, Jens (2013), "Standpunkt: Smart City: Herausforderung für die Stadtentwicklung," in *Difu-Berichte*, Deutsches Institut für Urbanistik, 2/2014.
- Radinković, Željko (2011), *Hermeneutik der Zukunft*, Dissertation, Universität Stuttgart; Sassen, Saskia (2012), "Talking back to your intelligent city"; online: <https://web.archive.org/web/20160108073226/http://voices.mckinseyonsociety.com/talking-back-to-your-intelligent-city/> (accessed 29 May 2025).
- Sennett, Richard (2012), "The Stupefying Smart City," *Urban Choreography*, 10.6.2013; online: <https://urbanchoreography.net/2013/06/10/richard-sennett-thestupefying-smart-city/>
- Simmel, Georg (1989), *Philosophie des Geldes* (1900), *Gesamtausgabe*, Vol. 6, Frankfurt am Main: Suhrkamp.
- Weber, Vanessa and Gesa Ziemer (2022), *Die Digitale Stadt: Kuratierte Daten für urbane Kollaborationen*, Bielefeld: transcript. (transcript-verlag.de, Open Access)
- Weizenbaum, Joseph (1978), *Die Macht der Computer und die Ohnmacht der Vernunft*, Frankfurt am Main: Suhrkamp.
- Wieglerling, Klaus (2011), *Philosophie intelligenter Welten*, München: Wilhelm Fink.
- Wieglerling, Klaus (2014), "Grundprobleme einer Hermeneutik der Leiblichkeit in Zeiten der Transformation des menschlichen Körpers," *Philosophy and Society*, 25, 4, pp. 137–154.
- Wieglerling, Klaus (2020), "Entgeschichtlichung und Digitalisierung," in Klaus Koziol (ed.), *Entwicklung der Wirklichkeit: Von der Suche nach neuen Sicherheiten*, München: (Kopaed), pp. 85–119.
- Wieglerling, Klaus (2021), "Exposition einer Theorie der Widerständigkeit," *Philosophy and Society*, 32, 4, pp. 641–661.
- Wieglerling, Klaus (2024), "Technische Autonomie und das Problem der Widerständigkeit," *Synthesis Philosophica*, 39, 1, pp. 73–90.

Graham Harman*

THE HUMAN CURATOR: AI AND THE FUTURE OF WRITING

ABSTRACT: While academics in the humanities and social scientists tend to be political progressives, they often behave more like reactionaries when it comes to technological innovation. Some of our professional furor against ChatGPT and other AI chatbots is certainly by the ease of plagiarism it enables, though presumably a consensus solution will eventually emerge for this problem. Yet the real anger directed at this technology apparently stems from a misplaced pre-emptive nostalgia about the role of the writer. We mourn in advance the solitary author summoning words onto blank paper *ex nihilo*, ignoring that the essential feature of writers lies elsewhere.

KEYWORDS: artificial intelligence, ChatGPT, Martin Heidegger, Ronald Purser, writing

* Graham Harman: Southern California Institute of Architecture (SCI-Arc), Los Angeles; graham_harman@sciarc.edu.

This is an Open Access article under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not transformed in any way and is properly cited.

I. ACADEMIC NOSTALGIA

It would be fair to say that most academics in the humanities and social sciences pride themselves on being politically progressive to a greater or lesser degree. This would suggest an openness to change, and more specifically, an awareness of the contingency of present social arrangements. Yet how quickly this openness vanishes whenever it is a question of technological change! Here I have found that my colleagues in the professoriate are often the first to resist the transformative implications of emerging media, forever focusing solely on what is about to be *lost* rather than gained. This discrepancy between an evident willingness (indeed, eagerness) to challenge the socio-political *status quo* and a reactionary clinging to existing technical procedures in intellectual life is not just a topic for abstract thought; it is also one that quickly instigates social crises among intellectuals. Namely, I have found that it often provokes outbursts of rage from those who apparently see themselves too clearly in the joint diagnosis “political progressive plus technological reactionary.” Recently I was shocked when a fellow professor of philosophy, someone of whom I have always thought highly, responded with a mixture of anger and (perhaps theatrical) disdain when I posted a ChatGPT exercise on social media. What I shared was rather innocent: ChatGPT’s staging, at my request, of an irreverent assortment of additional characters added to a famous scene from Shakespeare. The fellow professor in question was someone I have known for almost three decades. He had never aimed an unkind word in my direction at any time previously, or at least not to my knowledge. And this is just one such example of unexpected rage at the automation-in-progress of various intellectual functions at the hands of chatbots and other incarnations of artificial intelligence. While it is true that most teachers are at their wit’s end with such technologies due to undeniably rampant cases of student cheating, such overreactions to the newly automated strands of intellectual life must have deeper emotional roots. Where are those roots located? Here I will reflect on that question.

Let us begin with a thinker who was no political progressive: Martin Heidegger (1889–1976), widely known as both a major philosopher and a politically compromised Nazi, along with increasing evidence of much deeper anti-Semitic attitudes than was previously confirmed.¹ Heidegger’s retrograde grumblings in *The Question Concerning Technology* are

¹ M. Heidegger, *Ponderings II–VI: Black Notebooks, 1931–1938*, Indiana University Press, Bloomington, 2024.

proverbial, and are never really salvaged by the efforts of his disciples to make them sound less negative than they are, or by related efforts to imply that there is something shallow or vulgar about calling them negative in the first place.² Here I will cite from a different piece of work by Heidegger, the 1942–1943 Freiburg lecture Course on Parmenides, though it displays the same spirit as his better-known remarks on technology:

It is not accidental that modern man writes “with” the typewriter and “dictates” “into” a machine. The “history” of the kinds of writing is one of the main reasons for the increasing destruction of the word. The latter no longer comes and goes by means of the writing hand, the properly acting hand, but by means of the mechanical forces it releases. The typewriter tears writing from the essential realm of the hand, i.e., the realm of the word. The word itself turns into something “typed.”³

Aside from the excessive scare-quoting (an alarming premonition of Derridean stylistic tics-to-come), the tone of dour pessimism is familiar to everyone who has read Heidegger. While he tries his best to adopt the decisive tone of a sage whose superior wisdom ought to be self-evident, there are multiple problems with this passage. First among them is that the somber Black Forest philosopher never explains what writing has to do with “the essential realm of the hand,” even though it is anything but clear that the hand has any *essential* relation to writing at all. Presumably no one would argue that the printing press “tears writing from the essential realm of the monastic quill pen and Chinese wood-block,” since the latter devices are generally viewed as nothing more than quaint relics of a specific stage in the history of technology, no more essential in their own realm than swords and battleships in the military field. Likewise, it would be hard to see the ballpoint pen as having torn writing “from the essential realm of the fountain pen,” even if the latter still has its devotees. It would also be odd to denounce Microsoft Word for having torn writing from “the essential realm of the typewriter,” even though prominent Heidegger scholar Richard Polt is a sworn devotee of this still lingering technology.⁴

² M. Heidegger, *The Question Concerning Technology and Other Essays*, Harper & Row, New York, 1977.

³ M. Heidegger, *Parmenides*, Indiana University Press, Bloomington, 1992, pp. 80–81.

⁴ R. Polt, *The Typewriter Revolution: A Typist's Companion for the 21st Century*, Countryman Press, New York, 2015.

In fact, Heidegger offers no reason for why the human hand has any essential involvement in writing at all. If Homer existed as a human bard with no eyesight, then it stands to reason that the *Iliad* and the *Odyssey* were orally recited poems that no hand recorded until later in their history. We know this with greater certainty about the *Qur'an*, one of the most influential books in human history, which for devout Muslims counts as a direct dictation to the Prophet Muhammad mediated by the angel Gabriel. Although Muslims still revere the practice of memorization and oral recitation of the *Qur'an*, printed copies of the book are not despised, but are themselves handled with special care in every household devoted to Islam. Plato's powerful dialogue *Phaedrus* famously recounts the objections of some ancient Egyptians to writing itself, based on the claim that this new technology will be more poison than cure: that it will only serve to erode human cognitive skills associated with memory.⁵ In the passage cited above from his *Parmenides*, Heidegger strikes the pose of a deeply serious and classical figure in opposition to the purported frivolity of typists. Yet his ostensibly grave and wizened commitment to "essential" handwriting would only make him look like an empty, trend-chasing hipster to those ancient Egyptians in the *Phaedrus* who swore by memory alone.

More generally, we can say that the insistence on pausing at any given stage of technology risks the same arbitrariness as the Amish commitment on horses, buggies, and barn-raising culture, when one might return instead to far earlier Neolithic social systems. Why continue to have dogs, given that they were domesticated only around 30,000 years ago, after tens of thousands of years of human history without them? Indeed, why stick with behavioral modernity at all, since music, dance, abstract thinking, and burial of the dead are all relative latecomer practices among *homo sapiens sapiens*? Why retain even language, given that it "removes human communication from the essential realm of grunting gesticulation"? Why not enforce global nudism, given that clothing deprives us of an "essential" relation to our natural bodily forms? Some will say I exaggerate, since there may be cases where technology ought to be halted in the name of the general good. Organized labor often demands a stop to ongoing automation, whether on the factory floor or in the Hollywood

⁵ Plato, *Phaedrus*, Hackett, Indianapolis, 1995. For one helpful discussion see M. Coeckelbergh, D.J. Gunkel, *Communicative AI: A Critical Introduction to Large Language Models*, Polity Press, Cambridge, 2025, p. 8.

Screen Writers Guild. But while such demands may be needed for a time to prevent widespread professional disruption, they are dubious as a long-range strategy. A century from now it will surely be ludicrous to aspire to assembly-line work or a scriptwriter's career, which are likely to be obsolete targets of derision, much like those antiquated positions financed to this day by the British Monarchy: Royal Shoe-Wearer, Warden of the Swans, Groom of the Stool, or Royal Bargemaster. Perhaps the sole exception concerns offensive weaponry, since the increasing speed, power, and radioactivity of wars threatens to make every inch of the earth a battlefield, so that few would survive even an accidental discharge of nuclear weaponry. At any rate, Heidegger's histrionics notwithstanding, no comparable case can be made against the harmless typewriter.

The problem at play, of course, is that of nostalgia. Individual human lives, like human civilizations more broadly, move from one phase to the next with what often seems like alarming speed. As a natural part of this process, humans end up adopting certain new things while leaving others behind, though often enough these abandoned entities resurface later in new guise.⁶ Whenever this happens, we encounter the melancholic human tendency (especially among those who reached maturity amidst objects now disappearing) to focus primarily on what has been lost. I suspect it is a subspecies of the natural envy towards youth felt by aging people as they grow increasingly unhealthy and unattractive, failing to appreciate sufficiently their own compensatory gains in wisdom and inner peace. This tendency can even be detected in one of philosophy's oldest thought experiments: the Ship of Theseus. Imagine that the ship of the ancient Athenian hero Theseus is slowly renovated by replacing one piece of decaying wood with a new one, then another piece, and then another, until at some point we have a ship that no longer contains any of the wood from the original historical ship. Imagine further that someone else takes the discarded pieces of old wood and secretly reassembles them, until finally the old pieces of wood are restored in the form of the original ship. Which of these two counts as the "real" Ship of Theseus? So goes this old chestnut of a problem. Yet what both options assume is that a thing remains the same thing primarily by way of *preservation*, though this need not always be the case. More specifically, what if the real

⁶ Here Marshall McLuhan's thoughts on "retrieval" are helpful. See M. McLuhan, E. McLuhan, *Laws of Media: The New Science*, University of Toronto Press, Toronto, 1992. An earlier work on the same theme is M. McLuhan, W. Watson, *From Cliché to Archetype*, Penguin, New York, 1970.

Ship of Theseus consists not in a particular physical state that ought to be maintained, but is more like a disembodied principle that surpasses any particular set of pieces of wood, or any specific configuration of them? Elsewhere I have argued that a shipwreck rotting on the ocean floor might even give us a more essential version of the ship precisely *because* it has lost the majority of its details.⁷ An analogous argument failed with the French public, which overwhelmingly demanded that the Nôtre-Dame Cathedral be rebuilt with literal sameness following the devastating 2019 fire. Although understandable given the historical and religious significance of Nôtre-Dame, a number of architects proposed alternative restorations that, in my view at least, revived the deeper *spirit* of the cathedral in ways that the chosen replication probably has not. The point I am trying to make is that Heidegger's view of the "essential" aspect of any technology is far too antiquarian in spirit, in Nietzsche's sense of the term.⁸ Any genuine care for essence—and I am not among those inclined to mock this term—would require an active abandonment of a building's or a technology's most transient and accidental features. Far from being an essential aspect of writing, the human hand is likely to prove no more than a transient historical scaffolding that supported for a time the more essential features of the art.

The danger of nostalgia, then, is the risk of becoming overly enamored of one transient historical moment of a thing while paying little attention to its underlying spirit, which requires growth and reconception rather than mere preservation. When Amazon introduced the Kindle in 2007, I remember the appalled reaction of many other academic friends. There were some obvious advantages to the new technology, ranging from the ability to order and read books instantly anywhere with a Wi-Fi connection, and the chance to reduce the number of trees annihilated to produce paper for printed books. Furthermore, many of us know through personal experience the incredible hassle faced by intellectuals whenever they change homes, with the need to move thousands of paper books weighing many tons from one place to another. How much better to have them all accessible at the touch of a button!

⁷ G. Harman, "The Shipwreck of Theseus: Philosophy and Maritime Archaeology," in S. A. Rich, P. B. Campbell (eds.), *Contemporary Philosophy for Maritime Archaeology: Flat Ontologies, Oceanic Thought, and the Anthropocene*, Sidestone Press, Leiden, 2023, pp. 57–73.

⁸ F. Nietzsche, *Untimely Meditations*, Cambridge University Press, Cambridge, 1984.

Naturally, paper books have their charms just like any threatened medium, and among academics (who are often especial devotees of some past age or another) the nostalgic element tends to retain the upper hand. It is not simply a matter of the superior tactile experience of a book by contrast with current electronic devices. It is also easier to take notes in a paper book, and often easier to flip through one in search of a particular passage, despite the ostensible speed of electronic word searches. More than this, there is the inherent draw of used bookstores, filled with dusty treasures in the form of paper objects randomly and suddenly discovered. I have felt few greater delights on this planet than accidentally finding a first edition copy of Heidegger's *Habilitationsschrift* on the pseudo-Scotus for a mere \$25.⁹ The difficulty of obtaining certain paper books has also long served to justify research grants to travel to distant libraries: another of the treasured perks of the modern academic lifestyle that is gradually ending before our eyes amidst budget cuts. But here as always, by focusing on these losses we miss the opportunity to embrace a new world, settling and exploring it at an early date. Before developing this thought further, let us consider some possible counterarguments by a critic.

2. GRUMPY CAT CRITICAL THEORY

Users of social media (and who does this description exclude anymore?) will probably remember the "Grumpy Cat" meme, with its photo of a calico-tabby house pet with the sourest face imaginable. American President Barack Obama even referenced this meme in at least one speech.¹⁰ The cat in question was named Tardar Sauce, and was born in 2012 in Morristown, Arizona (with its human population of under 200) and deceased in the same place just seven years later. I must admit that I think of this cat whenever encountering the usual crabby academic dismay over the recent emergence of ChatGPT. It is not that I mean to ridicule the alarm felt in my line of work given the constant profusion of new and improved chatbots. The previous rules and standards of our classrooms were devastated instantly by the sudden release of ChatGPT to the public in November 2022. I am probably in the majority of the professoriate

⁹ For a non-vintage copy see M. Heidegger, *Die Kategorien- und Bedeutungslehre des Duns Scotus*, Sothis Press, Wallingford, 2023.

¹⁰ L. Koran, "Obama Compares Republicans to 'Grumpy Cat,'" *CNN* website, <https://www.cnn.com/2015/10/23/politics/grumpy-cat-obama-republicans>, (accessed 23 October 2015).

when I say that currently it is not possible to assign papers to students; at this point, I use a combination of in-class written exams and final oral exams. Although the accrediting agency that oversees the school where I teach does require us to develop the writing skills of our students, I have to be hopeful that they will understand our current need to suspend term papers while awaiting some new consensus about how to judge jointly crafted human/computer writing. I say this not in a spirit of excessive leniency toward the new generation of students, but in the recognition that at some point in the near future, more or less all writing will come from human/computer partnerships. Although this prospect seems as threatening to many academics as atomic warheads mounted on hypersonic missiles, or designer viruses killing off entire ethnic groups, I do not think AI chatbots and their more advanced successors will deserve to be judged by the criteria of weaponry. In the third and final section of this article I will explain why.

But before getting to that topic, let us consider the objections of a recent, entertaining, and widely read critique of ChatGPT: Ronald Purser's dramatically titled "AI is Destroying the University and Learning Itself."¹¹ Purser is currently Lam Larsen Distinguished Research Professor of Management at San Francisco State University, as well as a senior instructor at Dharma College in Berkeley. Despite his appointment in a Management department, a critical take on capitalism is evident from his most-cited publication: a book humorously entitled *McMindfulness*, published by Repeater in London, where I have published a book myself.¹² Purser's article is both polemical and witty; most academic readers will chuckle multiple times as they move through its pages. Yet the article is too reliant on low-hanging fruit, with several canned attacks on our supposedly benighted university administrators, and a smaller number of cheap shots. For instance, the author pointlessly derides Sam Altman as "Little Sammy," which would sound just as petty as Donald Trump's "Little Marco" (Rubio) if not for the fact that Altman stands a robust six feet, three inches tall. An OpenAI employee named Siya Raj Purohit is singled out for Purser's especial scorn, even when she does nothing more

¹¹ R. Purser, "AI is Destroying the University and Learning Itself," *Current Affairs*, <https://www.currentaffairs.org/news/ai-is-destroying-the-university-and-learning-itself>, (accessed 1 December 2025).

¹² R. E. Purser, *McMindfulness: How Mindfulness Became the New Capitalist Spirituality*, Repeater, London 2019; G. Harman, *Dante's Broken Hammer: The Ethics, Aesthetics, and Metaphysics of Love*, Repeater, London, 2016.

heinous than suggesting ways to use ChatGPT to improve the structure of syllabi. “The message was clear,” Purser cackles. “Let ChatGPT redesign your class.” Assuming we all agree that certain uses of chatbots by professors would count as professional fraud, it is hard to see Purohit’s advice as urging us to do anything illicit. Often enough the human mind falls into ruts and routines, and it strikes me as perfectly harmless to ask a chatbot to make suggestions to spice up a class, just as we might ask a human friend to do so. Although I have never used ChatGPT for this particular purpose, I have already found it useful for suggesting class discussion questions about certain assignments. Does this amount to letting a chatbot take over my class? Hardly. I am still the one who distinguishes good suggestions from poor ones, still the one who improves all such suggestions with tailored modifications, and still the one who chooses and prepares the course readings in the first place. One can certainly imagine lazy segments of the professoriate engaging in extensive abuses of the technology, as we now see students doing on a regular basis. But this problem cannot be addressed by banning chatbots, since it would be just as hard to enforce such a ban on professors as on students (as concerns the latter, I have not managed to catch a single instance of cheating since ChatGPT was introduced, though I have no doubt it occurred). Instead, both professors and students will have to be judged the same way as always: by the quality of their work. Admittedly, we are in a moment of academic crisis in which it is difficult to know whether the quality of work is owing to a human or to improper AI assistance. This is why I and others have scrapped term papers for now. Yet it is not at all obvious, despite widespread belief to the contrary, that this problem is a permanent one. At the very least, we should happily be able to expect perfect grammar and spelling from any future student work, even when written by non-native speakers of English, and to require them to reach such perfection “by themselves” may now be as futile and pointless as requiring them to learn Morse code. As for how to determine whether a student understands whatever they submit as academic work, this may require supplementary examination of such work: perhaps in the form of in-person discussion with the professor or (where appropriate) teaching assistants.

Admittedly, a number of Purser’s critiques of the recent direction of our universities seem to hit home. He cites Henry Giroux’s still convincing complaint that “public universities [are] being remade as vocational feeders for private markets,” something that most professors will

have long seen with their own eyes.¹³ Many will nod in agreement with the reference to Benjamin Ginsberg's assault on the "all-administrative university," though I happen to think a certain degree of mythology has exaggerated our disdain for evil corporate administrators.¹⁴ Purser will even convince many when he riffs on the late David Graeber's "bullshit jobs" slogan, speaking by analogy of "bullshit education."¹⁵ Most readers will also be appalled by Purser's recounting of one Chungin "Roy" Lee, who passed cynically through the halls of Columbia University, where he claimed that he and his friends learned "nothing" in their classes (and that he was there only to find a spouse and a business partner), before launching a start-up called Cluely: a website openly devoted to the encouragement and facilitation of cheating. In a famous and still jarring case, low-wage Kenyans were contracted to select and remove the most horrific images from AI databases, leaving some with long-term mental trauma. Journalist Karen Hao was "blackballed" after publishing a critical piece about OpenAI, raising the specter that big-money AI companies are already as sinister as insurance and oil companies are widely believed to be. Purser alarms us further when he repeats well-documented problems with the environmental impact of AI training runs. Perhaps even more unnervingly, Ohio State University has declared that use of AI will no longer count as a violation of academic integrity: a policy that seems, at the least, premature.

Such concerns are clearly worth examining. Many of them call out for rigorous public policy measures. Yet as far as I can tell, Purser offers no credible remedies for the problems he outlines. My point here is not that no one should criticize something for which they have no solution; critique is often the initial step toward a solution. The problem, instead, is something I find completely missing from Purser's critique: namely, ChatGPT is already very good at what it does, and is becoming better and better over time. From reading Purser, one almost gets the sense that the only possible reason to use ChatGPT is to cheat. This is a failure of imagination, one that suggests he has not yet used the technology enough. At the very least, it is the best research assistant you have

¹³ Purser oddly refers to the Wikipedia article on Giroux rather than any publications, but one comprehensive source is C. Robbins (ed), *The Giroux Reader*, Paradigm Press, Boulder, 2007.

¹⁴ B. Ginsberg, *The Fall of the Faculty: The Rise of the All-Administrative University and Why it Matters*, Oxford University Press, Oxford, 2011.

¹⁵ D. Graeber, *Bullshit Jobs: A Theory*, Simon & Schuster, New York: 2018.

ever had, even if its findings still require significant double- and triple-checking review. Just as critics say, it does “hallucinate.” It has credited me with ideas that are not my own, and when asked to rank my books in order of quality, it omitted many of my own, while including books by others as well as utterly non-existent books. But that is only a partial and inadequate picture of the situation. In my own field, philosophy, I have found ChatGPT profoundly useful for a number of tasks that no human assistant could perform in a reasonable amount of time. Just try asking the bot if there is any intellectual connection between two philosophers who seem to have none at all; quite often it will provide useful leads that trigger breakthroughs in one’s own conception of the history of philosophy. If you cannot remember which author made a specific statement, ChatGPT will often remind you in less than a second. Ask it to rank the five most important theories of knowledge of the seventeenth century, and while you may disagree with the rankings, you at least have a quick reminder of the most likely candidates. Certain playful uses of chatbots can also bear intellectual fruit: although ChatGPT is known to give corny responses to requests for jokes, there is an easy workaround simply by asking it to respond to questions in the style of famous comedians. I have learned a great deal by having ChatGPT give comedy routines on the philosophies of, say, David Lewis and Saul Kripke in the voices of such prominent comics as Dennis Miller, Chris Rock, or Richard Pryor. Nothing about this exercise is frivolous. If done properly, it pushes our intelligence in speculative directions that would not have been possible with any previous stimuli.

Rather than looking for ways to restructure our intellectual situation in view of this powerful new technology, Purser prefers to appeal to bland and familiar academic tropes. He insists in anodyne fashion on “faculty input,” which sounds democratic enough, unless you have ever served in university administration and seen for yourself that such input often takes the form of complaint, obstruction, self-righteous shouting, angry demands for more faculty funding, or bizarre conspiracy theories about the motives for university decisions. Purser also ensures that “critical thinking” makes its expected appearance, as if the *primary* role of human thought were to object to whatever it hears. Multiple appeals are made to “resistance,” along with claims that the young student generation stands at the forefront of a promising resistant upsurge. But while today’s world is certainly rife with currents that deserve resistance, the idea that the mission of the mind is to say “no” seems very much a relic

of modern philosophy. That era's famous duality of subject and object wrongly implies to many that unless the subject is purified of any entanglement with objects, the result can only be "alienation." Yet quite often it is the procession of inanimate objects through history, or even the forces of reality unleashed by commercialism (that demon ritually loathed by intellectuals) that force intellectual change more than an ocean full of critical thinking. In recent decades, academics have grown fond of the remark that it is now easier to imagine the end of the world than the end of capitalism. But as I once asked Mark Fisher, was Steve Jobs really less imaginative than Fredric Jameson?¹⁶ What AI gives us the opportunity to question is the model of the intellectual as someone who critiques whatever emerges from beyond the academy, especially when it emerges from the vicinity of wealth. Admirers of Karl Marx too often forget the part where he and Friedrich Engels credit capitalism with annihilating so many dead forms of life.¹⁷ The *primary* lesson of this time of ours, in which an increasing number of universities are financially on the ropes, may not be that tech bros and hedge fund managers are demons prancing in our midst, but that too many intellectuals are almost religiously devoted to a model of critical refusal and opposition that is no longer suited to our task.

Some evidence for this can be found in the specific way that Purser draws a distinction between tools and technology. Whereas tools help us achieve some task, he says, technologies reshape the very environment in which we act. This seems plausible enough. Technologies can also have obvious political effects, some of them clearly deleterious: Purser cites Langdon Winner, who of course wrote a famous article on how bridges were constructed in such a way as to prevent black bus riders from using certain beaches.¹⁸ But naturally, it does not follow that all artifacts are encoded with oppressive politics, or even that oppression and greed can ever be the *primary* effects of technology even when intended, except in certain narrowly defined cases. Whatever the motives of oligarchs or deranged Lex Luthors may be when introducing a given technology, its effects will probably outrun their best-planned schemes. Purser argues further that earlier technologies, unlike AI chatbots, "expanded human

¹⁶ M. Fisher, "Dialogue with Graham Harman," *k-punk* blog, http://k-punk.abstractdynamics.org/archives/2010_01.html, (accessed 6 January 2010).

¹⁷ K. Marx, F. Engels, *The Communist Manifesto*, Penguin, London, 2004.

¹⁸ L. Winner, "Do Artifacts Have Politics?," *Daedalus*, 109, 1, 1980, pp. 121–136.

agency” over a long period of time. It is not clear that either part of this sentence is true. No technology—whether it be the cotton gin, the dishwasher, or the car—can easily be assessed as either adding to or subtracting from human agency. Nor is it the case that such changes occur only over long periods of time, as can be seen in the case of ChatGPT itself. Far from needing a long incubation period prior to its social effects, this technology has induced near-total disorientation in the professoriate in just a few years: witness Purser’s own anecdotes about academic colleagues now contemplating sudden retirement. Indeed, I suspect that Purser made the claim about long periods of time being needed in order to bolster his call for additional faculty consultation and “research.” And anyway—needed *for what*? AI is infiltrating society at a speed that no sleepy professorial assembly will be able to oppose; our ability as a profession to “resist” is really very limited, despite our frequent delusions of political grandeur. We can stop assigning term papers, which I and many others have already done. We might also try to demand that university employers, like Purser’s own San Francisco State, stop making deals with AI companies over the heads of their faculty; this may or may not work, but even if it succeeds it will slow the spread of AI in society not a whit. We can behave like critical thinker Cassandras bemoaning the inevitable downfall of the human intellect. The better option, I think, is to recognize that our typical assumptions about how an intellectual should operate in (and benefit from) society was built under earlier conditions of society. When Purser asks rhetorically “why retain tenured professors at all?,” he apparently aims this barb at corporate sell-outs in university administrations who fail to “resist.” But surely he knows that our profession has reached a perilous economic situation that is, to at least some extent, independent of anyone’s particular good or ill will.

The university was born at a specific point in human history, and will end at a different specific point in human history, and that point may come much earlier than employees of universities would like. But intellectual life existed before the university and will continue to exist after it, even if some particular generation will end up drawing the short straw and inheriting horrible professional luck. On this note, I would like to shift from the possible fate of universities to the likely fate of writing itself.

3. FROM THE HUMAN CREATOR TO THE HUMAN CONNOISSEUR: CHATGPT AND AUTOMATED WRITING

Earlier I cited a passage from Martin Heidegger, who among all serious philosophers is surely one of the grumpiest of cats. Let's review his absurdly pompous passage against typewriters:

It is not accidental that modern man writes "with" the typewriter and "dictates" "into" a machine. The "history" of the kinds of writing is one of the main reasons for the increasing destruction of the word. The latter no longer comes and goes by means of the writing hand, the properly acting hand, but by means of the mechanical forces it releases. The typewriter tears writing from the essential realm of the hand, i.e., the realm of the word. The word itself turns into something "typed."¹⁹

I will not belittle the notion of "essence," which has received too much negative press in recent decades.²⁰ Yet I will not hesitate to ridicule Heidegger's romanticism of the human hand. For it is the very definition of nostalgia to overidentify the essential features of a thing with one particular transient stage through which it passes. It is a fact of human psychology that we tend to appreciate best those moments in the history of music, cinema, television, or sports that happened to coincide with our youth; or if not with our youth, then perhaps with some later pinnacle of our worldly success. There was seldom a greater white devotee of black American music than my father, but this holds only for the music of the 1960s; as for hip-hop, he detests it so much that he falls into a depressive mood even when he hears a few bars accidentally. Nor am I any better when it comes to those areas that interest me most, no matter how hard I try to stay abreast of changing fashions: for me, the conditions of the 1990s (in sports, music, and culture) will always represent what life ought to look like. As briefly mentioned above, nostalgia is exactly what Nietzsche means when he speaks of the "antiquarian" approach to history.²¹ It can be useful to study under certain elderly antiquarian professors, despite their often pedantic reverence for earlier routines, but only because

¹⁹ See footnote no. 3.

²⁰ One place where I develop this point further is in D. Varn, "Marginalia on Radical Thinking: An Interview with Graham Harman," <http://skepoet.wordpress.com/2012/06/01/marginalia-on-radical-thinking-an-interview-with-graham-harman/>, (accessed 1 June 2012).

²¹ F. Nietzsche, *Untimely Meditations*.

they have access to a treasury of facts and anecdotes that the generation just above us has forgotten.

I hope that most readers of this article will agree with me that to treat writing as having an “essential relation to the hand” is absurd. The world has now experienced any number of great writers who never used anything but a typewriter or a word processor. Most of us are willing to take Heidegger’s elderly-sounding Ludditism with more than a grain of salt. My worry, though, is that most are still overly attached to other features of writing with which it may be assumed to have an essential relationship. Namely, I hold that our image of the writer is bound up too closely with the idea of a solitary human being generating sentences *ex nihilo* on a blank screen. This is what we imagine as the ideal existence of a writer. This person closes themselves up in a room and wages a battle with nothingness, drawing words out of thin air in heroic fashion. This may happen in a rapid, caffeine-fueled way, as with Balzac; it may involve long stretches where nothing at all is written as the author pouts face-down on a couch, as with Flaubert. It may involve more eccentric routines, such as Jack Kerouac typing *On the Road* on a single uninterrupted roll of paper.²² Perhaps it needs the assistance of drug trips, or must be preceded (or followed) by marathon sex sessions. The room for autobiographical variation is limitless. But what we think writing *ought* to be always involves the personal generation of words from out of nothing. The human writer is the human creator, a kind of deity in miniature, especially in a country such as France that knows how to revere its writers properly.

Yet this is precisely what I would like to persuade readers to put behind them once and for all: this vision of the “essential relation” of writing with visionary personal experience in the face of blank paper. And ChatGPT, I believe, is the first sign that the role of the writer will change. We all have reasons to regret that we were not born a bit sooner than we were. In my own case, for instance, I regret that I was just 14 months old when the first moon landing occurred, so that I am left with no memory whatsoever of the event. But in other ways, we are all born at exactly the right time, and manage to benefit from it. My own stroke of historical luck is that I began publishing in 2002, fully twenty years before the appearance of ChatGPT. What this means is that nearly everything I have ever written is accessible to most chatbots. Thus I am able to ask them questions about what I might have done differently in certain books to

²² J. Kerouac, *On the Road*, Penguin, London, 1960.

have made them better; what subjects I have not written about that I really ought to write about; what aspects of my personality come through most clearly in my books, and which past authors I would likely have gotten along with best or worst.

Nonetheless, what I really look forward to in the coming decades is something well beyond such opportunities. What I would like to ask some future successor of ChatGPT is the following: “please write 100 draft manuscripts on [whatever topic] in the style of my previous books.” Now, from the standpoint of the present day, this looks like rampant cheating. That is why I will not even think of doing it until many others have done it before me; I do not have the temperament of an early adopter, or that of someone who enjoys provoking controversy and fighting through it. Such an episode would leave me drained of energy and sad. Even so, does it not indicate a sort of secret cheater’s mentality that I would even fantasize about AI doing all of the writing “for me,” so that I did nothing “by myself”? This is where my earlier point becomes relevant. Why do we assume that writing “by oneself” requires that we endorse an “essential relation” between writing and sitting face-to-face with a piece of blank paper? Did I not create the style and content of my previous books, written in a pre-AI universe? Would it really be “cheating” to ask AI to generate further products written in the same spirit, and to take ownership for myself? Imagine that I asked the more powerful successor of ChatGPT to generate one hundred book drafts about a subject that interests me, but about which I have never written a thing: say, climate politics. Is the world really best served by my spending three years in a library, followed by six months of torment trying to summon exactly the right words from the void? This is what we now insist a writer must do, and I must admit that I have enjoyed it, and will probably continue to enjoy it for a while longer. Yet it is increasingly clear that however much someone might try to romanticize the life of a writer as we currently know it, this life is probably not the best use of a writer’s time. The blank page is not only frightening; it can also be a waste of time in the name of heroic *ex nihilo* creation. Would it not be objectively better to work with AI-generated drafts, shaping and editing them on the basis of lifelong intellectual experience?

What I am trying to say is that what we romanticize most is the idea of the human *creator*, perhaps because we have been taught to view ourselves as made in the image of God. But what if that is not really what we are, at bottom? What I want to suggest is that human beings are at their

best not as creators, but as *connoisseurs*. As a rule, the connoisseur is seen as someone respectable, but not quite at the level of the creator. Dr. Johnson is a literary giant, but pales in the company of Shakespeare; the same for Clement Greenberg by comparison with Jackson Pollock, at least among those who admire Pollock. But what if the ability to create something from nothing was simply an unfortunate necessity for humans in the pre-AI age, when there was nothing else to do our creating for us. Even the creative artist has no greater skill than the ability to distinguish own their best work from their less-than-best work. And so what if some future human, entirely unskilled in executing works on canvas with oil or acrylic, nonetheless had the most immaculate taste in how to ask AI to do it? I once had a professor whose own lack of respect for hip-hop stemmed from the fact that they merely “sampled” rather than playing musical instruments with their own hands. But perhaps this was just an unknowing relapse into Heidegger’s claim that writing has an essential relation to the hand. The thesis I venture here is that our new essential relation is not to the creative process, but to connoisseurship.

REFERENCES

- Coeckelbergh, Marck, David J. Gunkel (2025), *Communicative AI: A Critical Introduction to Large Language Models*, Cambridge: Polity Press.
- Fisher, Mark (2010), “Dialogue with Graham Harman,” *k-punk* blog, http://k-punk.abstractdynamics.org/archives/2010_01.html, (accessed 6 January 2010).
- Ginsberg, Benjamin (2011), *The Fall of the Faculty: The Rise of the All-Administrative University and Why it Matters*, Oxford: Oxford University Press.
- Graeber, David (2018), *Bullshit Jobs: A Theory*. New York: Simon & Schuster.
- Harman, Graham (2016), *Dante’s Broken Hammer: The Ethics, Aesthetics, and Metaphysics of Love*, London: Repeater.
- Harman, Graham (2023), “The Shipwreck of Theseus: Philosophy and Maritime Archaeology,” in Sara A. Rich, Peter B. Campbell (eds.), *Contemporary Philosophy for Maritime Archaeology: Flat Ontologies, Oceanic Thought, and the Anthropocene*, Leiden: Sidestone Press, pp. 57–73.
- Heidegger, Martin (2023). *Die Kategorien- und Bedeutungslehre des Duns Scotus*, Wallingford, UK: Sothis Press.
- Heidegger, Martin (1992). *Parmenides*, trans. André Schuwer, Richard Rojcewicz, Bloomington: Indiana University Press.
- Heidegger, Martin (2024), *Ponderings II–VI: Black Notebooks, 1931–1938*, trans. Richard Rojcewicz, Bloomington: Indiana University Press.
- Heidegger, Martin (1977), *The Question Concerning Technology and Other Essays*, trans. William Lovittm, New York: Harper & Row.
- Kerouac, Jack (1960), *On the Road*, London: Penguin.

- Koran, Laura (2015) "Obama Compares Republicans to 'Grumpy Cat,'" *CNN* website, <https://www.cnn.com/2015/10/23/politics/grumpy-cat-obama-republicans>, (accessed 23 October 2015).
- Marx, Karl, Friedrich Engels (2004), *The Communist Manifesto*, trans. Samuel Moore, London: Penguin.
- McLuhan, Marshall, Eric McLuhan (1992), *Laws of Media: The New Science*, Toronto: University of Toronto Press.
- McLuhan, Marshall, Wilfred Watson (1970), *From Cliché to Archetype*, New York: Penguin.
- Nietzsche, Friedrich (1984), *Untimely Meditations*, trans. R. J. Hollingdale, Cambridge: Cambridge University Press.
- Polt, Richard (2015), *The Typewriter Revolution: A Typist's Companion for the 21st Century*, New York: Countryman Press.
- Plato (1995), *Phaedrus*, trans. Alexander Nehamas, Indianapolis: Hackett.
- Purser, Ronald (2025). "AI is Destroying the University and Learning Itself," *Current Affairs*, <https://www.currentaffairs.org/news/ai-is-destroying-the-university-and-learning-itself>, (accessed 1 December 2025).
- Purser, Ronald (2019), *McMindfulness: How Mindfulness Became the New Capitalist Spirituality*, London: Repeater.
- Robbins, Christopher (ed.) (2006), *The Giroux Reader*, Boulder: Paradigm Press.
- Varn, C. Derick (2012). "Marginalia on Radical Thinking: An Interview with Graham Harman," <http://skepoet.wordpress.com/2012/06/01/marginalia-on-radical-thinking-an-interview-with-graham-harman/>, (accessed 1 June 2012).
- Winner, Langdon (1980), "Do Artifacts Have Politics?", *Daedalus*, 109, 1, pp. 121–136.

Andrew Witt*

ARCHITECTURE, AUTOMATA, AND ARTIFICIAL MEMORY

ABSTRACT: Artificial memory is an essential precondition for automata, one with a rich collection of conceptual and technical precedents. This paper frames artificial memory as both a technical media and a mode of spatial experiment, arguing that the intersection of mechanical and spatial invention catalyzed early experiments in automaton memory. More specifically, seventeenth century architects played pivotal and surprising roles in the development of artificial “memory architectures” through their work in the creation of large-scale reprogrammable musical automata. In these early, perhaps original, forms of automata memory, we see not only the provocative roots of autonomous machines but also a powerful and expansive view of the architect as transdisciplinary inventor.

KEYWORDS: automata, architecture, memory, organs, Salomon de Caus

* Andrew Witt: Graduate School of Design, Harvard University, Cambridge, Mass.; Senseable City Lab, Massachusetts Institute of Technology, Cambridge, Mass.; andrew_witt@harvard.edu.

This is an Open Access article under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not transformed in any way and is properly cited.

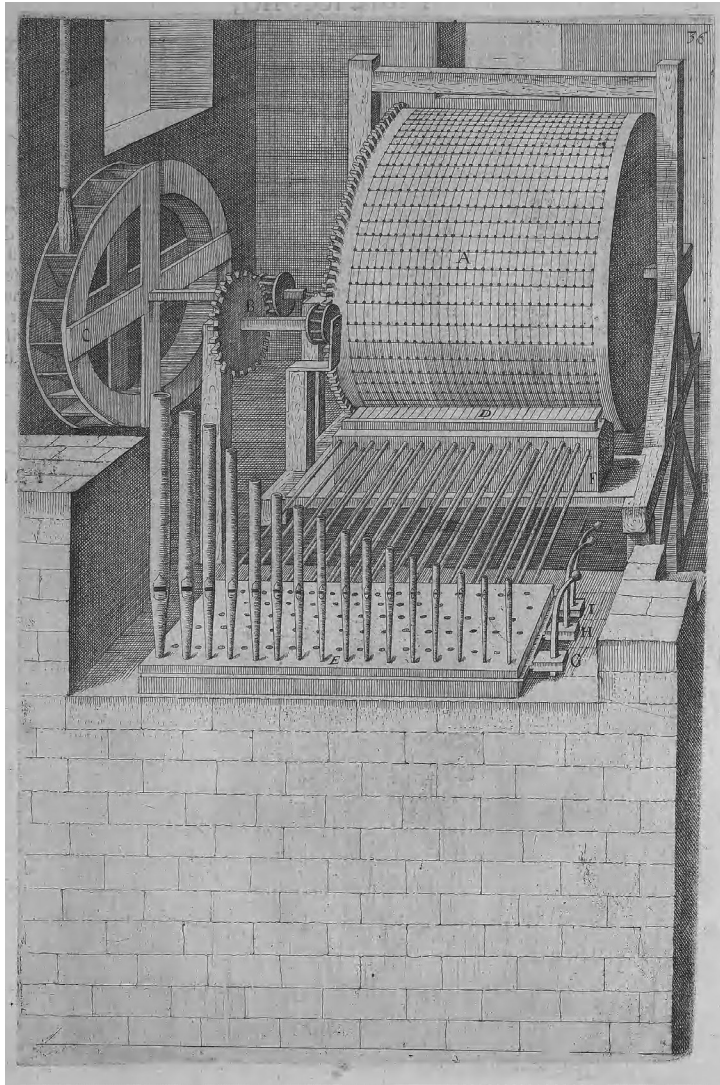


Figure 1. Salomon de Caus's Problem XXVIII in his First Book is "Machine by which one would make the sound of playing the Organ, by means of water" [*Machine par laquelle l'on fera sonner un jeu d'Orgues par le moyen de l'eau*]. Source: S. de Caus, *Les Raisons des Forces Mouvantes, Avec Diverse Machines t'ant Utiles que Plaisantes*, Jan Norton, Frankfurt, 1615.

When did machines begin to remember? When did they gain a language of instruction, inscrutable to human eyes, and the power to execute instructions, unbound to human hands? More than poetic musings, these questions hunt for the moment when machines, through memory, subtly shifted from inert objects towards lively and even intelligent beings. Without memory, an automaton is condemned to a Sisyphean repetition that is only a meager shadow of action. Memory—and particularly the potential of a machine to be reconfigurable, reprogrammable, and adaptive—is an essential foundation for more advanced automation such as autonomous robotics and artificial intelligence. When did memory become mechanized?

Today, “memory architecture” often evokes the chip-scale technical apparatus for storing digital data and code for computational processing. For architects, artificial memory comprises an array of technologies and media that shape creative practice through their modulation of recorded data and information. Yet architects have also long understood memory as amenable to its own artifice spatial typologies and experiments—libraries, museums, memory palaces, memory theatres, and beyond—that spatialize organizational regimes of memory. Artificial memory thus implicates a set of technologies on the one hand and a set of spaces on the other.

In certain extraordinary circumstances, artificial memory can be both a technology and a space simultaneously. Among the earliest formats of machine mnemotechnology were just such hybrids: architectural-scale mechanical contrivances that orchestrated the technological behavior of sophisticated automata. More specifically, I argue that a particular seventeenth century programmable automaton, the water organ, was one of the earliest instances of reconfigurable machine memory, and that its creation drew on the fluid and expansive disciplinarity of seventeenth century architecture. As they advanced in Renaissance Europe, musical automata were a locus of surprising encounters between a mathematically-inclined architects and experimental automation technologies. The scope of renaissance architecture that encompassed hydraulic proto-engineering on the one hand and an intimate understanding of geometry on the other allowed architects to contribute substantial innovations to this field. Moreover, the mathematical conventions of notation, knowledge organization, and memory offered a formative bridge between architecture and automation in the high Renaissance. Thus, architects at this pivotal moment drew on an unusual constellation of skills spanning

spatial design, mathematics, and engineering to make machine memory by architectonic means.

The memory of these programmable automata was arguably one of the first media intended to be read by a non-human entity. Machine readability requires the encoding of information in a format optimized for machinic processing, instead of human legibility. This encapsulation of instructions or protocols in machine-readable media make sophisticated processes portable and recombinant.¹ In effect, this encoding establishes a peculiar kind of language that allows the encoded content to vary and yet remain machine readable. The origin of machine memory is thus intimately tied to how instructions and protocols are embodied in machine readable media.

Far from an idiosyncratic historical curiosity, the question of how memory is both encoded and spatialized is foundational to concepts of machine intelligence. A reconstruction of the origins of machine memory can reveal insights into the singular role of memory in the construction of intelligent machines, as well as the peculiar role that architects have played in imagining artificial and automatic memory.

MEMORY TECHNOLOGIES, MEMORY SPACES

Architects have tended to frame artificial memory either as a technological medium or as a particular kind of experiential space. On the one hand, as technologies of recording and organizing information—that is, of creating artificial memory—have advanced, the aims and practices of human remembering have transformed. Historian Ann Blair has suggested that “what we take to be human capacities, say, of memory and recollection, change over time under the impact both of cultural expectations and the technologies with which we operate.”² Technologies of memory shape practices of memory. In an architectural context, historians and theorists point to the essential ways in which the media of recording and artificial memory—from writing to digitization—have augmented and altered the practice of design. In the modern period, Mario Carpo cites Leon Battista Alberti’s use of recording formats of texts and tables as a way of “preserve the fidelity to the original of the copies of images he valued, used,

¹ A. Witt, *Formulations: Architecture, Mathematics, Culture*, MIT Press, Cambridge, Mass., 2022.

² A. Blair, *Too Much to Know: Managing Scholarly Information Before the Modern Age*, Yale University Press, New Haven, 2010, p. 3.

and needed.”³ The Albertian role of artificial memory was to ensure exact reproducibility and endless identity, in effect to automate repetition.

More recently, advanced technical media of memory have allowed machines themselves to remember, raising the question of the relationship between human and machine memory. For a machine to remember is really for it to scan a recorded memory that is machine readable. Looking to the dawn of the computational age, Zeynep Çelik Alexander observes that “technologists working on the problem of scanning took for granted that the storage of sensory stimuli in the human nervous system and information in a machine were analogous”—that human and machine memory had some fundamental parallels.⁴ Yet the isomorphism of human and machine memory is not a given, and in fact early experiments would venture forms of machine memory quite different from our own.

Beyond the array of technologies it may implicate, the concept of artificial memory evokes for the architect ways in which regimes of memory might become spatial. An early example of such spatial memory regimes was Giulio Camillo’s theater of memory, an imagined space in which to behold the totality of human knowledge. Articulated in his posthumously published 1550 work *L’idea del teatro*, Camillo proposed a curious room with a tiered system of seven semi-circular racks on which were arrayed dense sequences of discrete images each evoking a distinct domain of knowledge. As historian Lu Beery Wenneker observed, “the images should be striking and out of the ordinary, extremely beautiful or extremely ugly, gory or comic, but above all, distinctive.”⁵ These images comprised a kind of panoramic collage of knowledge or memory, apprehended in spatial terms. The theatre was an atlas of idea-images unfurled onto the walls of an immersive space.

Architects have long mused on the architecturalization of Camillo’s curious speculation and adjacent memory spaces. Daniel Libeskind, for instance, attempted to construct a machine that implemented *L’idea del teatro*, a project for which Camillo himself left no drawings. For the 1985 Venice Biennale, Libeskind presented a triad of writing, reading, and memory machines that he called “Three Lessons in Architecture”

³ M. Carpo, *The Alphabet and the Algorithm*, MIT Press, Cambridge, 2011, p. 54.

⁴ J. May, Z. Ç. Alexander, *Design Technics: Archaeologies of Architectural Practice*, University of Minnesota Press, Minneapolis, 2020.

⁵ L. Beery Wenneker, *An Examination of L’idea del teatro of Giulio Camillo, Including an Annotated Translation, with Special Attention to his influence on Emblem Literature and Iconography*, PhD Diss., University of Pittsburgh, 1970, p. 4.

produced under his direction at the Cranbrook Academy of Art. The memory machine reframes Camillo's theatre as a room-scale wooden frame with a vertiginous menagerie of documents, images, and objects suspended above and around the occupant, animated by an intricate set of pulleys, wheels, and gears. The effect is that of a scaffolded and deconstructed music box. A hybrid machine-space-furniture, the Memory Machine articulates mechanism as a kind of tectonic, though Libeskind argues "the subversive element of mechanics (Teatro Farnese versus Teatro Olimpico) enters only in those places where memory has succeeded in retaining its surrogate authority."⁶

Libeskind's project show that what appears *prima facie* as a dichotomous distinction between memory technologies and memory spaces actually invites the creation of hybrid architectural memory machines. In fact, it is this possibility of a memory machine at an architectural scale that suggests a dialog between architectural and computational modes of creation, with the notion of artificial memory a generative catalyst.

PHILOSOPHIES OF ARTIFICIAL MEMORY

Beyond their architectural implications, philosophical investigations into memory have attended primarily to the status and functions of human memory, with occasional exotic forays into the space of animal consciousness. Indeed, to the extent artificial memory is considered philosophically, it has often been seen primarily in ambivalent or even malign relation to human memory. As an apparatus for the externalization of individual and collective human memory, artificial memory can be accused, perversely, of being a catalyst for human forgetting. French philosopher Bernard Stiegler, one of the foremost theorists of technological memory, argued this position to a certain extent: that "technicization is what produces loss of memory" through this process of externalization.⁷ For Stiegler this process of externalizing memory, as well as a range of other biological functions, is what most authentically characterizes the fundamental trajectory of human history.⁸ As Stiegler says, it is a trajectory of reciprocal creation, of "The technical inventing the human, the human

⁶ D. Libeskind, *The Space of Encounter*, Universe, New York 2001, p. 180.

⁷ B. Stiegler, *Technics and Time, 1: The Fault of Epimetheus*, trans. Richard Beardsworth and George Collins, Stanford: Stanford University Press, 1998, p. 3.

⁸ B. Roberts, "Cinema as mnemotechnics," *Angelaki: Journal of Theoretical Humanities*, 11, 1, 2006, p. 56.

inventing the technical.”⁹ Nevertheless, even for Stiegler, machine memory is to be understood primarily in relation to human capacities and, ultimately, existence.

To properly understand the role of machine memory in the longer trajectory of artificial intelligence, we must attend to machine memory not primarily in relation to human faculties but rather as an independent dimension of machine autonomy. In that respect, the distinctions from the philosophical analysis of human memory prove surprisingly rich when transposed to our understanding of machine memory.

In their influential 1966 paper “Remembering,” philosophers C. B. Martin and Max Deutscher propose a particularly useful tripartite taxonomy of memory: memory of experiences, of information, and of methods (how-to).¹⁰ Though they apply this schema to examine philosophical questions about when a particular person may truly be said to remember, the schema itself is quite useful apart from this anthropocentric application to taxonomize distinct types of memory, human or machine. Martin and Deutscher’s account is also helpful for calling attention to the distinct preconditions for each type of memory. Remembering demands not only the capacity for memory itself but also a range of other capacities that furnish the inputs and outputs for the act of remembering. For instance, a prerequisite for the memory of experiences is the capacity to have experiences, more particularly a capacity for sensation and perception, as well as the capacity to recall said sensations. This capacity would seem to require some form of embodiment. Likewise, the memory of information requires the capacity to store and recall information semantically; that is, it requires some mode of interaction with language. Finally, memory of methods requires a capacity to recall and implement sequential actions. Yet the faculties that each form of memory presupposes are essentially orthogonal. Memory of methods, for example, does not per force require sensation, perception, or an ability to process language. It is possible, then, to disaggregate these distinct types of memory and consider them separately.

Martin and Deutscher’s tripartite schema is particularly productive if transposed from humans to automata. With it, we gain useful functional differentiations among formats and uses of machine memory, as well as

⁹ B. Stiegler, *Technics and Time*, 1, p. 137.

¹⁰ C. B. Martin, M. Deutscher, “Remembering,” *The Philosophical Review*, 75, 2, 1966, p. 161.

an implied hierarchy of sophistication among distinct types of artificial intelligence. For example, memory of information is certainly a function of today's artificially intelligent large language models but also arguably a capacity of earlier information retrieval technologies. Memory of experiences would seem to require some aspect of sensation, implying that certain robots with appropriate sensor arrays may have access to this type of memory but not, for example, AI chatbots. This hierarchy of memory types has some historical relevance and illuminates developmental distinctions in the function of automaton memory, since the forms of machine memory were not developed in the same moment. Indeed, their evolution extends across centuries, arguably beginning with memory of methods, followed by memory of information, and finally memory of experiences, which depends on sensors that can electronically survey the environment.

One may argue that Martin and Deutscher's third category, machinic memory of methods, was first achieved by instruments that play back variable instruction sequences in a way that allowed the dynamic operation of the attached machine. I would argue that this specific type of machinic memory emerged in a particular form: early mechanical encodings of musical automata scores.

WATER, ARCHITECTURE, AUTOMATA

Techniques to record information in machine-readable formats developed iteratively over centuries through symbiotic innovations in both notation and media of encoding. Musical automata were significant catalysts for the evolution of automaton mnemotechnologies, as they built on accepted, discretized, and easily serializable musical notation. The scripting of a musical score, and implied the use of a variable notation that, while not precisely linguistic, imposed a definite syntax and structure on encoded musical instructions. The water organ, whose origins date to antiquity, was among the most sophisticated musical automata. It also was among the earliest automata to integrate encoded instruction sequences, and thus arguably ranks among the first machine-readable media. Water organs of the high renaissance also integrated variable programming with the possibility of recoding and reprogramming behavior at will.

Since antiquity, water has been a literal wellspring of automation technologies, a perpetual inspiration for both musical inventors and

architects. The tenth book of Vitruvius (ca. 75 BC–15 BC), specifically devoted to machines, considers water-moving and water-powered machines extensively, including the water wheels (Vitr. X, 5),¹¹ water pumps (X, 7),¹² and the hydraulis, a precursor to the water organ (X, 8).¹³ The hydraulis was a small pipe organ, the first instrument to employ a keyboard, and the beginning of a long iterative development of water-powered musicality. Most of these instruments were much smaller than contemporary organs, many being compact enough to be portable. In the Middle Ages, the Banu Musa brothers of Baghdad (Muhammad, ca. 803-873; Ahmad, d. ninth century; and Al-Hasan d. ninth century) developed a flute that automatically played under power of water. Recorded in their codex *Kitab al-Hiyal al-Naficah* (*The Book of Ingenious Devices*) the Banu Musa flute was among the first machines that employed a pinned drum for the sequential control of the flute. The brothers, who described the device as “the instrument that plays itself,” designed the drums to be swapped out to play a different song. Despite the evident benefits of this novel device, the application of this pinned drum to other instruments in the West took centuries further to be fully appreciated and leveraged.

Despite centuries of development, water organs that could play scores automatically and without human intervention were of a more recent vintage, only fully emerged in the late sixteenth and early seventeenth centuries. The development of water organs during this period, though not exclusively an architectural endeavor, was significantly advanced by architects who, in the Vitruvian tradition, saw infrastructural water works as an integral aspect of their discipline. Indeed, in the early renaissance architecture and hydraulic engineering had not yet diverged, and no less an architect than Leon Battista Alberti (1404–1472) attested to the importance of water in *De Re Aedificatoria*. He, like Vitruvius, devotes a significant portion of his tenth book to the uses and control of water. Aqueducts, Archimedean screws, canals, waterwheels, and dams all fell under the purview of the broadly practicing architect. Historian Pablo Gumiel Campos also observes that Alberti expands on the hydraulic engineering techniques propagated from Vitruvius, including piped

¹¹ Vitruvius, *The Ten Books on Architecture*, Harvard University Press, Cambridge, Mass., 1914, p. 294.

¹² *Ibid.*, p. 297.

¹³ *Ibid.*, p. 299.

distribution.¹⁴ Thus, though their expertise in hydraulic works, architects were well placed to imagine alternate applications of water power.

The late sixteenth century saw a convergence of developments that served to significant advance the sophistication of water automata in general and water organs in particular. This period was a time of exceptionally refined hydraulic works across the European continent, enabled in part by the advent of more sophisticated and powerful hydraulic pump designs.¹⁵ Beyond technical innovations, designers themselves were refining more sophisticated, and quasi-scientific methods in the elaboration of their work. This was encouraged by a deeper familiarity with precedent sources of antiquity, as well as more recent developments in the Islamic world. This wider diffusion of the hydraulic inventions of antiquity contributed to a culture of contemporary innovation in the sixteenth century.

The most significant source from antiquity was Hero of Alexandria (c. 150 AD), the ancient Greek polymath who was arguably the godfather of automata in the West. Hero's work was preserved through several treatises on a range of subjects including *Automata*, *Pneumatica*, *Dioptra*, *Metrica*, *Artillery Construction*, and *Catoptrics*. Still others are attributed to him, uncertainly. From the perspective of automata, Hero's two most salient works were the *Automata* and *Pneumatica*. *Pneumatica*, in particular, proved widely influential and terms of both content and method. Hero framed his work on mechanical invention in quasi-mathematical terms, enumerating theorems and problems in the manner of Euclid's *Elements*.¹⁶ In fact the practice of organizing mechanical research as a series of theorems predated Hero, as Ctesibius of Alexandria (c.300-230 BCE) compiled his own *Pneumatic Theorems*.¹⁷ *Pneumatica* furnished an extraordinary catalog of pneumatic devices that captured the imagination of later generations and secured his place in the pantheon of mechanical inventors. Intriguingly, Hero concluded *Pneumatica* with a description of the construction of a water organ.¹⁸ This organ proved a crucial seed in the later development of musical automata.

¹⁴ P. Gumiel Campos, *Aqua In Palatio: Water in 16th-century European palaces, The Iberian Peninsula as a case study* Vol. 1, (PhD Diss., 2024), 60.

¹⁵ *Ibid.*, 57.

¹⁶ K. Tjebjerg, "Hero of Alexandria's Mechanical Geometry," *Apeiron*, 37, 4, 2004, p. 31.

¹⁷ G. Francesco, *Hero of Alexandria's Automata: A Critical Edition and Translation, Including a Commentary on Book One*. PhD thesis, 2019, p. lxxxvi.

¹⁸ Hero of Alexandria, *The Pneumatics of Hero of Alexandria, from the original Greek*, Taylor, Walton and Maberly, London, 1851, p. 105.

Architects of the sixteenth century Renaissance took a keen interest in Hero's work and contributed to broader experimentation on the principles and methods he espoused. This included rendering Hero's key texts more broadly accessible. In 1589, Italian architect Giovanni Battista Aleotti published a translation of Hero's *Pneumatica* in Italian, the first in a vernacular tongue.¹⁹ Historian A. G. Keller noted that Aleotti had "that taste for the elaborate and the ingeniously cunning in which his age so much delighted."²⁰ This cultural attraction to ingenuity led to a fascination with Hero's work that peaked around 1580. Aleotti's translation was preceded by Federico Commandino's 1575 posthumous publication of a Latin translation of *Pneumatics*.²¹ In fact, Oreste Vannucci Biringucci had prepared an Italian translation by 1582, but it was never published due to his untimely death.²² Aleotti's translation demonstrated the extent to which architects were significant catalysts in the emerging field of pneumatic and hydraulic automation.

The late sixteenth century also saw ambitious built works that drew on Hero's techniques and in some cases advanced well beyond them. Among the most substantial was an extraordinary water organ at the Villa d'Este in Tivoli that could produce automated music. Though not easily reprogrammable, this organ, constructed in the late 1560s at the direction of Cardinal Ippolito II d'Este, had an unusually sophisticated hydraulic mechanism.²³ Developed by the engineers Claudio Venardo and Luca Clerico, the water organ was also built at an unprecedented scale.²⁴ Historians Simone M. Kaiser and Matteo Valleriani note that the organ's automated repertoire was fixed and limited, and that the "melody was very short, being composed of only 17 notes."²⁵ Even this modest feat proved a prodigious marvel to visitors, including Pope Gregory XII. The wonders of this organ indicated a wide latitude of innovation that was yet to come.

¹⁹ A. G. Keller, "Pneumatics, Automata, and the Vacuum in the Work of Giambattista Aleotti," *The British Journal for the History of Science*, 3, 12, 1967, p. 339.

²⁰ *Ibid.*, p. 342.

²¹ S. M. Kaiser and M. Valleriani, "The Organ of the Villa d'Este in Tivoli and the Standards of Pneumatic Engineering in the Renaissance," H. Fischer, V. R. Remmert, J. Wolschke-Bulmahn (eds.), *Gardens, Knowledge and the Sciences in the Early Modern Period*, Birkhäuser, Berlin, 2016, p. 85.

²² *Ibid.*

²³ *Ibid.*, p. 78.

²⁴ *Ibid.*, p. 82.

²⁵ *Ibid.*, p. 89.

By the early seventeenth century, the status of automatons themselves underwent a reputational rehabilitation of sorts. For at least the century prior, automata and their making were held in some suspicion, perhaps because they seemed to take on human faculties mechanically. Yet historian Alexander Marr argues that by the 1620s, “wonder-working” of automatons had emerged from the shadow of occult or magical associations to be understood more broadly as a mathematically informed enterprise of determined rationality.²⁶ Yet Marr notes that even well into the sixteenth century, automaton inventors could be condemned for sorcery, as was the case with English mathematician and polymath John Dee.²⁷ The more mathematical the articulation of autonomic theories, the more the automata were rehabilitated from the crepuscular realms of occult magic into the enlightened realm of the sciences. Though the creation of automata was still something of a liminal practice hovering between the practical and the fantastic, it was beginning to gain a measure of rigor and legitimacy.

Despite their dubious reputation, the techniques of automatic music performance were, by the late sixteenth century, well established in the West thanks not only to the musical automata that appeared in the gardens of the nobility but also to the various clockworks and music boxes that populated the *Kunstkammern* of European courts. By the early seventeenth century, the most distinguished and charismatic of these collections already included fairly advanced playback mechanisms such as pinned drums. These devices, derived from the Banu Musa flute and similar mechanisms, used a small metal cylinder with pins protruding radially. When the drum rotated, each pin would pluck one of an array of hammers that struck a bell, plucked a string, or similarly actuated a sonic response. The Habsburg *Kunstammer* of Rudolf II (1552–1612), perhaps the preeminent collection of its kind, featured such intricate pinned drum mechanisms by 1610.

In certain respects, musical automata were a salutary crucible for the development of mechanical memory. The need to record and perform musical scores offered a ready system for the encoding, and the expectations for a certain quality of performance encouraged exacting standards to the playback mechanism of these devices. The courtly interest

²⁶ A. Marr, “Understanding Automata in the Late Renaissance.” *Le Journal de la Renaissance*, 2, 1, 2004, p. 207.

²⁷ *Ibid.*, p. 208.

in automata also provided ample financing and incentive to experiment in this unusual medium. Moreover, music performance was a highly discrete task that nevertheless attracted enthusiasm and curiosity. Ultimately, this confluence of factors, intersecting with parallel developments in hydraulic engineering, led to some of the most significant advancements in machine memory during the seventeenth century.

REPROGRAMMABLE MEMORY: SALOMON DE CAUS

At this propitious moment of the scientization of automata, the French architect, inventor, and hydraulic engineer Salomon de Caus developed some of the most ambitious and unusual water organs, devices that drew on spatial and engineering methods to articulate a system of reconfigurable machine memory. In de Caus's work we see a case study of a multidisciplinary inventor leveraging architecture, hydraulic engineering, and the machinery of automation to build devices of considerable ingenuity. Across a career that ranged from England to Germany, de Caus achieved several striking works that bridged engineering and architecture, perhaps none more ambitious than the designs for a vast complex of gardens for Elector Palatine Frederick V. The palatial garden, known as the Hortus Palatinus (Figure 2), was planned to be enlivened by fountains, automata, and other water-driven delights. The Gardens, never fully completed and ruined during the thirty years war, nevertheless attested to de Caus's facility with mechanical invention in the service of design. With the Hortus Palatinus the environment itself became an automaton, a new fusion of mechanism, landscape, and architecture.

More than an accomplished designer and imaginative inventor, de Caus was a gifted systematizer who advanced hydraulic architecture through his significant theoretical work. Most notably, his singular 1615 treatise in hydraulic automation, *Les Raisons des Forces Mouvantes avec diverses machines tant utiles que plaisantes* (Figure 3), is one of the seminal works of the field and arguably the defining text of its kind in the early seventeenth century. The presentation of principles and devices is rigorously organized, with a mathematical style of thinking and presentation pervading the text. De Caus begins with a series of primary definitions, including those for fire, air, water and earth, followed by 18 deductive theorems characterizing the behavior of various machines and assemblies. The method is exactly reminiscent of the way that Euclid built a geometry of deductive reasoning derived from elemental premises. The

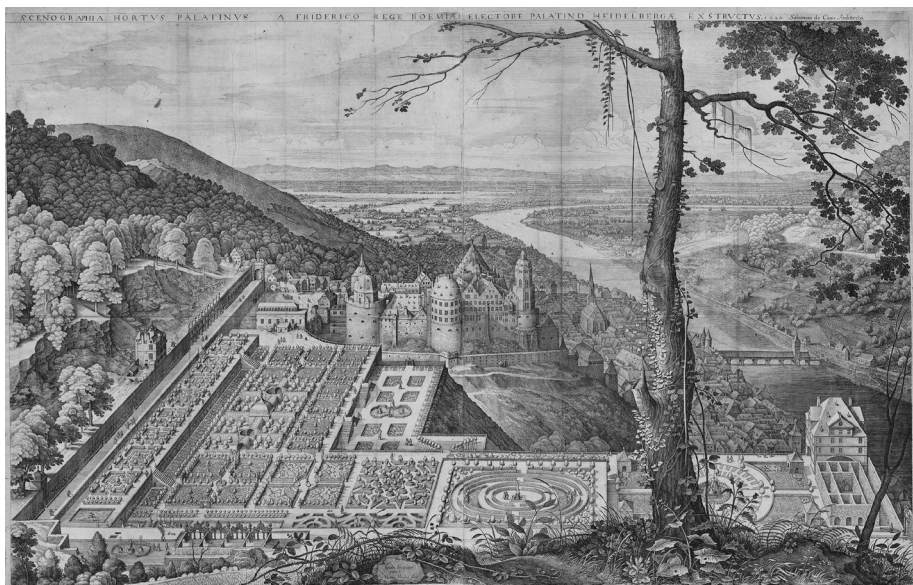


Figure 2. A perspective of the Hortus Palatinus as planned.

Source: S. de Caus, *Hortus Palatinus: A Friderico Rege Boemiae Electore Palatino Heidelbergae Exstructus*, Bry, Frankfurt, 1620.



Figure 3. Frontispiece of de Caus's *Les Raisons des Forces Mouvantes*. Source: S. de Caus, *Les Raisons des Forces Mouvantes, Avec Diverse Machines t'ant Utiles que Plaisantes*, Jan Norton, Frankfurt, 1615.

larger text is divided into three books, each consisting of 35, 20, and 13 “problems,” which were in fact pretexts for ingenious mechanical contrivances ranging including water-driven pneumatic bellows to intricate fountains.²⁸ These problems draw on the earlier theorems and present increasingly sophisticated machines as the solutions to the stated problems. These theorems thus offered the generative foundation for an extraordinary range of machinery, as well as aligning de Caus’s project tacitly with conventions of mathematical expression and knowledge organization.

Though they collectively deal with the design of mechanized hydraulics works, the three books address different topics. The first book ranges widely in an examination of diverse methods of moving and pumping water, mechanisms of water clocks, techniques of fountain design, water-driven wood fabrication tools, and water organs, including descriptions of de Caus’s method for storing and reading scores. The second book is devoted particularly to the architecture of fountains, grottoes, and follies that might house delightful automata. The third book delves more fully into organs and organ-making, including the fabrication and proportion of pipes, the use of bellows, and the design of water wheels. Throughout, de Caus furnishes a series of marvelous plates that bring his inventions to life with exquisite detail.

Among its many innovations, *Les Raisons des Forces Mouvantes* elaborated de Caus’s technique of machine memory: a reprogrammable pinned drum (Figure 4). Through his text and a series of marvelous analytic drawings, de Caus describes a large cylinder (“roue musicale”) with a horizontal axis furnished with a grid of holes to which various rectangular pegs could be affixed.²⁹ These pins would strike keys of a keyboard or other mechanical relay as the cylinder rotated, thereby opening the stops of associated organ pipes (Figure 5). The various lengths of the pins corresponded to different durations of activation for the connected pipes, and some were cut to fine points to allow instantaneous plucking of the keys. In order to allow dynamic reconfiguration, the cylinder was perforated with a series of holes that allowed the pegs to be “change the pegs when one would like to change the song.”³⁰ Thus unlike the pinned drums of the Banu Musa and other predecessors, de Caus’s drum was

²⁸ S. De Caus, *Les Raisons des Forces Mouvantes, Avec Diverse Machines t’ant Utiles que Plaisantes*, Jan Norton, Frankfurt, 1615.

²⁹ Probleme XXV, *ibid.*

³⁰ [...] “changer les cheuilles quand l’on voudra changer de chanson.” Probleme XXX, *ibid.*

reprogrammable at the level of the individual notes, allowing for editing and modification without the wholesale replacement of the drum. Arguably, this was one of the first instances of reprogrammable machine memory. Drawings of this apparatus indicate 30 notes, a narrower range than the contemporary harpsichord but nevertheless one that afforded broad combinatorial possibilities in musical performance. Moreover, this type of reconfiguration—the rearrangement of pegs in a large drum—required that the pegs themselves should be of sufficient size to be manually manipulable. Thus, to allow for reconfiguration, de Caus's device necessarily approached the architectural scale of a large room or even a building. In a way, reconfigurable machine memory invited architectural thinking.

Two aspects particularly distinguish de Caus's organ from its predecessors. First, the scale of his proposed water organs was substantial, taking on architectural dimensions. Although the exact size of these devices is not explicitly stated, tacit indications such as the inferred dimensions of the drawn brick enclosure confirm that this is a proposal at a building scale.³¹ Moreover water organs required larger and more robust mechanisms than the delicate devices of small courtly automata, so a larger scale is in keeping with those demands. To a degree, De Caus was designing public automata, or at least automata that were housed in their own building enclosures and thus had a different kind of presence than those safely locked in a courtly *wunderkammer*. They were artifacts of a larger design ambition for the palace as a whole. Second, de Caus devised a system by which the scores could essentially be reprogrammed through a rearrangement of variable pegs on a fixed drum. While the pinned barrel, was, by the seventeenth century, one of the conventional mechanisms for the automatic production of music, most prior devices had pins permanently attached to the drum. Thus, any modification of the drum necessitated the full replacement of one pinned drum with another. At an architectural scale of de Caus's organ, this would have been wholly impractical. De Caus's method allowed the dynamic change of even editing of sections of the score, on a large common drum. In effect, he developed one of the first systems of reprogrammable machine memory.

The principles of de Caus's work were widely influential, not only in terms of hydraulic engineering but also as precedents for the automata that entertained the courts of Europe. Aside from its immediate influence on water automata, the pinned drum mechanism was also a tangible

³¹ *Ibid.*

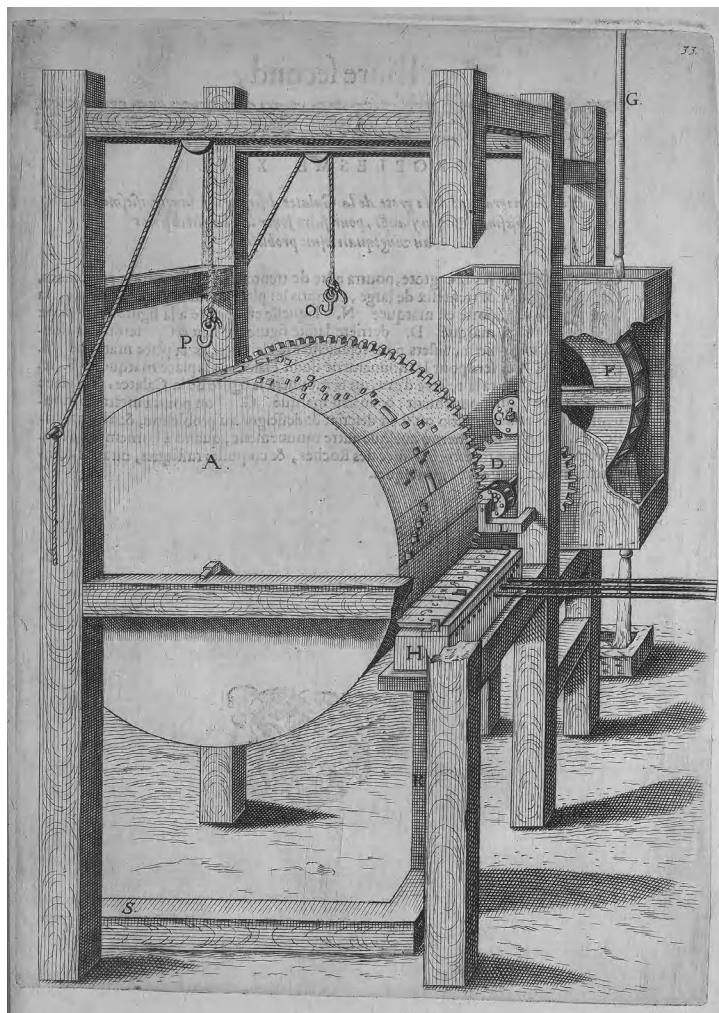


Figure 4. A perspective of de Caus's organ, showing the drum, keyboard, and driving mechanism. Source: S. de Caus, *Les Raisons des Forces Mouvantes, Avec Diverse Machines t'ant Utiles que Plaisantes*, Jan Norton, Frankfurt, 1615.

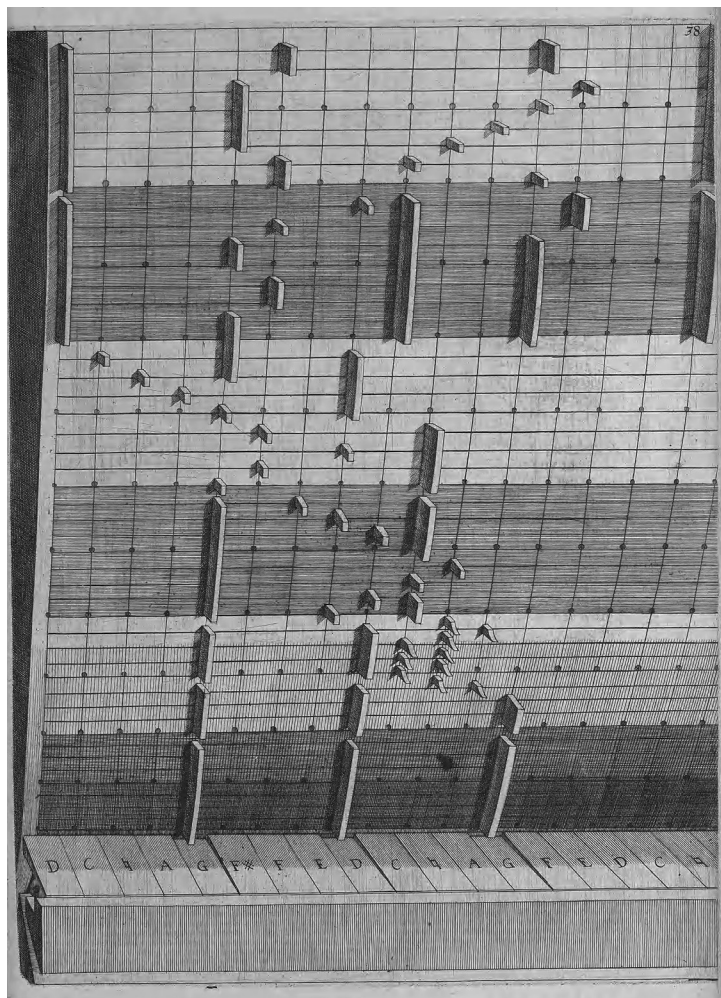


Figure 5. A detail of the reconfigurable pegs on the larger cylindrical pegboard of de Caus's organ. Source: S. de Caus, *Les Raisons des Forces Mouvantes, Avec Diverse Machines t'ant Utiles que Plaisantes*, Jan Norton, Frankfurt, 1615.

precedent for the Jacquard Loom and later punch card methods of automated tabulation and calculation. Though he did not overtly describe his work in these terms, de Caus showed the capacity of tectonic thinking to advance automation and, in particular, endow automata with marvelous new memory.

In some curious respects, de Caus's work paralleled Libeskind's later Memory Machine. The robust wooden frames, obvious musical associations, and geared articulation give the two a certain formal affinity. They also represent the polarities of human and machine memory. Though highly machinic in its articulation, Libeskind's aleatoric work suggests a dream state of human memory, with a tangible but mysterious structure. In contrast, de Caus's organ exhibits a strict organization that communicates its machine legibility. Yet both ultimately exist in the liminal space between technology and space, artifacts that probe the possibilities of artificial memories through architectural imagination.

THE ARCHITECTURE OF MACHINE MEMORY

In 1813, none other than the eminent German composer Ludwig van Beethoven (1770–1827) set out to write a piece commemorating Napoleon's Waterloo defeat not for orchestral performers but for an elaborate musical automaton: the Panharmonicon.³² This extraordinary device, capable of reproducing the sounds of a small orchestra, as well as gunshots and cannons, was the invention of Johann Nepomuk Mälzel (1772–1838), Beethoven's erstwhile patron. Mälzel was a polymathic showman and one of the most charismatic and prolific inventors of musical automata of the early nineteenth century. An itinerant exhibitor around Europe and later the United States, he was credited with a wide range of musical machines. He popularized a variant of the metronome, though his claims to have invented it were questionable.³³ Ever the showman, Mälzel acquired the Mechanical Turk, the legendary false automaton in which a grandmaster chess player would sit, concealed, and play an unsuspecting opponent, guiding what appeared to be the actions of an automaton.³⁴ Mälzel purchased the Turk from its original inventor,

³² T. Riley, "Composing for the Machine," *European Romantic Review*, 20, 3, 2009, p. 368.

³³ "Observations upon the Automaton Chess Player, now exhibiting in this city, by Mr. Maelzel, and upon various Automata and Androides," *The Franklin Journal and American Mechanics Magazine*, 3, 1827, p. 132.

³⁴ *Ibid.*, p. 130.

Baron von Kempelen, who created it in 1769 and subsequently exhibited it widely.³⁵ Unfortunately, the collaboration ended in a scandalous lawsuit in which Beethoven accused Mälzel of falsely claiming the rights to Beethoven's piece.³⁶ Mälzel, for his part, surreptitiously inscribed Beethoven's partial score on a recording cylinder, which he then used at performances of the Panharmonicon for years.³⁷ If one were to consider Beethoven's score as a kind of software for the Panharmonicon, the conflict might be considered an early hint of software copyright disputes to come.

Despite the acrimonious conclusion of this episode, the fact that an artist of Beethoven's reputation was willing to compose for a mechanical automaton shows not only the degree to which musical mechanisms had become a culturally accepted or even mundane aspect of the material culture of the early nineteenth century. It also indexes the extent to which the technology of recording and programming the actions of automata—particularly through means such as pinned cylinders—had been diffused and applied in a range of contexts. In the nearly two hundred years since the designs of the water organ of de Caus, automatic musical instruments—and the artificial memory they employed—had become part of the wider culture. Historian Terrance Riley noted that “By 1800 even the minimally-informed knew that all true automata, both anthropomorphic and non-anthropomorphic, were fundamentally musical machines—that beneath the costume of the android and the cabinet of the music box lay identical, and quite familiar, clockwork mechanisms,” mechanisms of artificial memory.³⁸

In the two hundred years since Beethoven, machine mnemotechnologies advanced enormously. The pinned cylinders of de Caus and others became direct progenitors of wax recording cylinders, phonograph records, hard and floppy computer disks, compact disks, and the entire family of computational storage that followed. Born of the entanglement of wonder-making and engineering, they occupy a remarkable position as a moment of convergence between architectonic thinking, automated performance, and emerging formats of mechanized memory. In very real respects, musical automata set in motion the possibility of machine memory that is now foundational to our technological culture.

³⁵ *Ibid.*, p. 125.

³⁶ T. Riley, “Composing for the Machine,” p. 370.

³⁷ *Ibid.*

³⁸ *Ibid.*, p. 369.

The imaginings of de Caus may seem far removed from the micro-electrical switching and hyperscale digital memory that powers today's computational automatons. Indeed, water and hydraulic automation has been relegated to something of a byway in the history of technology, long superseded by electricity and silicon. Yet occasionally, the abortive lineage of water organs as computational device resurfaces, to intriguing effect. William Phillips's 1949 MONIAC (Monetary National Income Analogue Computer), a device that used water flows to simulate macroeconomic scenarios, is a well-known example. Through a series of gates and flues, an operator could alter the distribution of funds in a dynamic simulation of the national economy of the UK.³⁹ MONIAC might be seen as a kind of water organ for playing the economy in simulation. Speculative visions of water-computing devices have also periodically enlivened architectural speculation. American architecture collective Ant Farm's well-known Dolphin Embassy project featured a powered by water and operated by dolphins.⁴⁰ More pragmatically, new technologies like microfluidics and "lab on a chip" devices offer a kind of return to mechanistic fluid engineering at an impossibly tiny scale, a kind of de Causian hydraulics in miniature. Perhaps we have not seen the last of water in computational automata.

Beyond its contribution to water-powered automata and its advancement of machine memory precursors, de Caus's work prompts reflections on alterities of automated architecture. How would architecture itself have been different if de Caus's inventions had been integrated into the disciplinary repertoire? What if the pinned drums and automated actuation had driven the behavior of the building itself? What if not only automata but buildings themselves gained a mechanical memory? As an architect, it seems at least possible that de Caus may have mused on such questions. If those consequences had been played through, one could imagine entire buildings driven by kinetic scores, coming alive in animated reverie. Finally, de Caus's trajectory also offers one index of the fluid boundaries of early modern architectural disciplinarity that led to remarkable experiments in kinetic automation. De Caus represents an intriguing moment of alignment between architecture and proto-computational automation, a moment which mechanical memory became nearly architectural.

³⁹ C. Bissell, "Historical perspectives: The Moniac A Hydromechanical Analog Computer of the 1950s," *IEEE Control Systems Magazine*, 27, 1, 2007, pp. 69–74.

⁴⁰ A. Witt, "An A(I)lien Embassy," *AD: Architectural Design*, 94, 3, 2024, pp. 54–61.

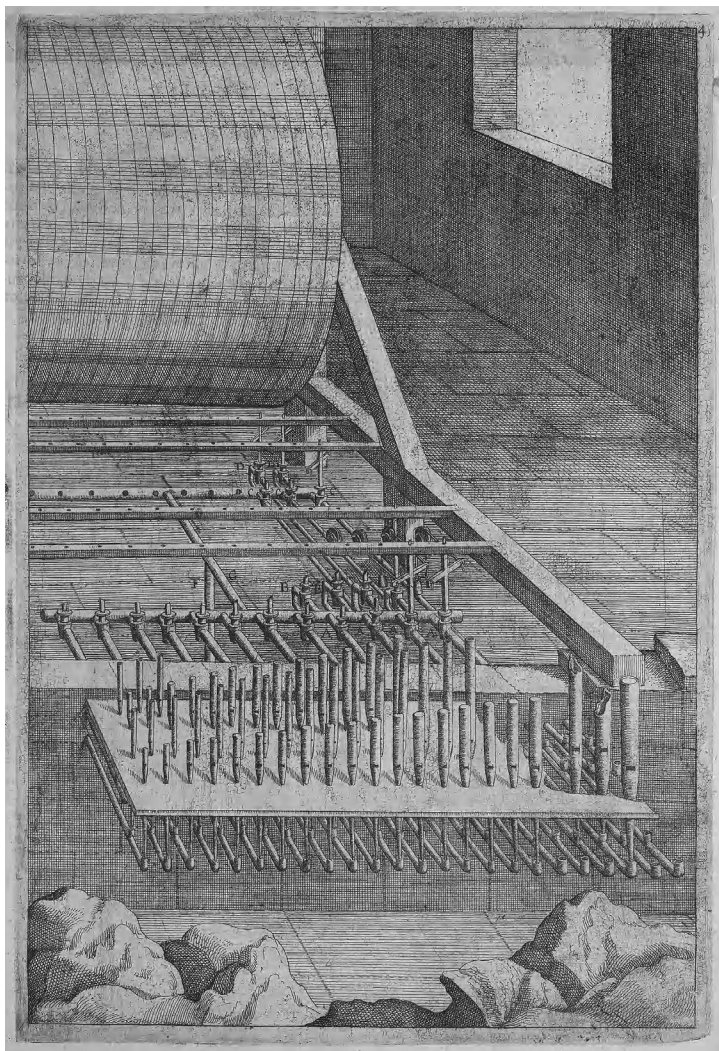


Figure 6. Details of the pipework for de Caus's organ.
Source: S. de Caus, *Les Raisons des Forces Mouvantes, Avec Diverse Machines t'ant Utiles que Plaisantes*, Jan Norton, Frankfurt, 1615.

REFERENCES

- Bissell, Chris (2007), "Historical perspectives: The Moniac A Hydromechanical Analog Computer of the 1950s," *IEEE Control Systems Magazine*, 27, 1, pp. 69–74.
- Blair, Ann (2010), *Too Much to Know: Managing Scholarly Information Before the Modern Age*, New Haven: Yale University Press.
- Campos, Pablo Gumiel (2024), *Aqua In Palatio: Water in 16th-century European palaces, The Iberian Peninsula as a case study* Vol. 1, PhD Diss.
- Carmo, Mario (2011), *The Alphabet and the Algorithm*, Cambridge: MIT Press.
- De Caus, Salomon (1615), *Les Raisons des Forces Mouvantes, Avec Diverse Machines t'ant Utiles que Plaisantes*, Frankfurt: Jan Norton.
- Grillo, Francesco (2019), *Hero of Alexandria's Automata: a critical edition and translation, including a commentary on Book One*. PhD thesis, Glasgow: University of Glasgow.
- Hero of Alexandria (1851), *The Pneumatics of Hero of Alexandria, from the original Greek*, trans. Joseph George Greenwood, London: Taylor, Walton and Maberly.
- Kaiser, Simone M. and Matteo Valleriani (2016), "The Organ of the Villa d'Este in Tivoli and the Standards of Pneumatic Engineering in the Renaissance," in Hubertus Fischer, Volker R. Remmert, Joachim Wolschke-Bulmahn (eds.), *Gardens, Knowledge and the Sciences in the Early Modern Period*, Berlin: Birkhäuser, pp. 77–102.
- Keller, A. G. (1967), "Pneumatics, Automata, and the Vacuum in the Work of Giambattista Aleotti," *The British Journal for the History of Science*, 3, 12, pp. 338–347.
- Libeskind, Daniel (2001), *The Space of Encounter*, New York: Universe.
- Marr, Alexander (2004), "Understanding Automata in the Late Renaissance," *Le Journal de la Renaissance*, 2, 1, pp. 205–222.
- Martin, C. B., Max Deutscher (1966) "Remembering," *The Philosophical Review*, 75, 2, pp. 161–196.
- Riley, Terrance (2009), "Composing for the Machine," *European Romantic Review*, 20, 3, pp. 367–379.
- Roberts, Ben, (2006) "Cinema as mnemotechnics," *Angelaki: Journal of Theoretical Humanities*, 11, 1, pp. 55–63.
- Stiegler, Bernard (1998), *Technics and Time, 1: The Fault of Epimetheus*, trans. Richard Beardsworth, George Collins, Stanford: Stanford University Press.
- Tybjerg, Karin (2004), "Hero of Alexandria's Mechanical Geometry" *Apeiron*, 37, 4, pp. 29–56.
- Vitruvius, (1914) *The Ten Books on Architecture*, trans. Morris Hicky Morgan, Cambridge, Mass.: Harvard University Press.
- Von Neumann, John (1945), "First draft of a report on the EDVAC," Philadelphia: Moore School of Electrical Engineering, University of Pennsylvania.
- Witt, Andrew (2002) *Formulations: Architecture, Mathematics, Culture*, Cambridge, Mass.: MIT Press.
- Witt, Andrew (2004), "An A(I)lien Embassy," *AD: Architectural Design* 94, no. 3, 54–61.

Matias del Campo*

ARCHITECTURE AND THE ONTOLOGY OF INFERENCE

ABSTRACT: In this essay we develop *inferentialism* as a methodological orientation for architectural practice by staging a philosophical confrontation with the discipline's inherited models of form-making. Beginning with Kempelen's chess-playing Turk as an early diagram of distributed cognition, the argument moves through Simondon's critique of hylo-morphism, the imposition of form upon passive matter, to demonstrate why conventional design methods, including *parametricism*, fail to account for the epistemic conditions introduced by artificial intelligence. Simondon's concept of ontogenesis provides the theoretical scaffold: form crystallizes from within, the way a crystal precipitates in a saturated solution, structure and content emerging simultaneously from the primordial soup of data, energy, and potential. This framework dismantles the assumptions embedded in deterministic design paradigms and exposes how schemas imposed onto material suppress its own tendencies, its own metastability, and its own capacity for individuation. The residual hylo-morphisms of contemporary computational design, including top-down constraint hierarchies and convergence toward predetermined formal objectives, reveal an inherited Rationalist commitment to optimization that inference renders obsolete. The argument culminates in the articulation of inferentialism: a mode of practice where form, agency, and knowledge emerge through probabilistic transductive processes distributed across human cognition, computational models, and material environments.

KEYWORDS: inferentialism, individuation, transduction, latent space, hylomorphism, distributed agency, ontogenesis

* Matias del Campo: Digital Architecture and Planning, TU Wien, matias.campo@uwien.ac.at

This is an Open Access article under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not transformed in any way and is properly cited.

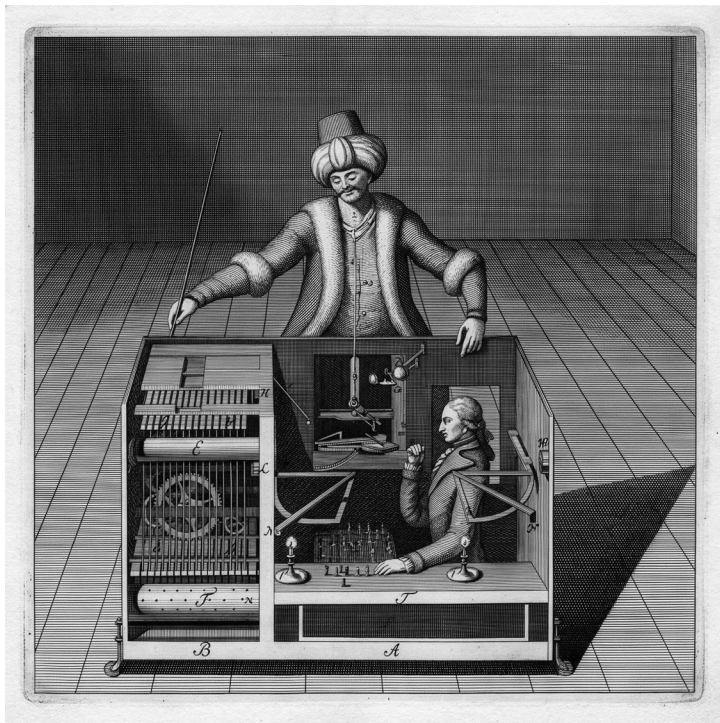


Figure 1. Joseph Racknitz, *Cross-section of Wolfgang von Kempelen's chess-playing automaton, The Turk*, 1789. Published in *Über den Schachspieler des Herrn von Kempelen*, Leipzig and Dresden, 1789. The apparatus serves as a foundational example of a technical ensemble, where the "hidden operator" within the cabinet facilitates a transductive operation across a structured field of play.

MOVEMENT I: THE MECHANICAL INTERVAL

Recently, I toured the Schönbrunn Palace in Vienna. Architecture friends on the visit asked me to show them around. While strolling through these wonderful Rococo spaces, we stumbled upon an important room. It was reserved for the entertainment of the royal family, and this is the room where a child Wunderkind named Wolfgang Amadeus Mozart played blindfolded to the astonishment of Empress Maria Theresa and her entire family. Legend has it that after the concert, the little lad fell into the arms of Marie Antoinette (the one of French Revolution fame) and asked her to marry him. However, for this essay far more important was another event in the same room a couple of years later. In 1770 Wolfgang von Kempelen presented a chess-playing automaton in this very room, before the assembled court of Maria Theresa.¹ The apparatus consisted of a seated figure of a pipe smoking Ottoman, placed behind a wooden cabinet topped by a chessboard. The life-sized mannequin was carved out of maple wood and robed in a lavish Turkish costume, including a large turban crowning the head (Figure 1). He was able to move his pipe to the mouth with his left arm (whilst pondering a move) and had an astounding dexterity in the right arm and hand to pick up and move the chess pieces around.

The doors could be opened to reveal a complex assembly of gears, levers, and rotating components. The audience was invited to examine the complex mechanism up close. The strange machinery appeared to account for all its operations. When the doors closed, the figure started playing. It responded to moves, adjusted pieces, and concluded games. On its tour through the courts of Europe the automaton encountered luminaries such as Benjamin Franklin and Napoleon Bonaparte, who tried to cheat, but the Turk was not having it, and wiped all the figures from the board annihilating the game. The performance suggested an interior governed by mechanical necessity. In the aftermath of the Turk's circulation, the machine inspired serious interrogation of automation, from a wide array of figures such as Edgar Allan Poe,² Henri-Louis Jaquet-Droz,³

¹ T. Standage, *The Turk: The Life and Times of the Famous Eighteenth-Century Chess-Playing Machine*, Berkley Books, New York, 2003.

² E. A. Poe, "Maelzel's chess-player," *Southern Literary Messenger*, 2, 5, 1836, pp. 318–326.

³ The strongest evidence that Wolfgang von Kempelen and Henri-Louis Jaquet-Droz were aware of one another occurs in London. In 1783, Kempelen brought The Turk to London, exhibiting it at Savile Row. At that same time, Henri-Louis Jaquet-Droz was also in

Charles Babbage,⁴ and Ada Lovelace,⁵ who is considered the very first programmer.⁵ A later revelation of a concealed human operator, crouching inside the wooden box (probably in the most uncomfortable way imaginable!), transforms the interest in the device. It morphs the question. The chess-playing figure occupies an ambiguous technical position. The fascination emerges from its dualist pose: mechanism cosplaying as intellect, intellect routed through a mechanical device. The operations of this machine depend on a composite arrangement: mechanical interface, human cognition, spatial concealment and staged visibility. Each of these elements participates in the performance, but none accounts for it alone. The machine produces an effect of autonomy through the orchestration of distributed processes whose boundaries do not coincide with the visible form of the object. What we are seeing in the example of the Turk, is a form of unified agent: an ensemble sustaining the operations, a cabinet suggesting containment, boundary, or frame. Resulting in a performance depending on circulation across components whose relations remained partially obscured. The Turk, this curious machine, allowed agency to circulate across the system, thus making the specific event—the preconceived motions of the chess pieces, the awe of the audience, the disbelief of the losing party—possible in the first place. In doing so, the device displaced the source of action within a structured interval between mechanism, operator, and spectatorship.

In contrast, the architectural discourse has long relied on a stable distinction between *instrument* and *author*.⁶ Thus, tools mediated between conception and execution, and machinic processes extended the reach of intention into domains of calculation, fabrication, and coordination. In doing so, the discourse maintained a narrative where the operative field of a project appeared to precede its technical realization, and procedures translated that field into form. Kempelen's chess-playing figure destabilized the balance of this alignment. It staged a technical object whose

London, displaying his own famous automata—The Writer, The Draughtsman, and The Musician—at King Street, Covent Garden.

⁴ C. Babbage (1864). *The Chess Player: In Passages from the Life of a Philosopher*, Harlow, Longman, Green, Longman, Roberts, & Green, London, 1864, pp. 465–471.

⁵ In her famous “Notes” on the Analytical Engine (1843), Lovelace famously argued that the machine “has no pretensions whatever to originate anything.” This philosophical boundary—the difference between a machine following a program and a machine possessing true “agency”—was exactly the same debate that surrounded The Turk during its exhibition tours.

⁶ M. Carpo, *The Alphabet and the Algorithm*, The MIT Press, Cambridge, Mass., 2011.

operations are divorced from a single locus, thus the autonomy appeared superficial, and participation dispersed across the system that sustained it.

The relevance of this configuration extends beyond the eighteenth-century machine. Today, computational environments operate through distributed datasets, staged training procedures, iterative feedback, and concatenated sequences of calculation. The respective interfaces present surfaces of coherence unfolding across operations within infrastructures that exceed immediate visibility. The outputs of these operations stabilize provisionally as subsequent iterations recalibrate earlier results. *The result is that it becomes increasingly difficult to locate where decisions are made and how form settles into place.* Agency circulates across an ensemble of technical and human operations whose boundaries do not align with the visible frame of interaction. The Turk offers an early diagram of this condition. It marks a moment at which technical systems began to occupy an interval between instrument and participant. Participation thus disperses across models, data, interfaces, and interpretive acts. What persists is the structural ambiguity first staged by the device: a technical object whose operation cannot be reduced to either tool or autonomous subject. The implications for architectural methods follow from this displacement.

MOVEMENT II: INSTRUMENTS, INTERIORS, AND THE STABILITY OF METHOD

Circling back to the Instruments and Author trajectory of this essay, architectural practice has long been organized around a relatively stable relation between intention, instrument, and form.⁷ The tools of architectural trade mediated between conception and realization: drafting instruments such as compass and straightedge translated thought into geometry, while model-scale mock-ups tested spatial relations through material approximation. Along the historical lineage that followed, machines extended the reach of calculation and fabrication, allowing increasingly complex geometries and assemblies to be coordinated across expanding technical infrastructures. Procedures unfolded within recognizable sequences in which a project established its operative field, its historical milieu, techniques resolved consequences, and form stabilized as a visible result emerging from this alignment.

⁷ S. Allen, *Practice: Architecture, Technique and Representation*, 2nd ed., Routledge, London, 2009.

Across technological regimes, this structure maintained continuity:

- The drafting apparatus,
- The perspective device,
- The analog model,
- The parametric script.

These configurations reorganized the procedures through which architecture took shape while preserving a familiar alignment between authorship and execution. The operative field appeared in advance of its technical articulation, and instruments acting within that field generated variation that remained legible through an underlying relational order. Within this disposition technical systems functioned primarily as mediators, translating an established schema into spatial or material form. One might recall the arm of the Mechanical Turk traversing the chessboard, grasping a figure and relocating it within a bounded spatial frame, following rules that appeared internally determined while relying on technological devices assembled from pantographs and magnets. Even when procedures grew highly complex, capable of coordinating intricate geometries or large-scale construction processes, their ontological status retained stability, operating within boundaries defined elsewhere while resolving relations without fundamentally altering the formation of the operative field itself.

The Turk (Figure 2) introduced an early turbulence into this alignment. Its performance depended upon a configuration in which operation exceeded any single locus of control or authorship. Despite the visible mechanism suggesting containment, actual operation unfolded across a distributed arrangement extending beyond the cabinet's surfaces of wood, ivory, brass, and silk. Apparent unity emerged from an ensemble whose internal coordination produced the effect of autonomy. The apparatus reorganized the relation between visible mechanism and distributed action, displacing agency across a configuration whose coherence depended upon circulation among components rather than containment within a single object. A similar disturbance begins to register in contemporary design environments, where technical procedures operate alongside predefined relations while simultaneously participating in the formation of the milieu within which design unfolds. Elements of the design process (prompts, datasets, model chains, feedback mechanisms) gain coherence across iterative interaction. Data acquisition,

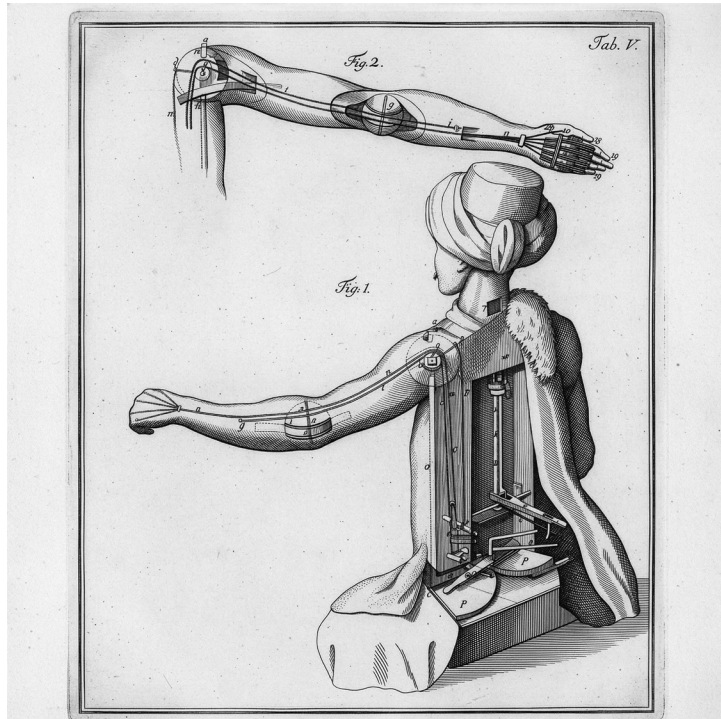


Figure 2. Joseph Racknitz, *Anatomical cross-section of the arm mechanism and upper torso of Wolfgang von Kempelen's chess-playing automaton, The Turk* (Table V, Figs. 1–2), 1789. Published in *Über den Schachspieler des Herrn von Kempelen*, Leipzig and Dresden, 1789. The cutaway reveals the pantograph linkages and internal drive components through which the figure's apparently autonomous gestures were mechanically coordinated—a precursor to contemporary inferential systems, where coherence arises through the progressive alignment of internal elements rather than the execution of a predetermined, hylomorphic form.

probabilistic modeling, staged training, and concatenated computational sequences enter into the formation of architectural proposals in ways that exceed any single originating schema. Consequentially the operative field emerges through successive alignments among heterogeneous components, producing a strange yet persistent stabilization of the design process across distributed systems. However, the conventional instrumental paradigm still operates. Drawings, scripts, and fabrication systems still mediate between conception and form, yet technical operations increasingly participate in shaping the field in which decisions arise. The boundary between instrument and participant acquires a new instability. Methods execute decisions while contributing to the conditions under which those decisions take shape. The Turk offered an early diagram of this condition, and contemporary computational environments extend it across infrastructures composed of datasets, model architectures, and iterative feedback loops. This leads to concerns about the ontological status of method itself: do technical procedures operate within a predefined field or participate in the ongoing formation of that field? A philosophical vocabulary attentive to processes of technical individuation provides a way to describe this shift with greater precision.

MOVEMENT III: MILIEU AND INDIVIDUATION

At this point the introduction of Gilbert Simondon becomes necessary, since the conceptual terrain outlined in the preceding pages already resonates with a vocabulary he developed in the middle of the twentieth century while reflecting on machines, technical objects, and the environments in which they take shape. Ontology, as the term is used in this essay, names neither a taxonomy of beings nor a search for ultimate substances, it refers to the question of emergence: the manner in which entities, relations, and operative fields come into existence, acquire consistency, then shift again under changing conditions. In the previous section, this problem in miniature was presented through *The Mechanical Turk*: presenting an object that appeared as an autonomous agent while its operation depended on a distributed arrangement of mechanisms, concealment, spectatorship, and human cognition. It is an arrangement whose coherence emerged in performance rather than residing fully in any single component. Starting off as an anecdote about a mannequin, a cabinet and a chessboard, the story turns into a more general question about architectural methods currently impacting with unusual intensity:

when technical systems participate in the formation of the field in which design takes place, what exactly comes into being, where does it stabilize, and through which processes does that stabilization hold?

To be clear, it is necessary to distinguish between two computational regimes often conflated in architectural discourse. The first consists of constituted modeling environments grounded in explicit hylomorphic control, such as *Parametricism*; the second comprises contemporary inferential systems where structure emerges through statistical alignment across data.

Returning to Simondon's philosophical project, that precisely gravitates around this question of emergence, developing a vocabulary through which technical and material formations can be understood as processes rather than finished entities. Central to this vocabulary stands the concept of individuation, a term he uses to describe the manner in which a technical object, a biological organism, or a social configuration comes into being through the progressive alignment of heterogeneous elements; for Simondon, "individuation is the operation of a system that is in a state of disparity and that, to resolve this disparity, finds a new structure."⁸ This happens within what he calls an associated milieu, an environment composed of energies, materials, interfaces, and operational conditions that participate directly in the formation of the object itself; or as Simondon put it, "the process of concretization is precisely this invention of an associated milieu, which is at once the condition of the technical object's functioning and the milieu of its relationship with the external world."⁹ Within such a milieu, coherence arises through phases of metastability, temporary equilibria sustained by ongoing exchanges among components whose relations remain open to further transformation, allowing the system to maintain performance while retaining potentials for change. Transduction¹⁰ names the process through which this alignment propagates across the system, a chain of operations through which a structural transformation in one domain induces adjustments in another, producing a continuity of becoming across the ensemble. Against the classical hylomorphic model in which form is imposed upon passive matter, Simondon proposes an account

⁸ G. Simondon, *L'individuation à la lumière des notions de forme et d'information*, Jérôme Millon, Paris, 2005, p. 5.

⁹ G. Simondon, *On the Mode of Existence of Technical Objects*, Minneapolis, Univocal, 2017, p. 60.

¹⁰ G. Simondon, *Sur la technique*, Presses Universitaires de France, Paris, 2014.

of technical becoming grounded in reciprocal adjustment between material conditions and operational procedures, an account in which form emerges through modulation within a field of relations that remains active throughout the process of formation. This reorientation allows technical objects to be understood as configurations whose identity takes shape through the processes that bring them into operation, rather than as instruments positioned for deployment within a predefined field, a perspective that bears directly on the contemporary condition in which computational systems participate in the formation of the very milieus through which architectural design emerges.

Within architectural discourse, this line of thought entered gradually, circulating through multiple theoretical and design trajectories that absorbed aspects of this vocabulary while translating them into distinct disciplinary concerns. Bernard Cache's work on objectile geometry¹¹ drew upon the notion of modulation (borrowed unapologetically from Deleuze) to frame form as continuous variation within a field, allowing geometry to unfold through transformation rather than through the application of fixed templates. Greg Lynn's early explorations of animate form and calculus-based modeling¹² resonated with the idea of metastable variation, treating geometry as a field of continuous adjustment, and subsequent work by figures such as Philip Morel,¹³ Achim Menges,¹⁴ and others¹⁵ operating at the intersection of computation and material processes continued to explore design as an evolving alignment among material behavior, fabrication logics, and computational procedures. Parallel developments associated with the work of Zaha Hadid and Patrik Schumacher articulated these tendencies at an urban and disciplinary

¹¹ B. Cache, *Earth Moves: The Furnishing of Territories*, The MIT Press, Cambridge, Mass., 1995.

¹² G. Lynn, *Animate Form*, Princeton Architectural Press, Princeton, 1999.

¹³ P. Morel, *Empiricism & Objectivity: Architectural Investigations with Mathematics*, Editions HX, Orléans, 2004.

¹⁴ A. Menges, *Material Synthesis: Fusing the Physical and the Computational*, Wiley, Hoboken, 2015.

¹⁵ For others, see also the list of participants in the exhibition *Architectures non-standard*, curated by Frédéric Migayrou and Zeynep Mennan at the Centre Pompidou (Musée National d'Art Moderne), Paris, from December 10, 2003, to March 1, 2004. Participating offices included: Asymptote (Hani Rashid & Lise Anne Couture), dECOi (Mark Goulthorpe), DR_D (Dagmar Richter), Greg Lynn FORM, Jakob + MacFarlane, Kolatan/Mac Donald Studio, Kovac Architecture (Tom Kovac), Lewis.Tsurumaki.Lewis, Objectile (Patrick Beaucé & Bernard Cache), NOX (Lars Spuybroek), ONL (Kas Oosterhuis & Ilona Lénárd), and R&Sie(n) (François Roche & Stéphanie Lavaux).

scale through the doctrine of *Parametricism*,¹⁶ framing architecture as modulation across interdependent parameters within a legible and internally coherent model space. These trajectories collectively displaced the understanding of architectural form as the execution of a prefigured schema, shifting attention toward processes through which coherence arises within a field of interacting elements while leaving unresolved the question of how technical systems themselves participate in the formation of the operative field within which architectural thinking unfolds, a question that acquires renewed urgency in the context of contemporary machine learning environments whose operations extend across distributed datasets, model architectures, and iterative feedback structures. Despite their emphasis on emergence, these approaches operate within fields whose relational structure precedes computation itself.

The position developed in these pages emerges from a slightly different orientation. It shifts focus from the technical object toward the chain of operations through which objects, images, models, and environments gain temporary coherence across distributed systems. In earlier computational design systems, computation modulated relations already defined; in contemporary machine learning systems, computation infers relations that did not previously exist as formal entities. Thus, contemporary machine learning environments introduce a further displacement in which the operative field of design takes shape through recursive alignment across data, models, infrastructures, and interpretive practices rather than appearing in advance as a predefined domain within which variation unfolds. Previous regimes of computational design methods operated within constituted parametric environments whose internal relations structured the horizon of possible transformations, enabling geometry to evolve through modulation across explicit dependencies among variables while preserving a legible field whose organization preceded the operations performed within it. Contemporary inferential systems unfold within a different ontological configuration in which the field itself arises through iterative exposure, recalibration, and evaluation across distributed technical ensembles, allowing patterns to register gradually across datasets and model architectures whose interactions generate phases of temporary coherence. A dataset gathers

¹⁶ P. Schumacher, "Parametricism as a style—Parametricist manifesto," presented at the 11th Architecture Biennale, Venice, 2008.

heterogeneous instances into a structured collection through which recurrent relations become perceptible to a computational system, while a model refers to a trainable architecture capable of registering such relations through adjustment of internal parameters across repeated exposure, allowing tendencies to stabilize provisionally across successive epochs and broader training cycles that structure the temporal dimension of learning within contemporary computational environments. Inference occupies a central role within this landscape as the mechanism through which outputs arise through statistical estimation grounded in prior exposure, producing architectural proposals whose coherence emerges through recursive alignment across data, models, simulations, and guided by human interpretive acts that steer trajectories and select among provisional outcomes.

Form and field take shape together through processes of *individuation* unfolding across distributed technical systems, producing an ontological condition in which stabilization registers as a phase within an evolving milieu whose coherence arises through alignment across heterogeneous operations rather than through the execution of a prefigured schema. Architectural practice operates inside this evolving configuration as a participant in processes of calibration, selection, and interpretation through which the operative field acquires consistency across concatenated procedures that sustain its formation across time.

MOVEMENT IV: FROM MORPHOGENESIS TO INFERENCE

Simondon's concept of *ontogenesis*¹⁷ sharpens the argument at this juncture, particularly when placed in relation to the architectural discourse on *morphogenesis*¹⁸ that shaped digital design thinking across the final decade of the twentieth century and the early years of the twenty-first. Morphogenetic approaches, developed through Greg Lynn's *Animate Form*, through the research culture surrounding the Architectural Association in London in the work of Michael Hensel, Achim Menges, Michael Weinstock, and George Jeronimidis,¹⁹ and through parallel

¹⁷ G. Simondon, "The position of the problem of ontogenesis," *Parrhesia*, 7, 2009, pp. 4–16.

¹⁸ A. Menges, "Computational Morphogenesis: Integral Form Generation," in A. Okeil, A. Al-Attili, Z. Mallasi (eds.), *Em'body'ing Virtual Architecture, Proceedings of the 3rd International ASCAAD Conference*, pp. 725–744.

¹⁹ M. Hensel, A. Menges, M. Weinstock, *Emergent Technologies and Design: Towards a Biological Paradigm for Architecture*, Routledge, London, 2010.

explorations by figures including Neri Oxman,²⁰ Kas Oosterhuis,²¹ and Alberto Estevez,²² described architectural form as the outcome of dynamic processes unfolding within generative environments structured by parametric relations, material behaviors, or biological analogies whose internal logic remained available to inspection through parameters, rules, and constraints. Geometry evolved through continuous modulation across relational fields whose organization structured the horizon of possible variation, allowing form to register as differential transformation operating inside constituted systems whose internal coherence preceded the transformations occurring within them. Morphogenesis provided architecture with a language for emergence, variation, and dynamic formation, while retaining an ontological assumption according to which the field of generation functioned as the ground within which variation unfolded, stabilizing the horizon of possible outcomes even as specific configurations continued to transform across successive iterations.

Simondon's description of crystal growth offers a lucid demonstration of ontogenesis as a process through which form and field emerge together through progressive alignment across a medium initially undifferentiated yet charged with potential for structural resolution. In a supersaturated solution, a seed crystal initiates propagation through which molecular relations extend across the surrounding medium, generating a structured lattice whose coherence registers gradually through successive phases of alignment rather than through the execution of a predetermined form imposed upon inert matter (Figure 3).

Growth unfolds through local interactions that establish conditions for further alignment, allowing structure to emerge through reciprocal engagement between seed and milieu as the surrounding environment acquires organization through the very process of crystallization. Instead of occupying a pre-existing field, the crystal structures the medium through the transductive propagation of internal relations, or, as Simondon observes, "each already constituted molecular layer serves as the base for the next layer being formed, and the result is an amplifying structure."²³

²⁰ N. Oxman, "Structuring materiality: Design Fabrication of Heterogeneous Materials," *Architectural Design*, 80, 4, 2010, pp. 78–85.

²¹ K. Oosterhuis, *Hyperbodies: Towards an E-motive Architecture*, Birkhäuser, Basel, 2003.

²² A. T. Estévez, "Genetic Architectures: New Bio-digital Techniques," *Architectural Design*, 81, 6, 2011, pp. 54–61.

²³ G. Simondon, *Individuation in Light of Notions of Form and Information*, University of Minnesota Press, Minnesota, 2020.

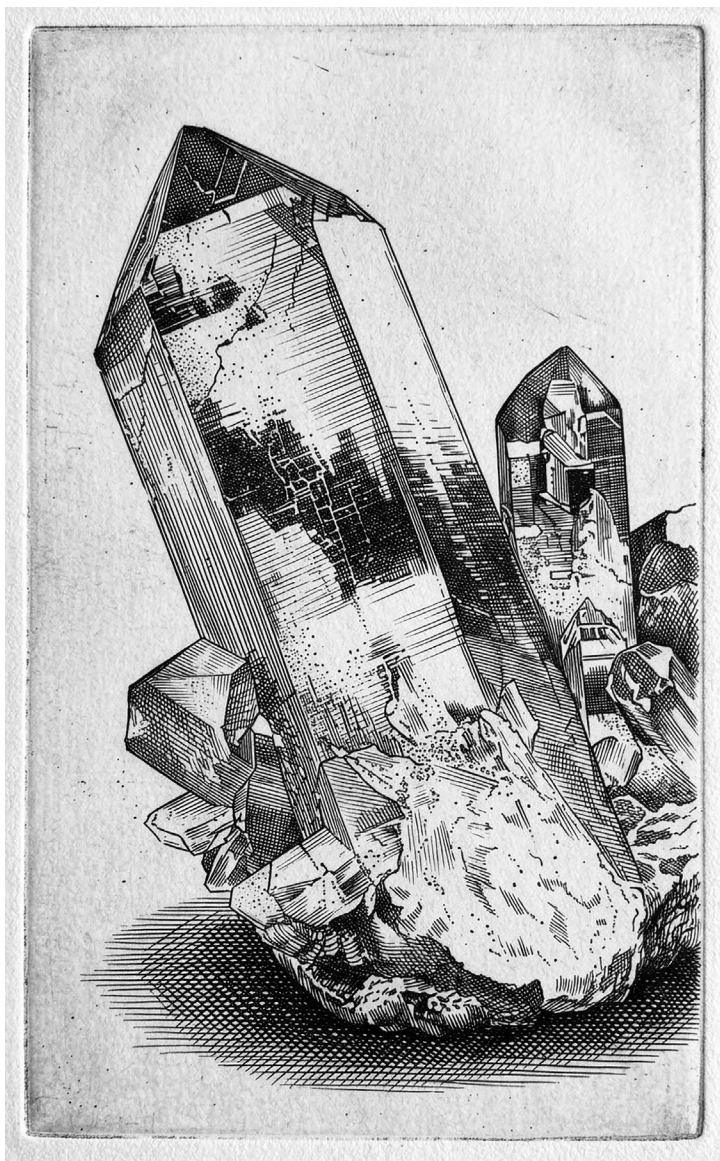


Figure 3. Matias del Campo, *Crystallization as Transductive Individuation*, 2018. Courtesy of the author. A diagram of ontogenesis where structure emerges through transductive propagation across a metastable medium: form as provisional settlement arising through reciprocal engagement between seed and milieu rather than external imposition.

This account provides a useful lens through which contemporary computational environments in architecture may be understood. A large image corpus assembled for training a generative model gathers heterogeneous visual instances into a field of potential relations whose consistencies register gradually through repeated exposure across model architectures capable of amplifying latent correspondences. In contemporary diffusion-based image models, visual data undergoes a transformation in which images are decomposed into numerical matrices composed of pixel intensities and color values, then subjected to iterative processes that introduce progressive noise until recognizable structure dissolves into a stochastic field of values whose distribution retains traces of prior relations without preserving explicit form. This noisy field functions as a saturated informational medium within which latent correspondences persist in distributed form. Generation unfolds through a reverse process of iterative denoising in which local correlations propagate across the field under the influence of conditioning inputs that guide alignment without prescribing final structure, allowing recognizable configurations to emerge gradually as relations stabilize provisionally across successive passes through the model (Figure 4). The resulting image appears as a temporary settlement within a field whose structure develops through recursive alignment across noise, data, and guided interpretation, recalling the propagation of crystalline structure across a saturated solution in which local ordering establishes conditions for further alignment across the surrounding medium.

At this juncture it becomes necessary to acknowledge that contemporary image models operate within regimes structured not only by statistical relations among pixels or spatial configurations, but also by semantic associations introduced through captioning, tagging, and other forms of linguistic alignment that bind visual data to descriptive vocabularies. Such processes shape the formation of relations within the dataset and influence the trajectories through which inference unfolds, introducing layers of meaning that interact with formal tendencies across the training corpus. The present argument maintains its focus on the ontological question of how form and field emerge through recursive alignment across distributed systems, leaving a sustained examination of semantic labeling, linguistic conditioning, and the politics of dataset construction to a separate inquiry whose scope extends beyond the present investigation, even as its relevance to the broader ecology of contemporary computational design remains evident.

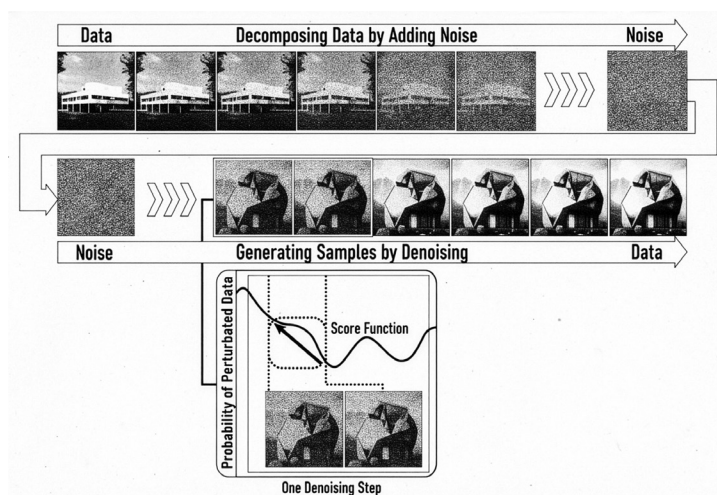


Figure 4. Matias del Campo, *The Transductive Cycle of Diffusion*, 2023. Courtesy of the author. A schematic representation of the iterative movement between structured data and stochastic noise. The sequence illustrates the forward process of decomposing data by adding noise and the generative reverse process of generating samples by denoising along a Markov chain. In the context of the ontology of inference, this cycle serves as a diagram for ontogenesis: form is not a static imposition upon matter, but a provisional stabilization—an individuation—emerging from the recursive alignment of noise, latent correspondences, and conditioning inputs.

Outputs emerging from such environments register as provisional stabilizations within a field whose organization develops through recursive alignment across data, model behavior, simulation, and guided interpretation, allowing architectural proposals to appear as temporary settlements within a milieu that continues to evolve through ongoing interaction. One might consider the series of latent-space explorations conducted through generative adversarial models trained on *Piranesi's Carceri* engravings (PiranesiGAN, Figure 5), where spatial configurations arise through iterative navigation across a probabilistic field whose internal relations register gradually through repeated probing, selection, and recalibration. Individual images acquire temporary coherence as architectural propositions, then reenter the evolving field as further iterations redirect the trajectory of exploration, allowing form to register as a provisional settlement within a continuously transforming informational landscape. Stabilization appears as a phase within an ongoing process, since the conditions that produced coherence remain active and capable of further transformation, situating architectural practice within a milieu whose structure develops through the operations that sustain it. Design activity unfolds through calibration, selection, and interpretation across distributed technical systems whose interactions generate phases of coherence that orient subsequent decisions. Datasets, model architectures, simulation environments, and interpretive judgments enter into dynamic alignment, allowing proposals to register through recursive propagation of relations across concatenated processes that gradually acquire recognizable consistency. Form settles provisionally as alignments accumulate across operations guided by human interpretation, registering as a temporary crystallization within an environment whose internal structure continues to evolve through sustained interaction rather than through the execution of a predetermined schema.

A useful way to render this trajectory more tangible appears in the formation of river deltas (Figure 6), where sediments travel within a fluid medium shaped by currents whose direction and intensity shift over time, depositing material in zones where alignment among forces allows particles to settle into temporary formations whose contours register the history of their emergence. Channels branch, merge, redirect; layers accumulate unevenly; emergent landforms acquire provisional stability while remaining open to further reconfiguration under changing conditions. The resulting landscape appears structured and legible, yet its coherence arises from distributed flows interacting across time rather



Figure 5. Matias del Campo and Sandra Manninger (SPAN), *PiranesiGAN: Spatial Individuation*, 2020. Courtesy of the artists. Generative adversarial network output trained on Giovanni Battista Piranesi's *Carceri d'invenzione* engravings. Architectural configurations emerging from a latent space trained on the *Carceri* series: each image acts as a provisional settlement—a momentary stabilization of form within a probabilistic milieu that continues to evolve through recursive navigation and data alignment.

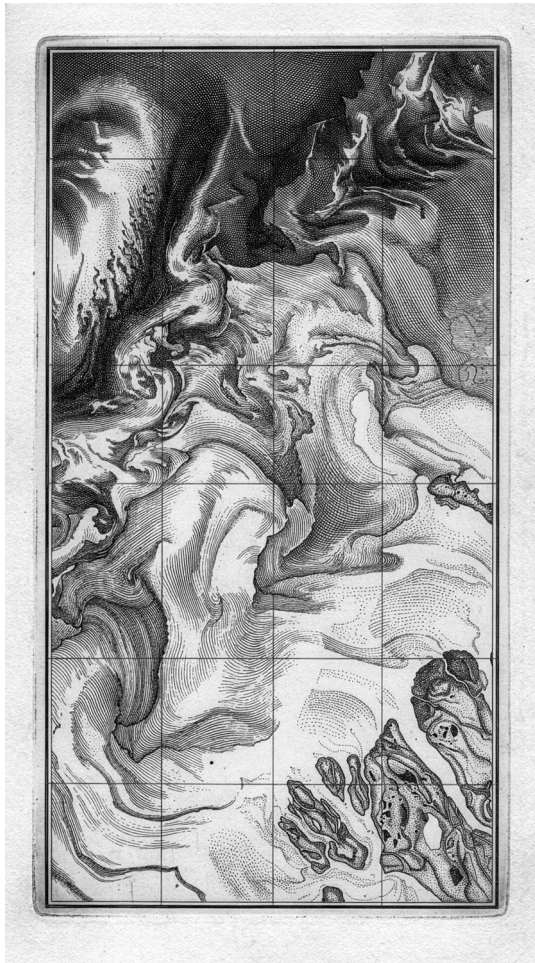


Figure 6. Matias del Campo, *The Delta as Inferential Landscape*, 2026. Courtesy of the author. An aerial cartographic rendering of a river delta illustrating structure emerging from distributed flow rather than external imposition. The delta functions as a diagram for the inferential milieu, where form acts as a temporary settlement stabilized through the recursive alignment of internal tendencies.

than from the execution of a singular geometry imposed from outside the system. Contemporary computational design environments unfold through comparable dynamics in which data circulates, models register tendencies, simulations redirect trajectories, and interpretive decisions modulate the flow, producing architectural configurations that settle provisionally within a milieu whose structure develops through the very operations that sustain it.

Elsewhere I have described engagement with generative systems as navigation across latent space, a terrain composed of statistical relations whose internal structure becomes perceptible gradually through iterative probing, selection, and recalibration, allowing certain regions of the space to condense into recognizable proposals while others dissolve back into the surrounding field of potential. Stabilization registers as alignments accumulate across successive operations, producing configurations that acquire temporary consistency without exhausting the potentials embedded in the environment that sustains them. What emerges resembles a landscape shaped by repeated flows of adjustment and interpretation, registering the traces of interaction that produced it while remaining open to further transformation across subsequent cycles of engagement.

Here, agency distributes across the ensemble rather than concentrating at a single locus. Computational processes register tendencies across data and simulations extend those tendencies into provisional spatial organizations. Evaluative acts redirect trajectories; and the architect operates within this evolving configuration as a mediator of alignments, introducing new material, filtering outputs, and steering the propagation of relations through which form and field acquire temporary coherence. Architectural proposals register as settlements within a moving landscape, stabilized through recursive interaction among heterogeneous components whose relations remain active, allowing the operative field to take shape through concatenated processes that continue to evolve as new alignments form. Seen from within this evolving configuration, the mechanism through which such alignments register becomes legible.

MOVEMENT V: INFERENCEALISM

Within contemporary computational environments associated with machine learning and generative modeling, inference functions as the operation through which patterns detected across prior exposure become actionable within present interaction, allowing outputs to emerge through

probabilistic evaluation grounded in accumulated relations (Figure 7) rather than through the execution of predefined geometric dependencies characteristic of earlier parametric systems. Inference registers tendencies across datasets, propagates correlations through model architectures, and presents configurations whose coherence reflects the alignment of distributed operations unfolding across time. Architectural proposals emerging from such environments may therefore be understood as the result of inferential processes through which relations stabilize provisionally within an evolving field, allowing design activity to unfold through orientation within a probabilistic milieu whose structure takes shape through the very operations that sustain its formation.

Decision-making within such a milieu assumes the character of orientation rather than prescription. The designer engages a field of tendencies whose responses register through gradients, clusters, and zones of stabilization that guide subsequent action, producing a mode of practice in which navigation across probabilistic terrain replaces traversal across a fixed geometric domain associated with earlier digital modeling paradigms (Figure 8). Method acquires the character of a technical ensemble composed of interacting procedures whose alignments generate phases of temporary coherence that sustain performance while retaining potentials for transformation. Architectural form appears as a momentary settlement within a landscape structured by inference, concatenation, and iterative stabilization.

A methodological orientation becomes legible through the operations described in the preceding sections, giving conceptual clarity to procedures already active within contemporary architectural work. What has been described across these pages can be gathered under the term *inferentialism*. The term designates a mode of architectural activity in which form emerges through probabilistic evaluation across concatenated technical systems whose interactions participate directly in the formation of the operative field. The field within which design unfolds takes shape through recursive alignment among datasets, model architectures, simulations, interfaces, and interpretive acts that guide trajectories across successive iterations. Stabilization registers as a phase within an ongoing process of individuation sustained by these interactions, allowing architectural form to appear as a provisional calcification within an evolving ensemble whose internal relations continue to shift across time.

In this sense the term consolidates the trajectory extending from the mechanical interval staged by Kempelen's Turk through Simondon's

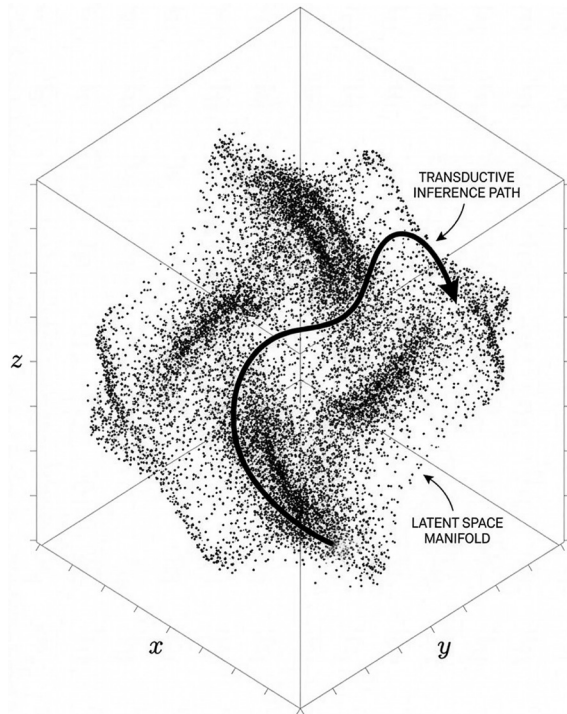


Figure 7. Matias del Campo, *The Topology of Inference*, 2026. Courtesy of the author. Diagram of a three-dimensional latent space manifold with a transductive inference path. The data density constitutes a pre-individual milieu where inference stabilizes provisional form. Coherence remains contingent on model weights: a robust path achieves transductive resolution, whereas underfitting lacks structural definition and overfitting is trapped by localized noise, failing to navigate generative potentials.

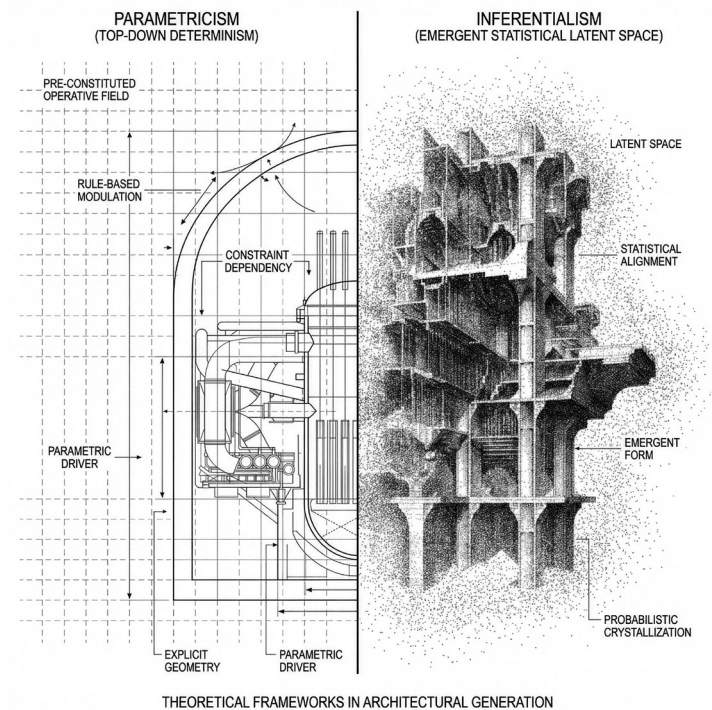


Figure 8. Matias del Campo, Theoretical Frameworks in Architectural Generation: Parametricism and Inferentialism, 2025. Courtesy of the author. Diagram. Left panel: Parametricism as top-down determinism—a pre-constituted operative field governed by rule-based modulation, constraint dependency, and explicit geometry driven by parametric inputs. Right panel: Inferentialism as emergent statistical latent space—probabilistic crystallization producing emergent form through statistical alignment within a distributed latent field. The diagram maps the epistemic difference between traversal across a fixed geometric domain and navigation across probabilistic terrain as the operative condition of contemporary AI-assisted architectural practice.

account of technical individuation to the contemporary situation in which inference functions as a generative mechanism within architectural practice. *Inferentialism* is a concept in which an architectural method unfolds through distributed processes whose coherence arises through recursive interaction across concatenated systems. Inference and concatenation mark the specific contribution presented here, situating architectural work within a landscape where agency circulates across data, models, simulations, and interpretive acts whose alignments generate phases of stabilization within an evolving field.

MOVEMENT VI: THE MIRROR ROOM

Picture yourself back in 1770, in that lustrous Mirror Room of Schönbrunn Palace (Figure 9). Dozens of candles reflected across the large mirror panels cladding the elaborate white-and-gold Rococo interior. Kempelen's concoction, the Turk, positioned at the center of the hall, its surfaces still carrying the scent of fresh paint while presenting the same assurance of containment: carved panels, hinged doors, a chessboard resting above a massive plinth that appears to hold the entirety of the operation within its wooden enclosure. The figure of a Turk wearing an oversized turban gazing toward the opponent across the board. Opening the box reveals an arrangement of gears, cams, springs, and sliding components whose visible complexity suggests mechanical sufficiency. When closed, the Ottoman mannequin resumes the game, scanning the board with his mechanical arm. Motion after motion of the chess pieces proceeds with measured delay, responses registering across the board as if calculation occurs inside the cabinet itself.

Despite these seemingly computational procedures, the performance never resided fully within the visible mechanism. Operation extended across a milieu that included the concealed human operator, the mechanical linkage, the spatial staging of the room, and the attention of those gathered around the board. Coherence emerged through coordination across this arrangement, and the figure presented at the center of the hall offered a surface of unity while sustaining relations that exceeded the frame appearing to contain them. What took place in the Mirror Room can therefore be read as an early diagram of distributed computation, one in which apparent autonomy stabilized through alignment across heterogeneous components rather than through containment within a single device.



Figure 9. Spiegelsaal (Mirror Room), Schönbrunn Palace, 2026. Interior designed by Nicolaus von Pacassi, 1743. Photograph by Matias del Campo. The room in which Wolfgang von Kempelen first demonstrated the chess-playing automaton before the court of Maria Theresia, here read as an early spatial diagram of distributed computation: apparent autonomy stabilized through alignment across heterogeneous components rather than containment within a single device.

Contemporary design environments shaped by artificial intelligence experience a comparable structural condition. Assembling datasets, training models, evaluating outputs, extending simulations, adjusting parameters, and reintroducing iterations compose concatenated chains of operation through which architectural proposals acquire temporary coherence across distributed systems. What appears at the interface as a resolved form registers as the outcome of inferential alignment across these chains, allowing the operative field of design to take shape alongside the processes acting within it, rather than preceding them as a predefined domain. Simondon's account of individuation provides the vocabulary necessary to describe this configuration with precision, situating technical formations within associated milieus whose relations remain active throughout phases of stabilization that sustain their operation. Within such conditions, inference registers tendencies across data, propagates relations through models, and guides provisional stabilization across iterative cycles of calibration in which human interpretation participates as one element among many.

These particular coordinates mark the intersection of the trajectories presented in this essay. The eighteenth-century automaton staged intelligence as a performance sustained by distributed coordination, while contemporary computational environments embed inference within the processes through which architectural relations align and stabilize across time (Figure 10). The fake Automaton once presented before the Viennese court now reads as a diagram through which the ontology of inference becomes legible: a condition in which form and field take shape together through concatenated operations whose coherence registers provisionally within an evolving milieu. Architectural practice unfolds within this interval, engaging technical systems that participate directly in the formation of the field in which design decisions acquire consistency. What emerges under the name *Inferentialism* concerns this orientation. Architectural program and form appear as a temporary stabilization within a field structured by inferential alignment, and the ontology at stake concerns the conditions through which relations emerge, propagate, and hold long enough to be called a project.

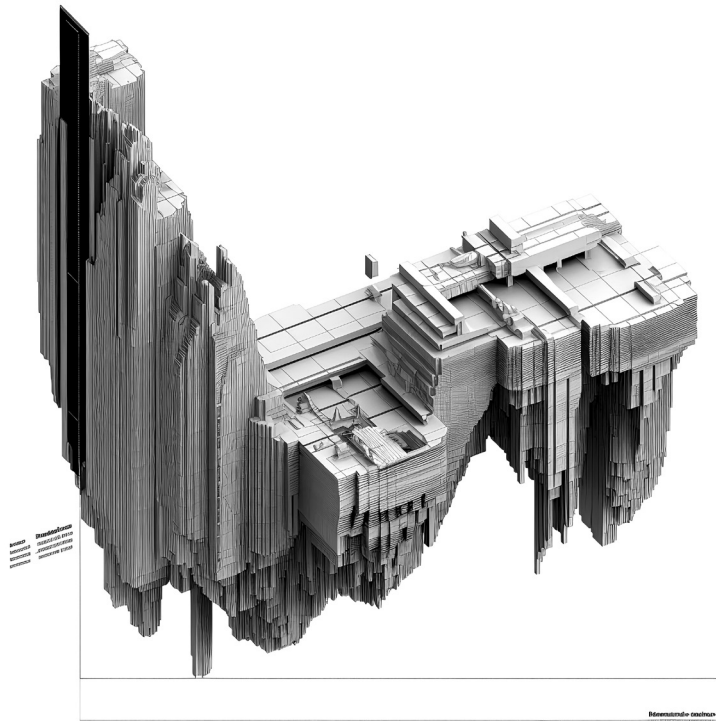


Figure 10. Matias del Campo, *Blocks_Boulder*, 2024. Courtesy of the author. The image registers what hylomorphic logic cannot account for: tectonic character that emerges from within probabilistic processes rather than descending from a form imposed upon passive matter. Mass, striation, overhang, and habitable ledge crystallize through iterative statistical alignment—a Simondonian individuation in which the milieu participates in producing the individual rather than receiving it. The project does not execute a prior intention but stabilizes, provisionally, within an inferential field. What appears as architecture is the momentary calcification of a process still in motion. This is the paradigm shift the essay proposes: not parametric determination operating top-down across a pre-constituted geometric domain, but *Inferentialism*: form as a phase condition emergent from probabilistic terrain.

GLOSSARY

Associated Milieu

A term developed by Gilbert Simondon to describe the environment that co-constitutes a technical object. A machine does not operate in isolation; it comes into being through alignment with materials, energies, interfaces, and users. In contemporary computational design, datasets, models, interfaces, and human interpretation form such a milieu. Architectural form emerges through interaction with this evolving field rather than within a predefined container.

Concatenation

A chaining of operations in which outputs from one procedure become inputs for another. In inferential design environments, prompts, datasets, model runs, simulations, and evaluations form concatenated sequences whose cumulative alignment generates provisional coherence. Architectural work unfolds through such chains rather than through a single linear act of design.

Dataset

A structured collection of examples used to train a machine-learning system. For architectural purposes, a dataset is not merely a storage of images or models but a field in which relations become statistically perceptible. It provides the informational medium through which patterns can be inferred and later re-activated in design processes.

Diffusion Model

A class of generative machine-learning systems that produce images or forms through iterative transformation between noise and structure. Starting from a field of randomness, the model gradually stabilizes recognizable configurations by aligning local correlations. For architecture, diffusion models illustrate how form can emerge as a provisional stabilization within a probabilistic field.

Distributed Agency

A condition in which action does not originate from a single author or instrument but circulates across an ensemble of human and technical components. In contemporary computational environments, decisions arise through interactions among datasets, models, interfaces, and designers. Agency appears as a relational effect rather than a singular source.

Field (Operative Field)

The domain within which architectural relations take shape. In earlier computational design, this field often preceded the operations performed within it. In inferential environments, the field itself emerges through iterative alignment across data, models, and interpretation. Form and field therefore take shape together.

Generative Model

A machine-learning system capable of producing new images, forms, or spatial configurations based on patterns learned from prior data. Rather than executing explicit instructions, generative models operate by inferring statistical tendencies and stabilizing them into provisional outputs. They function as engines of variation within probabilistic environments.

Inference

The process through which a trained model produces an output based on patterns learned from prior exposure. Inference does not retrieve stored forms; it estimates relations that are statistically likely given the input. Within the essay, inference names the mechanism through which architectural proposals emerge from probabilistic alignment across datasets, models, and interpretive acts.

Inferentialism

The methodological orientation proposed in the essay. Inferentialism describes an architectural practice in which form emerges through probabilistic evaluation across concatenated technical systems. The operative field of design takes shape alongside these processes rather than preceding them. Architectural work becomes an activity of navigating and stabilizing relations within an evolving inferential milieu.

Individuation

A concept from Simondon describing the process through which entities and relations come into being. Rather than assuming fixed objects or forms, individuation emphasizes emergence through alignment within an associated milieu. Architectural proposals can be understood as phases within such processes, acquiring temporary consistency through iterative stabilization.

Interface

The surface through which humans interact with computational systems. Interfaces present outputs as coherent objects while concealing the distributed processes that produced them. In design environments shaped by artificial intelligence, the interface stabilizes perception while underlying operations remain dispersed across data and models.

Latent Space

A mathematical representation of relations learned by a machine-learning model. Instead of storing images or forms directly, the model organizes patterns within an abstract space of tendencies and correlations. Designers navigating such a space explore regions where recognizable configurations condense from statistical structure.

Machine Learning

A computational approach in which systems learn patterns from data rather than following explicit instructions. In architectural contexts, machine learning allows relations to be inferred across large datasets, producing outputs that reflect statistical tendencies rather than predetermined geometries.

Metastability

A state of provisional equilibrium that holds temporarily while remaining open to transformation. Simondon uses the term to describe systems that sustain coherence through ongoing exchange with their milieu. Architectural proposals emerging from inferential environments may be understood as metastable formations within evolving fields.

Model (AI Model)

A trainable computational architecture capable of registering and reproducing patterns across data. A model does not contain fixed forms; it encodes relations that can be activated through inference. In architectural workflows, models participate in shaping the field within which design decisions take place.

Morphogenesis

A term adopted from biology to describe processes through which form emerges through dynamic interactions within a system. In architectural discourse, morphogenesis often refers to parametric or generative

procedures that produce variation within a defined field. The essay contrasts this with inferential environments in which the field itself emerges through data-driven alignment.

Ontogenesis

Simondon's term for the coming-into-being of form and field together. Ontogenesis emphasizes processes through which structures emerge from interactions within a medium rather than being imposed from outside. Contemporary computational environments can be understood through this lens, as form stabilizes through recursive alignment across data and models.

Parametric Model

A computational design system in which geometry is generated through explicit relationships among defined parameters. Parametric models operate within a preconstituted field whose structure precedes variation. The essay distinguishes this regime from inferential systems, where relations themselves are learned rather than predefined.

Probabilistic Field

An environment structured by statistical tendencies rather than deterministic rules. In machine-learning contexts, outputs emerge through evaluation of likelihoods across learned relations. Architectural work within such fields involves orienting and stabilizing tendencies rather than executing fixed geometries.

Prompt

An input provided to a generative model that guides the direction of inference. Prompts function less as commands than as vectors that orient the model within a probabilistic field. They participate in concatenated sequences of interaction through which outputs gradually stabilize.

Stabilization

The process through which provisional coherence emerges within an evolving system. In inferential environments, stabilization occurs when alignments across data, models, and interpretation temporarily settle into recognizable form. Such stabilizations remain open to further transformation.

Technical Ensemble

A configuration of interacting technical and human components whose coordination produces a coherent effect. Rather than treating tools as isolated instruments, the essay frames architectural practice as engagement with such ensembles. Design unfolds through alignment across these distributed systems.

Transduction

Simondon's term for the propagation of structure across a system through successive transformations. A change in one domain induces adjustments in another, producing continuity of becoming across the ensemble. In computational environments, transduction can be observed in the way outputs feed back into datasets, models, and subsequent iterations.

REFERENCES

- Allen, Stan (2009), *Practice: Architecture, Technique and Representation*, 2nd ed., London: Routledge.
- Babbage, Charles (1864), "The Chess Player," in *Passages from the Life of a Philosopher*, London: Longman, Green, Longman, Roberts, & Green, pp. 465–471.
- Cache, Bernard (1995), *Earth Moves: The Furnishing of Territories*, trans. Anne Boyman, ed. Michael Speaks, Cambridge, Mass.: MIT Press.
- Carpo, Mario (2011), *The Alphabet and the Algorithm*, Cambridge, Mass.: MIT Press.
- Estévez, Alberto T. (2011), "Genetic Architectures: New Bio-digital Techniques," *Architectural Design*, 81, 6, pp. 54–61.
- Fuegi, John, and Francis, Jo (2003), "Lovelace & Babbage and the Creation of the 1843 'Notes'," *IEEE Annals of the History of Computing*, 25, 4, pp. 16–26.
- Hensel, Michael, Menges, Achim, and Weinstock, Michael (2010), *Emergent Technologies and Design: Towards a Biological Paradigm for Architecture*, London: Routledge.
- Lynn, Greg (1999), *Animate Form*, New York: Princeton Architectural Press.
- Menges, Achim (2007), "Computational Morphogenesis: Integral Form Generation," in Ahmad Okeil, Aghlab Al-Attili, Zaki Mallasi (eds.), *Em'body'ing Virtual Architecture, Proceedings of the 3rd International ASCAAD Conference*, pp. 725–744.
- Menges, Achim (2015), *Material Synthesis: Fusing the Physical and the Computational*, Chichester: Wiley.
- Morel, Philippe (2004), *Empiricism & Objectivity: Architectural Investigations with Mathematica*, Orléans: Editions HXX.
- Oosterhuis, Kas (2003), *Hyperbodies: Towards an E-motive Architecture*, Basel: Birkhäuser.

- Oxman, Neri (2010), "Structuring Materiality: Design Fabrication of Heterogeneous Materials," *Architectural Design*, 80, 4, pp. 78–85.
- Poe, Edgar Allan (1836), "Maelzel's Chess-player," *Southern Literary Messenger*, 2, 5, pp. 318–326.
- Simondon, Gilbert (2005), *L'individuation à la lumière des notions de forme et d'information*, Grenoble: Jérôme Millon.
- Simondon, Gilbert (2009), "The Position of the Problem of Ontogenesis," trans. Gregory Flanders, *Parrhesia*, 7, pp. 4–16.
- Simondon, Gilbert (2014), *Sur la technique*, Paris: Presses Universitaires de France.
- Simondon, Gilbert (2017), *On the Mode of Existence of Technical Objects*, trans. Cecile Malaspina and John Rogove, Minneapolis: Univocal.
- Simondon, Gilbert (2020), *Individuation in Light of Notions of Form and Information*, trans. Taylor Adkins, Minneapolis: University of Minnesota Press.
- Standage, Tom (2003), *The Turk: The Life and Times of the Famous Eighteenth-century Chess-playing Machine*, New York: Berkley Books.

Simon Weir*

ARCHITECTURE AND SURREALIST AUTOMATISM: FROM HISTORICAL MODELS TO AI SPECULATIONS

ABSTRACT: This paper develops a theoretical framework for understanding contemporary architectural automation through the historical lens of surrealist automatism. Tracing a shift from André Breton and Philippe Soupault's automatic writing to René Magritte's problem-pairing and Salvador Dalí's paranoiac-critical method, it argues that surrealist automatism progressively moved from motor execution toward perceptual, associative, and interpretive forms of cognition. On that basis, the paper identifies three modes of architectural automatism—architecture built with automatist construction, architecture modelled on the inner world, and architecture persuasively interpreted as surreal—using them to reassess contemporary AI-assisted design. Rather than treating such systems as neutral instruments of efficiency, the paper asks whether they remain tools governed by intention or whether they reorganise design thought in ways that exceed simple instrumentality. Surrealist automatism, it argues, offers architecture a critical framework for understanding automatism/automation as an opaque, generative and speculative force.

KEYWORDS: surrealism, automatism, architectural cognition, paranoiac-critical method, unconscious processes, generative AI, design methodology

* Simon Weir: School of Architecture, Design and Planning, The University of Sydney; simon.weir@sydney.edu.au.

This is an Open Access article under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not transformed in any way and is properly cited.

I. INTRODUCTION

Surrealist automatism offers architecture a theory of cognition that engages with perceptions, associations, and thoughts that operate prior to conscious intervention. Read in this way, automatism is not only a historical technique but a model for understanding how design may be extended by processes that exceed deliberate control. This perspective sharpens a contemporary question raised by AI-assisted design: do such systems remain instrumental tools, or do they alter design thinking in ways that make them something categorically different? Long before machine learning, surrealist artists engaged with misbehaving, opaque, and partially autonomous processes that redistributed agency and partly disabled intention. Reconsidered today, these practices allow automation to be understood not as an instrument of control, but as a speculative and creative condition that externalises cognition.

The paper proceeds in four steps. First, the paper reconstructs early surrealist automatism in André Breton and Philippe Soupault as grounded in psychological theories of the unconscious mechanisms of perception. Next, it shows how the translation of automatism into painting and architecture exposed the technical limits of this kind of automatism, and prompted translation into perceptual and interpretive automatism. Third, it examines Salvador Dalí's paranoiac-critical method and René Magritte's problem-pairing as models of relocated automatism. Finally, it uses these models to propose a framework for architectural automation as modes of externalised cognition, that is relevant today for engaging with AI-assisted design.

2. EARLY SURREALIST AUTOMATISM

Surrealism is the twentieth century art form most aptly associated with automatism. Surrealism began at the end of the First World War, then called the Great War, and was influenced by the burgeoning field of psychoanalysis seeking to understand the human psyche, specifically the part of the human psyche called the unconscious or the subconscious. Julien Levy, the legendary surrealist, writer and gallerist, wrote the book, *Surrealism*, in 1936, the year his Manhattan gallery held exhibitions by Magritte, Dalí, Leonor Fini and Max Ernst. Here, Levy wrote that surrealism's "fundamental doctrine is that poetry originates in, and appeals to, the subconscious."¹

¹ J. Levy, *Surrealism*, Black Sun Press, New York, 1936, p. 6.

On this topic, Levy quoted Dalí's unpublished 1934 lecture given in New York, at a museum that was then barely five years old, the Museum of Modern Art:

The subconscious has a symbolic language that is truly a universal language, for it [...] speaks with the vocabulary of the great vital constants, sexual instinct, feeling of death, physical notion of the enigma of space—these vital constants are universally echoed in every human.²

The date of surrealism's intertwining with automatism is plausibly 1919, when Breton and Soupault, months after returning home from military service, wrote a series of poems collected into a book they called, *Les Champs magnétiques*.³ Critic and author Edmond Jaloux reviewed the book a few months later, saying that it contained "charming phrases, obscure, beautiful, and sweet."⁴ As yet, no mention of automatism.

Next, in 1921 and 1922, Ernst travelled from Cologne to Paris and introduced Breton to Hans Prinzhorn's work, notably the 1922 book, *Artistry of the Mentally Ill*.⁵ Prinzhorn's study catalogued art made by psychiatric patients, and his text gave the images theoretical weight, treating these productions as aesthetic objects shaped by unconscious processes. He argued that Gestalt perception revealed unconscious organising forces at work in every act perception and creation. For Prinzhorn, the absent-minded doodling that produced these artworks emerged from an iterative series of unconscious interpretations that are present in any creative process. He wrote:

There is a psychological urge whereby „every shape, no matter how undefined and unobjective, demands interpretation“ [...] Only the individual stroke springs from a blind impulse, while the integration of the separate parts is guided by their author, no matter how inattentive he may be.⁶

² *Ibid.*, p. 7.

³ A. Breton, P. Soupault, *Les Champs magnétiques*, Au Sans Pareil, Paris, 1920.

⁴ L. Jenny, T. Trezise, "From Breton to Dalí: The Adventures of Automatism," *October*, 51, 1989, p. 106.

⁵ T. Bauduin, "Surrealism and Madness," *Papers of Surrealism*, 8, 2010, p. 1.

⁶ H. Prinzhorn, *Artistry of the Mentally Ill: A Contribution to the Psychology and Psychopathology of Configuration*, New York, Springer-Verlag, 1972, pp. 16–17.

Gestalt psychology proposed that the mind harbours unconscious organising mechanisms which guide how objects are perceived, for example the recognition of whole objects when only partial fragments are visible, and the recognition of groups and parts. Onto these mereological foundations, six unconscious drives operated on the objects and their parts, the urge to play, to interpret, to ornament, to order, to imitate, and to assign symbolic meaning. The description of these cognitive mechanisms *as unconscious* had two separate consequences for artistic automatism. First, aesthetic images that brought these unconscious drives to our attention were thereby revealing the unconscious by revealing the automatic actions of the mind. Second, the artist could be said to be operating automatically when they pursued absent-minded doodling.

Hearing of this from Ernst in 1922, Breton, now two years after writing the poems of *Les Champs magnétiques*, declared that they were the result of automatism. Art historian Laurent Jenny takes this rhetorical turn as a key moment in surrealist automatist theory, “the rhetoric of automatism... does not resemble any known rhetoric. This is because it inaugurates what one might call a *negative rhetoric*.”⁷ Automatist authors claim to *not* have used their conscious, logical, rational approaches. But what is never clear is what *exactly* they have and have not done. Subjectively, artists may recognise automatism as a practice, but objectively, automatism is “a practice of writing which is, in fact, unverifiable.”⁸

Nonetheless, the aim of figurative surrealism was to locate something pervasive, something more common and available than knowledge of European art history. Dalí’s speech at MoMA on the value of the unconscious as a subject for art is made in these terms: “To understand an aesthetic picture, training in appreciation is necessary, cultural and intellectual preparation. For surrealism the only requisite is a receptive and intuitive human being.”⁹ Magritte expressed the same sentiment when he wrote that the design of surrealist objects sought to convey the “feeling of the mysterious life of things (which should not depend on classic or literary allusions), which is only captured with a certain perceptiveness.”¹⁰ Yet while they both sought to create images appealing to all sensitive people, the theories they developed about their techniques and the

⁷ L. Jenny, T. Trezise, “From Breton to Dalí,” p. 107.

⁸ *Ibid.*

⁹ J. Levy, *Surrealism*, p. 7.

¹⁰ A. Danchev, *Magritte: A Life*, London, Profile Books, 2023, p. 297.

objects they designed were rooted in contemporary discourses of philosophy and psychology.

3. BRETON'S AUTOMATISM

Forms of literary automatism had been used by authors like Horace Walpole and Carlo Gozzi nearly two centuries earlier, but Breton's background in psychoanalysis afforded a new perspective.¹¹ Whereas contemporary spiritualists were using automatic writing to connect to spirits, independent entities, to other persons, Breton instead understood automatism as a means to contact and explore one's own unconscious.¹²

Breton learned about automatic writing during his military service. Working in a medical unit, he was exposed to what today may be called psychological trauma, and the new medical approaches at the time were based on psychoanalytic techniques.¹³ The author most associated with surrealism and psychoanalysis is Sigmund Freud,¹⁴ but other psychoanalysts and philosophers were referenced often in surrealist writings on creative automatism.

According to Soupault, Breton's automatic writing technique was most closely aligned with the work of psychoanalyst, Pierre Janet.¹⁵ Yet Breton, in his published writings, never mentioned Janet. There was, I suggest, a hint of Janet in Levy's remark that poetry appeals to the subconscious,¹⁶ "subconscious" is more Janet. Freud and Breton preferred "unconscious."¹⁷

According to psychiatry historian, Alexandra Bacopoulos-Viau, Breton did not acknowledge Janet's influence because "Janet's model

¹¹ A. Susik, "Chance and Automatism: Genealogies of the Dissociative in Dada and Surrealism," in D. Hopkins (ed.), *A Companion to Dada and Surrealism*, Chichester; Malden: Wiley, p. 402.

¹² K. Conley, "Surrealism and Outsider Art: From the 'Automatic Message' to André Breton's Collection," *Yale French Studies*, 109, 2006, p. 131.

¹³ J. Haan, P. J. Koehler, J. Bogousslavsky, "Neurology and surrealism: André Breton and Joseph Babinski," *Brain*, 2012, 15, 12, pp. 3830–3838; M. Polizzotti, *Revolution of the Mind: The Life of André Breton*, rev. ed. Boston, Black Widow Press, 2009, pp. 20–25.

¹⁴ For example: J. Levy, *Surrealism*, p. 5.

¹⁵ A. Bacopoulos-Viau, "Automatism, Surrealism and the Making of French Psychopathology: The Case of Pierre Janet" *History of Psychiatry*, 2012, 23, 3, p. 260.

¹⁶ J. Levy, *Surrealism*, p. 6.

¹⁷ The history of their use of these words is too complex for much certainty to be gleaned on this point. See: S. Gardner, "The Unconscious Mind," in T. Baldwin (ed.), *The Cambridge History of Philosophy 1870–1945*. Cambridge, Cambridge University Press, 2003, pp. 107–116.

of the mind would emphasise a struggle between two forces, synthesis (unity) and automatism (disintegration), with the former seen as ‘psychological strength’, the latter as ‘weakness’.¹⁸ Surrealism scholar, Abigail Susik extended this argument by contrasting Breton’s approach to art theorist Viktor Shklovsky’s notion of artistic defamiliarization. For Shklovsky, habits of perception produce perceptual automatization, and art’s function is to shake up this automatization to help reveal the real object beneath perception.¹⁹ Surrealist automatism, by contrast was the defamiliarization of this defamiliarization.²⁰ Breton instead considered automatism as a means of access to the inner psyche beneath social conditioning, beneath conscious intervention. Hence, surrealism scholar, Katharine Conley wrote that automatism is a “process that takes one down into the psyche, for Breton also constitutes a way back, as though in to a unique, original faculty to which easy access has been lost, except for those [...] outside culture.”²¹ Breton’s automatism was originally therapeutic, but the enduring interest was exploratory, aesthetic, political and philosophical.

Breton described the process of writing the poems of *Les Champs magnétiques* in the 1924 *Manifest du Surréalisme*. Here he called the poems “the first purely Surrealist work.”²² Breton described the process of generating these poems with Soupault as “systematic applications of automatic writing.”²³ Sometimes they wrote lines as quickly as possible, and passed the sheet of paper back and forth.²⁴ Sometimes they transcribed each other’s spoken words. When they pieced the texts together, there was no narrative, no logic, no thematic continuity. Soupault wrote that whatever rational approaches they already possessed regarding literary and poetic composition were irrelevant for the material they produced, so they understood themselves as being guided by unconscious

¹⁸ A. Bacopoulos-Viau, “Automatism, Surrealism and the Making of French Psychopathology: The Case of Pierre Janet,” p. 262.

¹⁹ A. Susik, “Chance and Automatism,” pp. 389–390.

²⁰ See here: D. Robinson, *Estrangement and the Somatics of Literature: Tolstoy, Shklovsky, Brecht*, Johns Hopkins University Press, Baltimore, 2008, p. vii.

²¹ K. Conley, “Surrealism and Outsider Art,” p. 136.

²² A. Breton, “Manifesto of Surrealism,” *Manifestoes of Surrealism*, University of Michigan Press, Ann Arbor, 1969, p. 37. For a detailed examination of Breton’s use of psychoanalytic language in his automatic processes, see: D. Lomas, *The Haunted Self: Surrealism, Psychoanalysis, Subjectivity*, Yale University Press, New Haven, 2000, and D. Lomas, *Simulating the Marvellous: Psychology, Surrealism, Postmodernism*, Manchester University Press, Manchester, 2013.

²³ A. Breton, “Manifesto of Surrealism,” p. 37.

²⁴ A. Susik, “Chance and Automatism,” p. 403.

preference.²⁵ Surrealism and automatism were here bound together and Breton defined surrealism in his manifesto as, “psychic automatism in its pure state, by which one proposes to express—verbally, by means of the written word, or in any other manner—the actual functioning of thought.”²⁶ Yet, any of surrealism’s many definitions need to be held lightly, because, as Levy, wrote:

Surrealism is not a rational, dogmatic, and consequently static theory of art—hence from a surrealist point of view there can be no accurate definition or explanation. The point of view is essentially anti-definitive and anti-explanatory. The surrealist object precedes definition.²⁷

Or, as we have seen, surrealist practice precedes definition. Indeed, at the end of the manifesto, Breton endorsed “automatic writing”²⁸ and far more general methods like “absent-mindedness” and “distraction” as modes of opening to the surreal unconscious.²⁹ If surrealism is a purely cognitive method, do the works of automatism have any definable and verifiable external qualities? Or is the difference only in the quality of the creative experience?

Breton was expansive in his definition of surrealism, “in any other manner,” yet how automatism could extend to complex problem solving like architectural design was far from clear. Breton’s frequently repeated automatist definition gives sufficient freedom for spoken words and poetry produced in trances or other mediumistic states, doodling, found-objects collaged together, but processes like architectural design, where the resulting object needed physical endurance, functional fit and structural soundness, require a large degree of reasoned premeditation and conscious control.

Christian Zervos recognised this tension in his 1928 article in *Cahiers d’Art*. While automatic writing could plausibly evade premeditation, painting and by extension architecture could never entirely eliminate technical and compositional forethought, making pure automatism technically impossible. Within surrealism’s proscribed definitions, “it would be impossible to conceive of it for a painted work. The elimination of a

²⁵ A. Breton, P. Soupault, *The Magnetic Fields*, Atlas, London, 1985, pp. 7–11.

²⁶ A. Breton, “Manifesto of Surrealism,” p. 26

²⁷ J. Levy, *Surrealism*, p.5.

²⁸ A. Breton, “Manifesto of Surrealism,” p. 37.

²⁹ *Ibid.*, p. 47.

picture's most essential preconceived pictorial elements seems a chimera and race towards nothingness."³⁰ For Zervos, surrealist automatism admitted text, but the physical execution of painting imposed technical challenges, and thereby theoretical obstacles, too insurmountable for the work to be authentically surreal. The elimination of the "most essential" elements also makes design impossible.

Then Breton began championing architecture, Ferdinand Cheval's *Palais Idéal*.³¹ Photography does not always convey the fact that the building is something of an architectural sculpture, not completely fulfilling that third element of the Vitruvius triad: utility. It is rather too small to use domestically for anything beyond storing tools. The surreality of *Palais Idéal* was twofold. First the only function was monumental, perhaps a monument to architectural uselessness as the aura of architecture was so present. Uselessness is a virtuous surrealist position—by rejecting habitual use, the object resists being possessed by humans, and attains more of an individual and unique identity, thereby appearing a little more natural and alive. Second, surreality of *Palais Idéal* arose from the charmingly absent-minded nature of its construction. Ferdinand Cheval was the town postman, and each day he would pick up a stone or two that he found while delivering the mail. After decades of construction without a plan, it was a huge architectural doodle, undeniably a strange and wondrous artwork. Yet by breaking architectural function, it was open to Zervos' objection as the elimination of an essential element to permit the nomination as surreal. Here, it seems, architecture had to be *broken* to become surreal.

In 1925, Breton acknowledged perspectives like Zervos': "It has been said that there could not be a Surrealist painting."³² But, Breton argued, "the mistake lies in thinking that the model can only be taken from the exterior world [...] The plastic work of art [...] will [...] refer to a purely interior model."³³ The interior model here has several potential meanings.

³⁰ C. Zervos, "Du Phénomène Surréaliste," *Cahiers D'Art*, 3, 1928, p. 114; "il serait impossible de le concevoir pour une œuvre peinte. L'élimination préconçue dans un tableau des éléments picturaux les plus essentiels me semble une chimère et une course au néant."

³¹ E. Flavell, "The Distance from Dream to Reality: Ferdinand Cheval in the Surrealist Imagination, 1924-1945," *Aspectus*, 3, 2021, pp. 1-20.

³² A. Breton, "Le Surréalisme et la peinture," *La Révolution Surréaliste*, 4, 1925, p. 27. Breton may also have been referring to M. Morise, "Les yeux enchantés," *La Révolution Surréaliste*, 1, 1, 1924, p. 16.

³³ C. E. Gauss, "The Theoretical Backgrounds of Surrealism," *The Journal of Aesthetics and Art Criticism*, 2, 8, 1943, p. 40, citing A. Breton, *Le Surréalisme et la Peinture*, p. 12.

While the context of surrealism rightly suggests that the “inner world” connotes the unconscious, Dalí would later associate this also with Kantian synthesis,³⁴ an association made often about Cubism in the 1920s.³⁵ There are too many categories and pure intuitions to elaborate here, but in brief, in *Critique of Pure Reason*, Kant argued that human experience was shaped *a priori* by the mind’s intuitions of space and time, and by categories including substance and causality. That is, we perceive objects in space and time, linked by cause and effect, not because reality *must* be this way, but because our cognitive apparatus imposes these conditions. Hence the unconscious inner world includes Kantian intuitions and categories, and hence, incidentally, why space-time (a soft watch) is a surrealist object: it is a Kantian intuition (part of the inner world) made into an object.

In the Manifesto, Breton gave a more direct example of what might be meant by the model of the inner world, wherein he described a delirious vision:

removed from the sound of any voice, a rather strange phrase [...] I cannot remember it exactly, but it was something like: “There is a man cut in two by the window,” but there could be no question of ambiguity, accompanied as it was by the faint visual image of a man walking cut half way up by a window.³⁶

Breton describes his personal predisposition towards the literary rather than the visual, but that immediately after this vision, “With a pencil and white sheet of paper to hand. I could easily trace their outlines. Here again it is not a matter of drawing, but simply of tracing.”³⁷ Can a complete work of architecture be experienced in a similar delirious vision? Capturing such visions is without a doubt a surreal enterprise, indeed, in the manifesto, Breton imagined the future possibility of a mechanical transcription of dreams, a collective recording device that might reveal latent structures in a shared unconscious.

³⁴ S. Dalí, “The Object as Revealed in Surrealist Experiment,” *This Quarter*, 5, 1, 1932, pp. 197–207 in H. Finkelstein (ed.), *The Collected Writings of Salvador Dalí*, Oxford, Oxford University Press, 1998. Dalí likely borrowed this interpretation from H. Kahnweiler, *Der Weg zum Kubismus*, Delphin-Verlag, München, 1920, p. 12.

³⁵ D. O’Brien, “Cubism and Kant,” in C. Vaughan, I. Vidmar (eds.), *Proceedings of the European Society for Aesthetics*, Vol. 10, The European Society for Aesthetics, Fribourg, pp. 482–506.

³⁶ A. Breton, “Manifesto of Surrealism,” pp. 21–22.

³⁷ *Ibid.*, p. 21.

Architecture that referred to a “purely interior model” was expressed later by Dalí in the architectural forms based on eggs, a stomach, a chrysalis and intra-uterine spaces, but it is clear that the physical automatism that began surrealist practice was no longer strictly necessary for the fabrication of surrealist artwork, thereby enabling the possibility of surreal architecture.³⁸

Once the theoretical obstacle of automatism controlling the designer’s hands had been lifted, alternative automatisms were developed by Dalí and Magritte. What emerged in Dalí and Magritte is not a rejection of automatism, but its relocation from the mechanics of the hand to the automatisms of perception, association and cognition. Dalí’s theories offered surreal architectural functions and proposals for architecture based on “interior models.” Magritte developed a surreality based on creative process based on the relationships between objects.

4. PARANOID AUTOMATISM

Dalí’s entry into figurative surrealism began around 1927 when he experimented with a painting technique related to Prinzhorn’s doodles.³⁹ Onto a flat and simple background, blue sky and grey concrete ground, dozens of loose brushstrokes are added. These strokes merge into fragmentary images of fear and desire. In the paintings of this style, *Honey is Sweeter than Blood* (1927), *Apparatus and Hand* (1927), and *Little Cinders* (1928), animals, bodies and faces seem to crystallise from improvised brush strokes.

Dalí’s mature entry into surrealism was marked by a more academic painting style, and the description of a paranoiac-critical method, a method derived as much from psychopathology as from Gestalt psychology. What emerged in Dalí’s paranoiac-critical method was not a rejection of automatism, but its relocation from the hand to perception and interpretation. Prinzhorn had shown how the mind possesses unconscious organising mechanisms capable of integrating stray fragments into coherent wholes. But Dalí grew restless with the logical, constructional aspect of this process. Doodling conveys an irrationality that comes together step by step, like detectives assembling fragments of evidence. This progressive accumulation could not account for visions and perceptions

³⁸ S. Weir, “Salvador Dalí’s Interiors with Heraclitus’s Concealment,” in G. Marinic (ed.), *Interior Architecture Theory Reader*, New York, Routledge, 2018, pp. 195–201.

³⁹ Notwithstanding several transitional works in 1928.

that appeared complete in their first instance. Sometimes too, the steps taken in the construction are not recognised as bearing any artistic intention at all. Dalí and Brassáï presented photographs of scrunched up and discarded bus tickets as “involuntary sculptures.”⁴⁰

Arguably, Dalí’s most significant contribution to surrealist theory was his work on paranoia. Breton was aligned with the eminent German psychologist Emil Kraepelin, but Dalí turned instead to Paul Sérieux and Joseph Capgras, and then to Jacques Lacan’s doctoral research for an anti-constructionalist model of paranoia.⁴¹ In the anti-constructionalist view, paranoia was not a reasoning madness, paranoid delirium is not pieced together like faulty logic, instead, unconscious delirium is embedded in perception. This shift was decisive. By arguing that paranoid mechanisms are conscious *and* unconscious, that meant that unconscious associations were necessarily built into every object in conscious perception, and thereafter every thought and experience. In the constructionalist mode, delusions are built sequentially, requiring interpretation and assembly. In the anti-constructionalist mode, the delusion is immediately coherent: perception itself is delirious, already charged with systematic associations, as Dalí wrote, “‘paranoiac phenomenon’ (delirium of systematic interpretation) is consubstantial with the human phenomenon of sight.”⁴²

The paranoid-critical method was thereby not surrender to the unconscious, but its conquest. Irrational perception made concrete for artistic effect. An imagined object that is automatically, paranoiacally, unconsciously perceived is presented to the designer’s mind who critically intervenes by making it concrete, by representing and fabricating it, etcetera, and the object is herein designed so that the audience experiences inspiring subconscious stimulation. By locating *and* activating unconscious mechanisms of perception, as Charles Gauss wrote for *The Journal of Aesthetics and Art Criticism*:

Surrealist works are not important as poems, as pictures, as objects, but as being the residue left when we have stripped down our souls to the bare framework of the unconscious which is beneath all our selves and from which we never escape.⁴³

⁴⁰ See K. Conley, *Surrealist Ghostliness*, University of Nebraska Press, Lincoln, 2013, p. 69.

⁴¹ S. Weir, “Surrealist Architecture: Dalí’s 1958 *Crisalida*, San Francisco,” *Journal of Surrealism and the Americas*, 13, 1, 2022, pp. 118–120.

⁴² S. Dalí, *Dalí, Dalí!*, Julien Levy Gallery, New York, 1939, s. p.

⁴³ C. E. Gauss, “The Theoretical Backgrounds of Surrealism,” p. 39.

When describing his painting technique to a *New York Times* reporter, Dalí explained, “I do all my work subconsciously [...] I never use models or paint from life or landscapes. It is all imagination. That is, I see everything in a dream as I am working.”⁴⁴ When Dalí said, “all my work,” he did not mean that buying paint and cleaning the brushes was subconscious, but the design was subconscious, the selection and form of the subjects. On this point, the painter that Dalí regarded as “highly educational,”⁴⁵ was Magritte, and his creative method, described later, was not dissimilar. For the paintings of the 1930s, Dalí’s normal practice was still to paint landscapes and then locate “in a dream as I am working” surreal objects to add on top of the painted landscape. These new surreal objects however were not comprised of visibly doodled brush strokes, instead the component pieces are invisibly small. Therefore, Dalí referred to his famous soft watches as colloids.⁴⁶ Cheese, milk, fog and oil paint are examples of colloidal systems. This meant that the painted objects did not need to appear to be doodled, but could appear as compact and coherent objects irrationally (subconsciously) conceived.

The difference between passive automatism and active paranoia is exemplified in different readings of Art Nouveau architecture. Breton wrote: “What is Art Nouveau if not an attempt to generalise and adapt mediumistic drawing, painting and sculpture to the art of furniture and decoration?”⁴⁷ Dalí instead aligned Art Nouveau architecture with surrealist function, as accommodating irrational desires rather than opine on the mental states of the designers or builders.⁴⁸ Across both perspectives is a foundation for surreal architecture that is secure and transferable: the surrealists’ abilities to perceive unconscious agendas in the works of others. Art Nouveau’s architects, and Ferdinand Cheval, did not call themselves surrealists. Surrealists persuaded us to see that these architects were acting automatically, unconscious of their own desires.

⁴⁴ Anonymous, “Salvador Dalí Arrives: Surrealist Painter Brings 25 of His Pictures for Show Here,” *New York Times*, Nov. 15, 1934, p. 22.

⁴⁵ D. M. Johnson, “Influence, Dialog, Rivalry,” in *Dalí & Magritte*, Ludion, Bruxelles, 2019, p. 73.

⁴⁶ S. Dalí, *Notes pour l’interprétation du tableau / “La Persistance de la mémoire,”* unpublished manuscript, 1931, Gala-Salvador Dalí Foundation collection, Centre for Dalinian Studies, Figueres, s. p.

⁴⁷ D. Roberts, “Surrealism and Natural History: Nature and the Marvelous in Breton and Caillois,” in D. Hopkins (ed.), *A Companion to Dada and Surrealism*, Wiley, Chichester/Malden, 2016, p. 290.

⁴⁸ S. Dalí, “De la beauté terrifiante et comestible de l’architecture modern style,” *Minotaure*, 3–4, 1933, p. 72.

This process of analysing architecture such that unconscious intentions are seen as the main feature of the design became the model that Rem Koolhaas would use in *Delirious New York*. Accordingly, Koolhaas posited that *all architecture* is “inevitably” the result of the paranoiac critical method.⁴⁹ Here Koolhaas astutely points towards the psychoanalytically technical yet ultimately very general way that the paranoiac critical methods were defined and described, and the method’s tendency to resemble idea generation in general, rather than anything specifically *surreal*. Anyone having a vision of what their project might look like is having a surrealist moment, especially if they do their best to transcribe that vision.

5. PAIRING PROBLEMS

Around 1925, Magritte found surrealism, and moved to Paris in 1927 to work with Breton in the growing surrealist movement. That year he painted the first version of *La Clef des songes*, foregrounding odd combinations of words and images. Magritte painted many similar paintings with this name, all with simple, diagram-like compositions. Typically, a flat grey background is framed with four or six equal rectangles. In each rectangle is a familiar object like a shoe or a candle, painted in a direct and realistic style. A couple of words, written in flowing handwriting, are underneath each object. Underneath the leaf, for example, is written the French for “table;” underneath the suitcase, “sky.”

Over the next year, Magritte developed a series of eighteen relations between words and images, which he prepared as *Les Mots et les Images*, published in *La Révolution Surréaliste*.⁵⁰ In each of the eighteen relations, there are some qualities that repeat in nearly all. Typically, an image and a word are paired. And there are no words like “running” or “justice,” but words of objects that *could* be depicted in the adjacent image, like “sun,” and “horse.”⁵¹

Examining each of the eighteen possible relations that Magritte identified, and then their application into architecture might be a fruitful

⁴⁹ R. Koolhaas, *Delirious New York*, Oxford, Oxford University Press, 1978, p. 246. See: S. Weir, “Salvador Dali in Rem Koolhaas’ *Delirious New York*,” *The Journal of Architecture*, 27, 2–3, 2022, pp. 398–419.

⁵⁰ R. Magritte, “Les Mots et les Images,” *La Révolution Surréaliste* 12, 1929, pp. 32–33.

⁵¹ Michel Foucault analysed some of these conjunctions in his 1973 monograph on Magritte, *This Is Not a Pipe*.

project, but for this essay on automatism, the focus is on Magritte's process. When pairing objects, Magritte described the first object as a question, and the design problem was to find "another object as answer, secretly connected to the first by links which are sufficiently complex to serve as verification of the answer."⁵² For example, Magritte's answer for the problem of the grand piano, was a giant engagement ring large enough to surround the piano like an architectural frame. Magritte believed that there was some innate psychological connection between the two objects based on affinities. Consider next his written explanation, sent to Breton, on a breakthrough in the process.

One night in 1936 I woke up in a room where there happened to be a bird sleeping in a cage. A splendid misapprehension made me see the cage with the bird gone and replaced by an egg. I had grasped a new and astonishing poetic secret, because the shock I experienced was caused precisely by the affinity between the two objects: the cage and the egg, whereas previously I had provoked the shock by bringing together totally unrelated objects.

Preceding from this revelation, I tried to discover if objects other than the cage, thanks to the pinpointing of some elements peculiar to them, and a strictly predestined part of them, could not display the same obvious poetic quality that the egg and the cage had achieved through being brought together [...]

In the course of my researches, I became certain that the element to be discovered, the unique feature residing obscurely in each object, was always known me in advance, but that any knowledge of it was, so to speak, hidden in the depths of my thought.⁵³

Magritte's revelation of 1936 was that the second object can hold a sympathy towards the first, rather than an antipathy. Notable is Magritte's repeated assertion about the location of the second object, the answering

⁵² R. Magritte, *Selected Writings*, Surrey, UK, Alma Books, 2016, p. 140.

⁵³ R. Magritte, D. Sylvester, S. Whitfield, *René Magritte: Catalogue Raisonné, Volume 2. Oil Paintings and Objects, 1931–1948* The Menil Foundation, , Houston, 1993, pp. 16–17. There is reason to be skeptical of Magritte's account here, as the story of waking to imagine an egg inside a cage, Magritte wrote that this happened in 1936, but the painting was painted in 1933, P. Allmer, *Critical Lives: René Magritte*, Reaktion Books, London, 2019, pp. 176–178.

object, as within something unconscious, as somehow “hidden” inside his mind, hidden in thought, hence its surreality.

Applying this to architecture is a matter of framing the design questions accordingly, selecting first objects in need of a second object, or in need of a particular quality that can be conveyed with an object—insofar as a quality, blue, for example, may also be an object, blue paint. One of my own recent experiments in architectural speculation used Magritte’s problem-pairing for text-to-image AI. The first object, the architectural “problem,” was theatre and the answering object, shipwreck, producing unprecedented architectural forms based on latent affinities between the words.⁵⁴ This approach functions as a contemporary analogue to Magritte’s own “hidden” answers, an automatism shared between human ideation and the collected knowledge of a large language model and stable diffusion model. This might open another zone of surreality for architectural design, if the deep engagement with external cognition does not disqualify it from authentic human surreality.

The authenticity question may be resolved by substituting the larger question of whether or not we consider our tools to be coauthors, ink bleeds into paper to be creditable contributors to the watercolour, or the patterns in the timber and stone to be contributor to the architectural design. AI does possess an unconscious in the psychoanalytic sense, but their language contains latent, unexpressed quantities beyond human reason. But complexity is no measure of consciousness, these models are puppets. It is, in my view, an illusion that they appear to be casual agents in the data they generate, just as it is an illusion to say that the gun is the causal agent when a shot is fired.

6. SURREAL AUTOMATISM IN ARCHITECTURAL DESIGN

The history of early surrealist automatism shows that there are several ways that architectural design can engage automatist surrealism. Since surrealism is a diverse practice that evades definition, and automatism is a somewhat blurry concept—as many different parts of our cognition can be considered automatic—some kinds of automatism are contrary and/or incompatible with some others, but the method is reliable and effective for unsettling creative habits, and generating novel and authentic ideas.

⁵⁴ S. Weir, S. Rich, “Shipwreck Hauntography,” *Footprint*, 35, 2024, pp. 51–66.

There are three fluid kinds of automatist surreal architecture.

1. Architecture built with automatist construction.
2. Architecture modelled on the inner world.
3. Architecture persuasively interpreted as surreal.

6.1. The first kind of surreal architecture are the buildings constructed without a design and assembled in an automatist manner, like Ferdinand Cheval's constructions where the design was unplanned and the result of innumerable spontaneous acts. Both Dalí's 1928-1929 paintings and Kurt Schwitters's *Merzbau* fit this schema.⁵⁵ Bird's nests, bee hives, termite mounds and similar forms might also be included in this category if we push surreality out beyond human participation.

All these works can be said to legitimately harbour evidence of the designer's unconscious mechanisms because every part of the building was materially set in place with the designer's hands. Two questions follow. Is a work of architecture automatist if one part of the design work was automatist? And, can architecture be correctly considered automatist surrealism if the architect used automatist means and then another more reasonable person constructed the edifice? This would depend if the architect could specify automatist elements into the construction. The clearest example of this is the exquisite-corpse method, where typically four people contribute a quarter of the design of a human figure. In the end, a singular figure appears that is independent of the intentions of any of the individual contributors. The autonomy of the figure is also significant for the person who architects their chosen instance of the exquisite corpse exercise, selecting the contributors and the subject (subjects other than human figures are all equally plausible.) This is arguably another form of surreal autonomous art making. If the architect saw in this object a deep resonance with their own inner world, who could refuse the process? Yet whenever we step beyond pure automatist construction, we are entering diluted automatisms, legitimate automatisms but very diluted, making misdiagnosis more likely.

If we extend cognition outwards to include artificial intelligence partners, the use of Magritte's pairings as prompts for expansion and representation become more intriguing. Whether or not these processes can

⁵⁵ See: N. Spiller, *Architecture and Surrealism*, Thames & Hudson, London, 2016, pp. 26–29.

be understood as having any significance for the exploration of the unconscious psyche is debatable. It is certainly arguable that such processes are not capable of foregrounding latent knowledge within the designer, and are only capable of revealing latent knowledge within the language model, and thereby of revealing profound associations between concepts and knowledges that it has a unique capability to identify, compress and explain. In this way, models might be capable of producing new knowledge but the prompts might need some surrealist prompting to push this hidden knowledge to the surface. However, as models collect greater amounts of data from individual users, an AI model may be able to identify patterns of data in users consisting latent knowledge. Indeed, the capacity of AI to act as a form of extended memory is one of the most intriguing possibilities of not just using but living for decades with AI

6.2. The second kind of surreal architecture looks to the inner world as a model. Within this notion of the inner world are many aspects. In Breton's original model, automatic writing necessarily reveals the inner world, but the idea has three distinct branches.

6.2.1. The first branch of the internal model is the principle that architectural function is (or, perhaps, should be) aligned with human desires, both conscious and unconscious. Accordingly, architecture should respond to unconscious desires. Thomas Mical argued for surrealism's relevance to architecture theory as a countermeasure to the totalising power of modernism, which sought to rationalize away the irrational aspects pre-modern architecture.⁵⁶

6.2.2. The second branch of the internal model accounts for the transcription of visions and dreams. Whether or not the reader would regard designing a building as faithfully as possible after seeing it in a dream as automatist or not, there is inarguably some automatic subconscious design work occurring in the process. Once we recognise the fact that the design and fabrication of buildings and paintings are (almost) never a single creative experience but are instead a series of many kinds of work, the designation of a project to be surreal or not based on the amount of automatism in the process becomes impossibly fraught.⁵⁷

⁵⁶ T. Mical, *Surrealism and Architecture*, Routledge, London, 2004, p. 1.

⁵⁷ As noted by Magritte's friend, Paul Nougé, see: A. Danchev, *Magritte: A Life*, p. 175.

6.2.3. The third branch is Magritte's process of seeking answer objects within thought. Relatedly, the inner model for Dalí has three vital constants: the presence of death which can be seen in decaying objects; the anxiety of space-time shown as objects experiencing causal forces, like cheese *melting*, architectural beams bending and sagging; and the sexual instinct shown in the forms of human bodies and their varied appearance through pareidolia.⁵⁸ Across all these perspectives, the key observation is the great variety of ideas that can be drawn from the inner world.

Architecture, especially public architecture, too should be understood as attempting to appeal to, and engage meaningfully with, many universal or pseudo-universal human values: the human scale of objects and the space needed to comfortably perform bodily activities, still and windy, wet and dry, quiet and noisy, sunny and shady, warm and cold, cosy and bare, clean and dirty. Almost all works of architecture possess a sunny side and a shaded side, sunrise and sunset sides, these too I would call vital constants of architecture, to which people have fairly predictable unconscious responses. The placedness of objects was considered by Magritte too. When considering the object pairings, he wrote that there were not two, but "three points of reference: the object, the something linked to it in the obscurity of my consciousness and the light into which this something had to be brought."⁵⁹ So, if we accept the notion that surreality addresses vital constants, all architecture that responds to the site in a manner that is experienced unconsciously by the inhabitants, is thereby surreal; and here we arrive at Koolhaas' observation that surrealism seems to encompass all aspects of architecture, knowingly surreal or not.

6.3. The third kind of surrealist architecture is any work of architecture persuasively interpreted as surreal. This is aligned with the foundational notion of Koolhaas, but very infrequently executed successfully. Breton and Dalí provided exemplars with their writings on Art Nouveau. Such

⁵⁸ For more on this topic, see: C. Wang *et al.* "Pareidolia in a Built Environment as a Complex Phenomenological Ambiguous Stimuli," *International Journal of Environmental Research and Public Health*, 19, 9, 2022; and S. Chalup K. Hong, M. J. Ostwald, "Simulating Pareidolia of Faces for Architectural Image Analysis," *International Journal of Computer Information Systems and Industrial Management Applications*, 2, 2010, pp. 262–278.

⁵⁹ R. Magritte, D. Sylvester, S. Whitfield, *René Magritte: Catalogue Raisonné. Volume 2*, p. 17.

interpretation can be laborious and research intensive, or spontaneous and unpremeditated, like using a door handle as a coat hook. Interpreting architecture as having a surreal function is a strong mode of interpretation. The automatism of such surreality is variable.

An extended third kind I consider most remarkable where it can be made to connect to 6.1. (automatist technique) is generative text-to-image AI. To include this in the list of automatic design methods is contentious, as it requires an extended model of cognition that includes an expanded unconscious. What is at stake here is not whether such systems possess autonomy or intention, but whether they can still be understood as purely instrumental—executing predefined ends—or whether they instead operate as cognitive systems that reorganise perception, association, and design thinking in ways that exceed their use as tools. My speculation is that their broad acceptance as extended cognition is the probable outcome. Several scenarios make the acceptance of surrealist imagery made with generative AI seem plausible:

6.3.1. Someone dreams of an extraordinary and moving image in unusually fine detail then during the morning's work prompts the image's appearance into existence.

6.3.2. Someone sees an image and is enamoured with it, then later learns that it was generated AI, and in spite of the disapproval of those around her, who say AI cannot make art, she continues to adore, and be moved by, the image. This is an example of making an object surrealist through interpretation.

6.3.3. An artist assembles a collage and then later learns that one of the images used was made by AI. The mereological problem of how parts contribute to the identity of a whole is somewhat disabling in examples like these. Though a surrealist flourish in an ornamental detail of an otherwise entirely conscious crafting of an image is unlikely to be granted the surrealist nomination.

6.3.4. An artist makes an image of an object in AI, then fabricates an object that corresponds to the image.

6.3.5. An artist takes an AI image and claims it, in the manner of Marcel Duchamp, as a ready-made artwork, seeking recognition for the idea.

6.4. There are very many plausible situations where AI is bound up into some surreal scenario. As far as I know it has not yet happened, but it will eventually occur that someone builds a building to look like its original AI perspective. Many may call this “surreal” but whether or not its purported surreality will endure is another question, and it seems unlikely if the process becomes common. Surreality in architecture through automatism and pseudo-automatist processes, given the anticipation of the increased automation of design through AI, will expand architectural cognition into all the kinds of surrealist architectures with new unconsciously satisfying affordances.

The future of architectural automatism need not lie in replicating historical techniques, nor in attributing consciousness to machines, but in extending cognition, blurring the boundary between tool and collaborator. As new tools externalise intuition, expand perceptual possibility, and generate forms that exceed individual intention, they revive the surrealist ambition to expose the hidden determinants of thought and harness them for architectural cognition. The architect’s role will remain based in selective judgment, translation of ideas into appropriate form, and critical testing of options and proposals. Whether mediated through human hands, collective practices, or language and diffusion models, automatism persists as an engine of estrangement, an invitation to design with, against, and beyond ourselves. Architecture becomes surreal not because it imitates past methods but because it continues to discover new ways to be designed, new ways to be interpreted, and new unconscious desires to accommodate.

REFERENCES

- Allmer, Patricia (2019), *Critical Lives: Rene Magritte*, London: Reaktion Books.
- Anon. (1934), “Salvador Dali Arrives: Surrealist Painter Brings 25 of His Pictures for Show Here,” *New York Times*, Nov. 15, p. 22.
- Bacopoulos-Viau, Alexandra (2012), “Automatism, Surrealism and the making of French psychopathology: the case of Pierre Janet,” *History of Psychiatry*, 23, 3, pp. 259–276.
- Bauduin, Tessel (2010), “Surrealism and Madness,” *Papers of Surrealism*, 8, pp. 1–6.
- Breton, André, Soupault, Philipe (1920), *Les Champs magnétiques*, Paris: Au Sans Pareil.
- Breton, André (1925), “Le Surréalisme et la peinture,” *La Révolution Surréaliste*, 4, pp. 27–29.

- Breton, André (1969), "Manifesto of Surrealism," in *Manifestoes of Surrealism*, trans. Richard Seaver and Helen R. Lane, Ann Arbor: University of Michigan Press, pp. 1–47.
- Breton, André, and Soupault, Philipe (1985/1920), *The Magnetic Fields*. London: Atlas.
- Chalup, Stephan K., Kenny Hong, Michael J. Ostwald (2010), "Simulating Pareidolia of Faces for Architectural Image Analysis," *International Journal of Computer Information Systems and Industrial Management Applications*, 2, pp. 262–278.
- Conley, Katharine (2006), "Surrealism and Outsider Art: From the 'Automatic Message' to André Breton's Collection," *Yale French Studies*, 109, pp. 129–143.
- Conley, Katharine (2013), *Surrealist Ghostliness*, Lincoln: University of Nebraska Press.
- Dalí, Salvador (1931), *Notes pour l'interprétation du tableau / "La Persistance de la mémoire"*, unpublished manuscript, Figueres: Gala-Salvador Dalí Foundation collection, Centre for Dalinian Studies.
- Dalí, Salvador (1933), "De la beauté terrifiante et comestible de l'architecture modern style," *Minotaure*, 3–4, pp. 69–76.
- Dalí, Salvador (1932), "The Object as Revealed in Surrealist Experiment," *This Quarter*, 5, 1, pp. 197–207.
- Dalí, Salvador (1939), "*Dalí, ! Dalí!*" New York: Julien Levy Gallery.
- Danchev, Alex, Sarah Whitfield (2023), *Magritte: A Life*, London: Profile Books.
- Finkelstein, Haim (ed. and trans.) (1998), *Collected Writings of Salvador Dalí*, Oxford: Oxford University Press.
- Flavell, Ella (2021), "'The Distance from Dream to Reality': Ferdinand Cheval in the Surrealist Imagination, 1924–1945," *Aspectus*, 3, pp. 1–20.
- Foucault, Michel (1983), *This is Not a Pipe*, Berkeley: University of California Press.
- Gardner, Sebastian (2003), "The Unconscious Mind," in Thomas Baldwin (ed.), *The Cambridge History of Philosophy 1870–1945*, Cambridge University Press, pp. 107–116.
- Gauss, Charles. E. (1943), "The Theoretical Backgrounds of Surrealism," *The Journal of Aesthetics and Art Criticism*, 2, 8, pp. 37–44.
- Haan, Joost, Peter J. Koehler, Julien Bogousslavsky (2012), "Neurology and surrealism: André Breton and Joseph Babinski," *Brain*, 15, 12, pp. 3830–3838.
- Jenny, Laurent & Trezise, Thomas (1989), "From Breton to Dalí: The Adventures of Automatism," *October*, 51, pp. 105–114.
- Johnson, Danielle M. (2019), "Influence, Dialog, Rivalry," in *Dalí & Magritte*, Bruxelles: Ludion, pp. 57–75.
- Kahnweiler, Henri (1920), *Der Weg zum Kubismus*, München: Delphin-Verlag.
- Koolhaas, Rem (1978), *Delirious New York*, Oxford: Oxford University Press.
- Levy, Julien (1936), *Surrealism*, New York: Black Sun Press.
- Lomas, David (2000), *The Haunted Self: Surrealism, Psychoanalysis, Subjectivity*. New Haven: Yale University Press.
- Lomas, David (2013), *Simulating the Marvellous: Psychology, Surrealism, Post-modernism*. Manchester: Manchester University Press.

- Magritte, René (1929), "Les Mots et les Images," *La Révolution Surréaliste* 12, pp. 32–33.
- Magritte, René; Sylvester, David & Whitfield, Sarah (1993), *René Magritte: Catalogue Raisonné, Volume 2. Oil Paintings and Objects, 1931–1948*, Houston: The Menil Foundation.
- Magritte, René (2016), *Selected Writings*, ed. Kathleen Rooney, Eric Plattner, trans. Jo Levy, Surrey: Alma Books.
- Mical, Thomas (2004), *Surrealism and Architecture*, London: Routledge.
- Morise, Max (1924), "Les yeux enchantes," *La Revolution Surrealiste*, 1, p. 16.
- O'Brien, Dan (2018), "Cubism and Kant," in Connell Vaughan, Iris Vidmar (eds.), *Proceedings of the European Society for Aesthetics*, Vol. 10, Fribourg: The European Society for Aesthetics, pp. 482–506.
- Polizzotti, Mark (2009), *Revolution of the Mind: The Life of André Breton*, rev. ed., Boston: Black Widow Press.
- Prinzhorn, Hans (1972/1922), *Artistry of the Mentally Ill: A Contribution to the Psychology and Psychopathology of Configuration*, New York: Springer Verlag.
- Roberts, Donna (2016), "Surrealism and Natural History: Nature and the Marvelous in Breton and Caillois," in David Hopkins (ed.), *A Companion to Dada and Surrealism*, Chichester; Malden: Wiley, pp. 287–303.
- Robinson, Douglas (2008), *Estrangement and the Somatics of Literature: Tolstoy, Schklovsky, Brecht*, Baltimore: Johns Hopkins University Press.
- Spiller, Neil (2016), *Architecture and Surrealism*, London: Thames & Hudson.
- Susik, Abigail (2016), "Chance and Automatism: Genealogies of the Dissociative in Dada and Surrealism," in David Hopkins (ed.), *A Companion to Dada and Surrealism*, Chichester; Malden: Wiley, pp. 388–410.
- Wang, Chen, Liangchen Yu, Yiyi Mo, Lincoln. C. Wood, Carry Goon (2022), "Pareidolia in a Built Environment as a Complex Phenomenological Ambiguous Stimuli," *International Journal of Environmental Research and Public Health*, 19, 9, 5163.
- Weir, Simon (2018), "Salvador Dalí's Interiors with Heraclitus's Concealment," in Gregory Marinic (ed.), *Interior Architecture Theory Reader*, New York: Routledge, pp. 195–201.
- Weir, Simon (2022), "Surrealist Architecture: Dalí's 1958 Crisalida, San Francisco," *Journal of Surrealism and the Americas*, 13, 1, pp. 112–131.
- Weir, Simon (2022), "Salvador Dalí in Rem Koolhaas' Delirious New York," *The Journal of Architecture*, 27, 2–3, pp. 398–419.
- Weir, Simon, Sara Rich (2024), "Shipwreck Hauntography," *Footprint*, 35, pp. 51–66.
- Zervos, Christian (1928), "Du Phénomène Surréalistes," *Cahiers D'Art*, 3, pp. 113–114.

Alisa Andrasek*

ARCHITECTURE AS DESIGN COGNITION: SYNTHETIC INTELLIGENCE, AESTHETIC JUDGMENT, AND THE OPERATIVE TURN

ABSTRACT: This paper proposes design cognition as a conceptual framework for understanding architecture in the age of synthetic intelligence. Rather than framing artificial intelligence in terms of automation, efficiency, or authorship, it argues that architecture is undergoing a deeper epistemic shift: from a representational discipline to an operative cognitive practice. What is at stake is not the adoption of new tools, but a reorganisation of where and how architectural thinking itself occurs. Design cognition emerges neither within a singular human subject nor inside computational systems alone, but through distributed cognitive ecologies composed of human judgment, synthetic representations, material behaviours, and environmental processes. Drawing on computation theory, philosophy of mind, and the Agentic Architecture project exhibited at the 2025 Venice Architecture Biennale, the paper reframes creativity as the navigation of high-dimensional possibility spaces rather than expressive authorship. It introduces aesthetic cognition as an irreducible architectural capacity: an ability to perceive, project, and evaluate spatial futures that cannot be fully formalised or automated. By situating architecture within multi-scalar planetary systems, it argues that design becomes an epistemic actor capable of participating in the recomputation of material, energetic, and informational relations.

KEYWORDS: design cognition, synthetic intelligence, aesthetic judgment, architectural epistemology, distributed agency, complex systems, planetary design

* Alisa Andrasek: Royal Melbourne Institute of Technology; alisa.andrasek@rmit.edu.au.

This is an Open Access article under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not transformed in any way and is properly cited.

I. INTRODUCTION: BEYOND AUTOMATION

The contemporary discourse on automation in architecture remains largely misaligned with the transformation it seeks to describe. Debates continue to orbit questions of efficiency, authorship, and substitution; whether machines accelerate design processes, replace human labour, or automate decision-making. While not insignificant, these concerns fail to address a more fundamental shift currently underway. What is changing is not simply how architecture is produced, but where and how architectural thinking itself takes place.

Architecture is entering a phase in which design can no longer be understood as a cognitive activity confined to the mind of a single human subject, nor as the execution of predefined rules by automated systems. Instead, architectural practice is increasingly organised around synthetic cognitive substrates—computational, material, environmental, and ecological systems that participate directly in the generation, evaluation, and transformation of spatial possibilities. The architectural automaton is neither a tool nor an autonomous subject, but a cognitive ecology: a coupled system in which intention, noise, constraint, and emergence co-produce form.

This paper argues that the decisive shift is not from human to machine agency, but from design as representation to design as cognition. Architecture becomes a mode of thinking enacted through distributed systems—human and inhuman alike—operating across scales, temporalities, and perceptual thresholds that exceed individual comprehension. Within this framework, creativity is not diminished by synthetic intelligence. On the contrary, it is intensified and displaced—relocated from expressive authorship toward the navigation of high-dimensional possibility spaces where architecture operates as an epistemic practice in its own right.

It is worth stating what this argument is not. It does not propose the disappearance of the architect, the irrelevance of authorship, or the inevitability of technical optimisation as design's horizon. These framings, which recur in both enthusiastic and critical responses to computational architecture, share a common misreading: they treat synthetic intelligence as a substitution machine rather than as an expanded cognitive field. The question is not whether machines can design, but how design intelligence is redistributed across new cognitive ecologies—and what kinds of judgment, responsibility, and imagination this redistribution enables and demands.

2. FROM REPRESENTATION TO OPERATION

Architecture has always been entangled with complexity. Historically, however, this entanglement has been managed primarily through representation. Drawings, diagrams, models, visualisations, and abstractions rendered complex phenomena legible to human perception. Such representations functioned as cognitive compressions: they translated vast material, social, and environmental systems into manageable visual or symbolic forms. The drawing was not simply a record of intention—it was the medium through which intention became possible.

Today, this representational paradigm reaches a limit it cannot exceed by refinement alone.

Contemporary architectural conditions are shaped by processes that resist direct depiction: climatic feedback loops, energetic gradients, algorithmic infrastructures, biological metabolisms, temporal dynamics unfolding across vastly different orders of scale. To merely represent these systems is no longer sufficient. Architecture increasingly operates within complexity rather than about it. The shift is not merely technological. It is epistemic.

This marks what might be called an operative turn. Design moves from describing complexity to constructing conditions under which complexity can be explored, negotiated, and transformed. As Luciana Parisi argues, computation operates not only as discrete calculation but as a generative and relational field capable of producing novel configurations through interaction.¹ Recursive computational workflows, algorithmic combinatorics, simulation-driven synthesis, and multi-scalar modelling do not simply visualise systems; they actively participate in their evolution. Architecture becomes less a mirror of the world and more a synthetic environment in which new configurations can be tested and brought into being.

The operative turn also alters the temporality of design. Instead of a linear sequence in which analysis precedes form and performance is tested after the fact, design becomes recursive. Simulation, evaluation, and generation feed back into one another. Form is no longer an endpoint but a

¹ L. Parisi, *Contagious Architecture: Computation, Aesthetics, and Space*, MIT Press, Cambridge, Mass., 2013. Parisi's account of algorithmic computation as a generative and relational field—capable of producing novelty through interaction rather than through the execution of fixed rules—provides theoretical grounding for the operative turn described in this paper.

temporary crystallisation within a continuous field of negotiation. Representation does not disappear in this condition—it becomes one operative layer among many, participating in cognition rather than concluding it.

A further consequence concerns the invisible. Much of what shapes the built environment—microclimatic flows, energetic exchanges, biochemical processes, algorithmic operations—remains inaccessible to unaided perception. Synthetic intelligence enables these invisible dynamics to become operational: to be sensed, inferred, simulated, and transformed into generative dimensions of design. Forces that were once treated as background—pressure gradients, thermal differentials, wave dynamics—become primary materials. Architecture extends beyond the limits of the human sensorium not by abandoning legibility, but by developing new instruments for perceiving and acting within what was previously opaque.

3. SYNTHETIC DESIGN COGNITION AND THE EXTENDED MIND

If architecture is no longer primarily representational, the central question becomes: where does design thinking now occur?

Design cognition, as it emerges today, cannot be located exclusively in the mind of the architect, nor attributed solely to computational systems. It arises through recursive coupling between human intention, synthetic representations, and evaluative processes distributed across multiple agents and substrates. This is not a metaphor. It is a structural description of how contemporary architectural practice actually works.

Andy Clark and David Chalmers proposed that cognitive processes do not end at the boundary of the brain—or even the body.²³ When external resources are reliably available, actively used, and directly coupled to action, they function as genuine components of a cognitive system rather than mere aids to it. What changes with synthetic intelligence is not the principle—architecture has always extended cognition into

² A. Clark, D. Chalmers, “The Extended Mind,” *Analysis*, 58, 1, 1998, 7–19. Clark and Chalmers argue that cognitive processes extend beyond the brain into the environment when external resources are reliably coupled to action. Architecture’s long history of externalising cognition into drawings and models finds its theoretical extension here: computational environments are not aids to thought but components of it.

³ M. Polanyi, *The Tacit Dimension*, Doubleday, New York, 1966, p. 4. Polanyi’s formulation describes the condition of skilled practice in which recognition precedes articulation. Aesthetic cognition in architecture operates within exactly this structure—and is intensified, not replaced, by the expansion of synthetic design environments.

drawings and models—but the scale, resolution, and generative capacity of the extension. Latent computational spaces, surrogate models, optimisation engines, and feedback-driven simulations do not assist thought; they constitute a new cognitive terrain. Within this terrain, design decisions are not linearly authored but iteratively negotiated. Possibilities are generated, perturbed, filtered, and recombined through processes that exceed explicit articulation.

Michael Polanyi's observation that *we know more than we can tell* acquires new architectural significance here. Synthetic design cognition unfolds through exactly this condition: computational systems produce fields of possibility that are not simply chosen from, but navigated. The designer operates within these systems—steering, discovering, filtering, committing to trajectories—exercising a kind of tacit intelligence that is trained and situated even when it cannot be fully verbalised. Cognition becomes distributed across an ecology of generation and judgment.

Crucially, noise plays a constitutive role in this ecology. Stochastic variation, probabilistic inference, and algorithmic drift introduce productive instability into design processes. Noise is not an error to be eliminated but a generative force that enables exploration beyond established patterns and deterministic solutions. It resists premature closure and allows novel spatial and organisational configurations to emerge. Environmental logics, material constraints, and programmatic relations do not converge neatly. They interfere, overlap, and pull in different directions. Synthetic design cognition operates within this field of negotiated indeterminacy—and thrives there.

4. AESTHETIC COGNITION: THE IRREDUCIBLE CAPACITY

Within this expanded cognitive field, traditional notions of authorship become inadequate—not because the architect disappears, but because the architecture of authorship itself must be rethought. The architect as singular creative origin gives way to a distributed assemblage of contributors: computational agents, material behaviours, environmental forces, algorithmic logics. Agency is no longer centralised but dispersed across interacting systems.

Yet, this redistribution of agency does not imply the disappearance of human creativity. What remains irreducible is not authorship as ownership, but judgment as a cultivated capacity for evaluating meaning, coherence, and consequence in the absence of fully defined rules.

This claim requires defence. The proliferation of generative AI tools—diffusion models, image synthesis systems trained on vast architectural datasets—might seem to challenge it. If machines can generate images that are aesthetically convincing, that audiences find beautiful or spatially compelling, does this not demonstrate that aesthetic judgment can be formalised and reproduced computationally?

The answer, grounded in both theoretical argument and practical experience working within these systems, is no—and the reason matters. What diffusion models produce at scale is a sophisticated statistical average of existing aesthetic preferences. They are extraordinarily powerful instruments for finding the centre of a distribution: the most likely form, the most recognisable style, the image that satisfies the largest number of trained expectations simultaneously. This is precisely why AI-generated architectural imagery so rapidly converges on a particular generic quality—technically proficient, spatially plausible, aesthetically familiar, culturally inert. It is fluent in the known. It is structurally incapable of the genuinely unfamiliar. This is AI slop at civilisational scale: a proliferation of competent averages mistaken for creative output.

Aesthetic cognition in architecture is not the ability to satisfy existing preferences. It is the capacity to perceive latent order within configurations that have no prior typological grounding—to recognise coherence before it is legible, to anticipate spatial logic before it is formalised, to commit to a trajectory whose destination is not yet visible. This is not intuition in the romantic sense. It is a trained, historically situated, practically grounded capacity that operates at the edge of what can be verified. It is precisely what the practitioner does in the gap between what the system generates and what the work becomes.

Working within expanded design search—navigating high-dimensional generative fields, filtering vast combinatorial outputs, recognising which configurations carry genuine architectural force—makes this capacity more demanding, not less. The architect operating within synthetic systems is not relieved of aesthetic judgment by computational assistance. She is placed in a position where that judgment must operate faster, across a wider field, and with higher stakes than conventional practice requires. The system expands the space of possibilities; it does not select among them. That selection—that commitment to a trajectory rather than another—remains irreducibly human, and irreducibly difficult.

This is also the distinction between design cognition and mere design generation. Generation is abundant; it costs almost nothing to produce

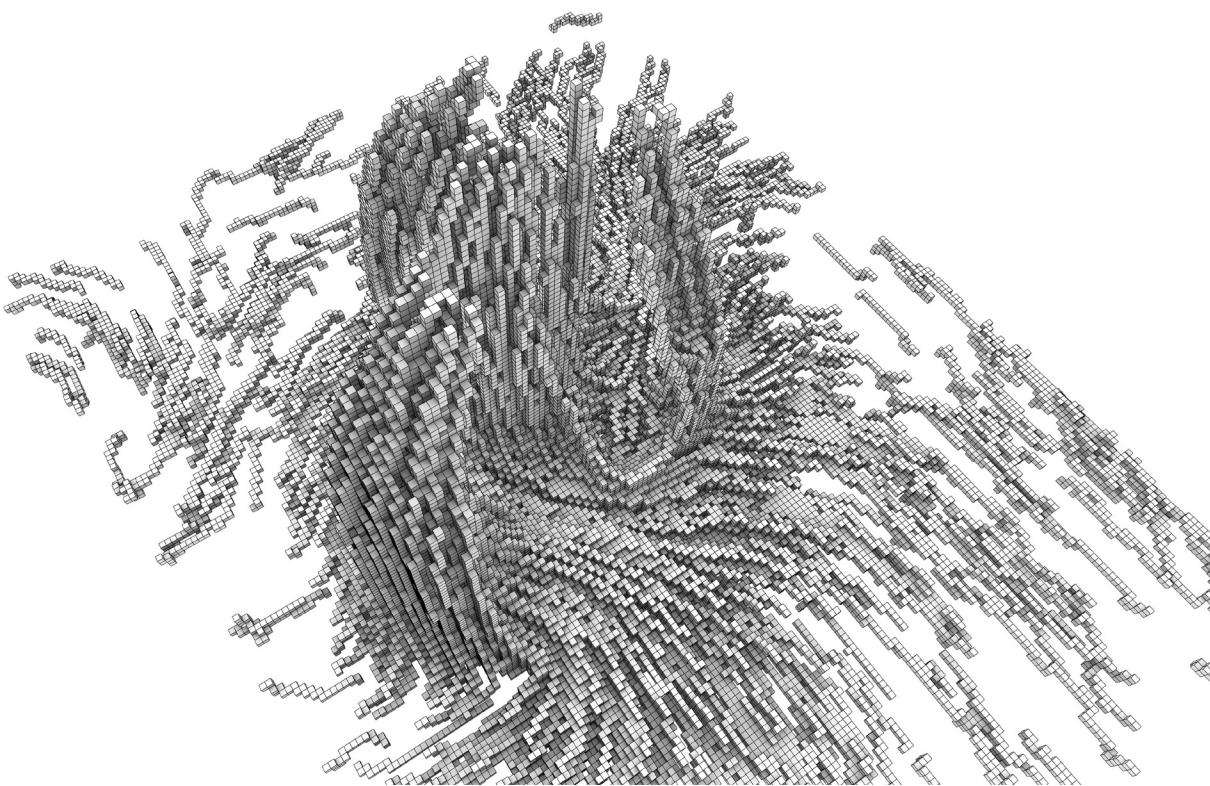


Figure 1. Alisa Andrasek, *Agentic Architecture*, implemented design search system. Courtesy of the author. High-resolution voxel substrate, environmental field simulations, and multi-objective agents operate as a generative field. Environmental exposure, spatial adjacency, material logic, and programmatic gradients are negotiated simultaneously, producing site-specific massing, vertical ecology, and discrete AI Timber assemblies within a coherent system.



Figure 2. Alisa Andrasek, *Vertical field typology: superimposed architectural states*. Courtesy of the author. Discretised timber and glass assemblies generated through AI-driven combinatorics and environmental imprinting. Representation unfolds as a field of coexisting architectural configurations, expressing multiplicity intrinsic to design search.

a thousand variants. Cognition—the ability to know which of those thousand variants is worth pursuing, and why, and what it opens—is scarce. It is where the practice lives.

5. AGENTIC ARCHITECTURE: DESIGN COGNITION IN PRACTICE

These claims are operative in practice. The Agentic Architecture project, developed for the *Intelligens* exhibition at the 2025 Venice Architecture Biennale, acts as a demonstrator of design cognition as a distributed, synthetic process—and as a test of these arguments against the resistance of an actual site, with actual material, environmental, and tectonic constraints.⁴

Curated by Carlo Ratti under the theme *Intelligens: Natural, Artificial, Collective*, the Biennale sought to explore new forms of intelligence emerging from the convergence of planetary systems, technological processes, and collective agency. Agentic Architecture was conceived as a direct response: positioning architecture itself as a form of collective intelligence—not as a claim about institutional scale, but as an argument about how design can organise relations among human, nonhuman, and environmental actors within a shared generative framework.

The project is situated in Miami's volatile coastal ecology: rising sea levels, storm surge, extreme heat and humidity, and intensifying pressures on high-density collective living. Rather than treating these as constraints to be mitigated after the fact, the site is operationalised as part of the project's deep ecology. Environmental forces—wind, water, solar exposure, views—participate directly in architectural formation. They are not analysed and then applied. They enter the generative field from the outset.

The design process begins by voxelising site data, producing a high-resolution substrate capable of hosting multiple interacting datasets. An electromagnetic field simulation establishes macroscale organisational logic through which site determinants interact. The resulting massing exhibits spiral and torsional tendencies arising from first-principles interactions among context-specific vectors: orientation, solar exposure, views, wind patterns, wave dynamics. A multi-agent system is then deployed within this re-voxelised field—introducing stochastics, local negotiation, and high-resolution differentiation. These agents are

⁴ A. Andrasek, "Agentic Architecture: Synthesising Complexity for Regenerative Futures," Exhibition *Intelligens. Natural. Artificial. Collective.*, curated by Carlo Ratti. 19th International Architecture Exhibition—La Biennale di Venezia, Venice, 2025.

structured as multi-objective actors, negotiating overlapping and sometimes competing criteria: environmental exposure, spatial adjacency, material logic, and the connective vector logics of prefabricated elements.

The architect operates within this field by searching for emergent expressions and aesthetics—exercising agency through pattern recognition rather than prescribing outcomes. Within a combinatorially vast and rapidly evolving field, aesthetic cognition must move faster than articulation, committing to trajectories before their consequences are fully known. Judgment and generation are simultaneous rather than sequential. This is the practice of design cognition in its most concentrated form.

Water is present from the outset as a primary design agent. The building's lower condition—its wet feet—develops as a high-resolution field of differentiated pylon foundations, piers, and boardwalks: not a repetitive substructure supporting an independent object, but a distributed interface through which hydrological pressures are absorbed, scattered, and neutralised. The logic draws on mangrove root systems, where resistance is achieved through intricacy, redundancy, and distribution rather than mass. Architecture here does not oppose environmental force—it resolves force through organisation.

Air operates in continuity with this hydrology. External wind forces inform the depth, porosity, and differentiation of the building skin. Internally, Miami's hot and humid climate drives a multiplied chimney effect: a dense network of vertical air channels and connective voids generating internal airflow and passive cooling. The building develops its own internal weather. What is typically treated as invisible background—pressure gradients, thermal difference, airflow—enters directly into architectural formation.

The material system through which this becomes constructable is AI Timber: a research trajectory linking discrete timber systems with computational design and machine-learning-driven combinatorics. Timber joints operate as structural, geometric, and computational agents. Their vector-based connection logic allows large families of differentiated assemblies to emerge from a shared set of discrete elements—enabling prefabricated timber to move beyond repetition and standardisation, producing heterogeneous, context-sensitive architecture that remains fully buildable.

The resulting vertical organisation does not align with established typologies. It operates as a vertical field—a complex built ecology structured through gradients of light, exposure, ventilation, and protection.

Space is organised through environmental and performative intensities rather than through the repetition of floor plates. What emerges cannot be named with existing terminology. It is not a tower. It is not infrastructure. It is a yet-to-be-named typology, legible only through the logic of its own formation.

Representation is integral to this process. Quantum render does not function as a final visualisation layer but as an operative mode of seeing the system—rendering architecture as a field of gradients, porosities, and environmental effects, where material and atmosphere are inseparable. The project exists as a superposed field of architectural states rather than a singular resolved image. Representation sustains openness rather than closure, and participates in design cognition rather than following it.

6. SCALE, INVISIBILITY, AND THE LIMITS OF PERCEPTION

A further consequence of synthetic design cognition concerns its relationship to scale. Architectural practice has historically operated within dimensions legible to human perception. Contemporary design increasingly engages phenomena unfolding across vastly different registers: micro-material assemblies, atmospheric systems, digital infrastructures, planetary cycles.

Many of the forces shaping architectural environments remain imperceptible: data flows, energy exchanges, biochemical processes, algorithmic operations. Synthetic intelligence enables these invisible dynamics to become operational—allowing forces, flows, and fields to participate directly in design logic and to crystallise into legible spatial and aesthetic structures. In doing so, architecture extends beyond the limits of the human sensorium without abandoning the requirement for spatial and material coherence.

Time becomes elastic within this expanded field, accommodating geological durations alongside real-time feedback. Space becomes relational, defined by gradients and fields rather than discrete objects. Scale ceases to be a fixed category and becomes an adjustable parameter within design cognition. These are not rhetorical expansions but actual operations: the voxelised site model, the electromagnetic field simulation, the multi-agent system, the fluid-dynamic analysis—each engages a different temporal and spatial register, contributing to an architectural outcome unreachable by any single register alone.

7. PLANETARY DESIGN COGNITION: SYSTEMIC INTERVENTION AT SCALE

The convergence of these transformations situates architecture within a broader planetary context. But the term planetary requires precision. It does not mean that every project addresses global systems, or that architecture should aspire to the scale of infrastructure policy. It means something more specific and more tractable: that architectural thinking can identify systemic problems embedded in the built environment at civilisational scale, intervene at the level of their underlying logic, and propagate solutions that are replicable, adaptable, and genuinely consequential.

One of the most persistent and underacknowledged of these systemic problems is the monotony of prefabricated architecture. For decades, the industrialisation of building production has delivered efficiency at the cost of differentiation. Prefab systems close down the combinatorial possibilities of construction in the name of standardisation, producing environments that are technically adequate and spatially impoverished—unloved buildings at a planetary scale. This is not a cosmetic failure. It is a structural one: the logic of the system produces uniformity as a necessary output, and the resulting environments fail precisely because they are not loved.

AI Timber proposes an intervention at exactly this level. The connector—the joint—is where the combinatorial logic of prefabricated construction is either closed down or opened up. Conventional prefab fixes the joint geometry, which fixes the assembly logic, which fixes the spatial possibilities of the resulting buildings. AI Timber applies vector mathematics to joint geometry, transforming each connection from a fixed constraint into a combinatorial agent. The result is a system that retains the economic and environmental advantages of prefabrication—standardised elements, rapid assembly, material efficiency—while recovering the spatial richness and differentiation that prefab has historically foreclosed.

This is what it means to work at a planetary scale without grandiosity: to identify the exact point in a system where a precise intervention unlocks consequences that propagate outward. The connector is humble. Its implications are not. If timber prefabrication is to be deployed at the scale the climate crisis demands—replacing concrete and steel globally, absorbing carbon, enabling rapid and affordable housing in regions that need it most—then the quality of the environments it produces is not an aesthetic preference. It is a condition of whether people will accept,

inhabit, and maintain what is built. Democratising good design is not a luxury aspiration layered on top of real problems. It is constitutive of any solution to them.

Agentic Architecture is a pilot for this logic, not a singular building. The design system it embodies—environmental forces as generative inputs, discrete material elements as combinatorial agents, multi-objective search across high-dimensional fields, aesthetic cognition as the irreducible filter—is transferable. It is a methodology before it is a building. The Miami project demonstrates what the system can do under specific coastal conditions; the same framework is applicable to high-density urban infill, to post-disaster reconstruction in climate-vulnerable regions, to the urgent global problem of affordable housing that is also spatially dignified. The planetary claim is grounded in exactly this replicability.

But replicability alone is not the argument. What the methodology also produces—and what matters as much as its transferability—is a genuinely new aesthetic: spatial experiences of wonder and awe that could not have been designed by any other means, because they could not have been imagined by any other means. They emerge from the logic of the system itself, from the interaction of environmental forces, material combinatorics, and the architect's sustained act of recognition within a generative field. This is not aesthetics as surface or style. It is form that carries the intelligence of its own making, spatial sequences whose richness is inseparable from the epistemic process that produced them.

This distinction matters because aesthetic cognition—the trained, irreducible capacity to know which configuration carries force and which is merely generated—is not an automatic consequence of access to powerful tools. The methodology described here is accountable to both dimensions simultaneously: it must operate at a planetary scale, and it must produce environments worth inhabiting, worth experiencing, worthy of being called architecture. These demands are not in tension. Each is the condition of the other.

8. FROM TYPOLOGY TO FIELD

As architectural cognition shifts, established typologies begin to dissolve—not because classification is impossible, but because the logics that generated historical types are no longer the primary drivers of form. Buildings emerging from the entanglement of environmental dynamics, generative computation, human judgment, and multispecies inhabitation resist

classification as tower, block, or infrastructure. Their logic is not static but relational, evolving over time in response to changing conditions.

Architecture increasingly takes the form of emergent assemblages: spatial organisations arising from the interaction of discrete agents, stochastic processes, and ecological relations, rather than from fixed programmes. Gilbert Simondon's account of individuation is instructive here: form, in his framework, does not pre-exist the process that produces it but emerges through the resolution of tensions within a field of pre-individual potentials.⁵ What Agentic Architecture demonstrates operates by exactly this logic. When design is organised as a search process within a generative field, outcomes genuinely exceed available typological categories not by chance but by necessity—because the process is oriented toward configurations that existing categories have not yet stabilised. The indeterminacy is procedural rather than rhetorical.

The move from typology to field does not signal the end of architectural form, but its redefinition. Form becomes provisional, contingent, and responsive—an emergent property of cognitive processes distributed across systems. Architectural language must evolve accordingly, developing concepts capable of articulating these dynamic conditions without either celebrating indeterminacy as an aesthetic value in itself or retreating to the false clarity of resolved objects.

9. CONCLUSION: WHAT REMAINS OPEN

Architecture, as it emerges through synthetic intelligence, is no longer defined primarily by objects or images. It is defined by its capacity to think through systems, through materials, through time. Design cognition becomes a distributed, synthetic process in which human and inhuman agencies co-evolve—not toward a predetermined resolution, but through sustained engagement with complexity as a medium.

The arguments assembled here—the operative turn, the extended cognitive field, the irreducibility of aesthetic judgment, the systemic logic of planetary intervention—converge on a single claim: that architecture is currently undergoing an epistemic transformation more fundamental

⁵ G. Simondon, *Individuation in Light of Notions of Form and Information*, University of Minnesota Press Minneapolis, 2020. Simondon's account of individuation as a process of resolving pre-individual tensions within a metastable field—rather than the instantiation of a pre-given form—offers a rigorous philosophical basis for understanding how architectural form can emerge through generative processes without being prescribed by them.

than any shift in style, technology, or professional organisation. What is changing is not the appearance of architecture but its nature as a cognitive practice. This transformation is already underway in work that does not yet have an adequate theoretical account of itself. This paper is an attempt at that account.

But it would be a mistake to end on resolution. The framework proposed here opens at least as many problems as it closes, and those problems are more interesting than the conclusions.

If aesthetic cognition is irreducible—if the capacity to recognise latent order within a field of possibilities cannot be formalised without loss—then it remains genuinely unclear how this capacity is transmitted, cultivated, or shared. Design education has never had a satisfactory account of how aesthetic judgment develops; for most of its history, it could pretend the question was answered by the master-apprentice relation or by the authority of established taste. Neither answer is available now. Synthetic intelligence floods every practitioner with more generated possibilities than any previous generation of architects encountered in a career. The question of how to train the eye—the real eye, the one that knows what it is looking for before it can say—becomes urgent in a new way.

If the planetary claim is genuine—if architecture can identify systemic leverage points and propagate solutions at civilisational scale—then the profession faces a question it has largely deferred: what obligations follow from that capacity? The ability to democratise good design is not the same as the will to do so. Synthetic cognition, deployed within existing economic structures, is perfectly capable of concentrating spatial quality further rather than distributing it. The tools do not determine the direction.

And if design cognition is genuinely distributed—if thinking happens across the coupled system of human judgment and synthetic process rather than within either alone—then the boundary of the work is no longer clear. The building is a crystallisation, not a completion. The design system that generated it continues to exist, continues to be navigable, continues to produce states that were not reached. The work is always in-formation.

Perhaps this is the most honest thing that can be said about architecture in the age of synthetic intelligence: it is a practice that has become capable of generating more futures than it can pursue, and the question of which futures are worth building—and for whom—has never been more open, or more consequential.

REFERENCES

- Andrasek, Alisa (2025), “Agentic Architecture: Synthesising Complexity for Regenerative Futures,” Exhibition *Intelligens. Natural. Artificial. Collective.*, curated by Carlo Ratti, 19th Venice: International Architecture Exhibition—La Biennale di Venezia.
- Clark, Andy, David Chalmers (1998), “The Extended Mind,” *Analysis*, 58, 1, 1998, pp. 7–19.
- Parisi, Luciana (2013), *Contagious Architecture: Computation, Aesthetics, and Space*, Cambridge, Mass.: MIT Press.
- Polanyi, Michael (1966), *The Tacit Dimension*, New York: Doubleday.
- Simondon, Gilbert (2020), *Individuation in Light of Notions of Form and Information*, trans. Taylor Adkins, Minneapolis: University of Minnesota Press.

Marco Vanucci*

ART IN THE AGE OF GENERATIVE MACHINES

ABSTRACT: This paper proposes that generative machines force a reconsideration of creativity not as originality *ex nihilo*, but as a structured practice of imitation, variation, and recursion. Rather than inaugurating an unprecedented form of artistic autonomy, contemporary AI systems revive pre-modern theories of art as *mimesis*, in which invention emerges through the transformation of inherited forms. Tracing a lineage from cybernetic feedback and conceptual art to machine-learning aesthetics, the paper argues that automation displaces authorship from the production of images to the construction of processes, aligning artistic agency with the epistemological act of framing, limiting, selecting, and choreographing generative systems. Situating contemporary AI practices within longer histories of imitation and programmed aesthetics, the paper contends that intelligent automata operate less as autonomous authors than as experimental apparatuses, epistemic devices, or serendipitous machines that expose the structural, mimetic, and recursive foundations of human creativity itself.

KEYWORDS: art, agency, artificial intelligence, cybernetics, mimesis

* Marco Vanucci: Anglia Ruskin University/London South Bank University; marco@open-systems-a.com.

This is an Open Access article under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not transformed in any way and is properly cited.

CYBERNETICS, SERENDIPITY, AND PLAY

On 2 August 1968, at the Institute of Contemporary Arts in London, the exhibition *Cybernetic Serendipity: The Computer and the Arts* opened, marking a turning point in the visual arts in the age of information technology. As stated in the press release, the exhibition aimed to offer an overview of “all aspects of computer-assisted creative activity: art, music, poetry, dance, sculpture, animation.”² Robots, poetry machines, musical and painting devices programmed through randomized algorithms were presented alongside early computational graphics and animations. Chance was employed as a way to inject surprise and awe into otherwise deterministic computational protocols.

The term *serendipity* derives from *The Three Princes of Serendip* by Horace Walpole (1754) and describes the ability to make happy and accidental discoveries. The curators’ ambition was to show the influence of cybernetics—already discussed and applied in scientific fields—within contemporary artistic practice. The encounter between art and science did not merely introduce new tools but new epistemologies and structures of thought. Form, play, authorship, creativity, movement, automation, and participation became central concerns of a new artistic paradigm.

Art in the age of information technology thus addressed, simultaneously, very old and very new questions. The term *cybernetics*, from the Greek *κυβερνήτης* (helmsman), was first used in 1948 by Norbert Wiener as the title of his book *Cybernetics: Communication and Control in Animals and Machines*. According to Heinz von Foerster, one of the founders of the discipline and a collaborator of Wiener, cybernetics is the ability to think in a circular way: in cybernetic systems, “A leads to B, B to C, and C can return to A.”³ A device responds to external stimuli and, through feedback, influences the environment that generated those stimuli in the first place.

It is precisely on this principle of feedback loops that early computational artworks sought to develop new forms of participation. The devices on display at the ICA reacted to their environment and interacted with visitors by producing sound, light, or movement. Alongside Nam

¹ J. Reichardt (ed.), *Cybernetic Serendipity: The Computer and the Arts*, Studio International, London, 1968.

² *Ibid.*, p. 5.

³ H. von Foerster, “Ethics and Second-Order Cybernetics,” *Understanding Understanding: Essays on Cybernetics and Cognition*, Springer, New York, 2003, p. 289.

June Paik's *Robot K-456* and Jean Tinguely's *Meta-Matics*—painting machines that ironized authorship and mechanized gesture—the exhibition displayed algorithms and machines capable of producing music, poetry, graphics, and animation. Michael Noll experimented with computers and printers to generate images resembling Mondrian's *Compositions with Lines*, inviting visitors to distinguish the copy from the original.

Play and participation thus became central strategies. Artists sought opportunities to develop new forms of engagement and discovery through computational machines. In architecture, Cedric Price and Gordon Pask, a polymath cybernetician based in London, developed a project as innovative as it was utopian: the *Fun Palace*. Through cybernetic mechanisms establishing human–machine feedback, users could reconfigure the building according to their needs. Pask described architecture in the computer age as a predictive discipline, capable of incorporating an increasing number of parameters and adapting to change.⁴

If the enhancement of human capabilities through machines often generated scepticism or fear, machines nevertheless functioned as tools of knowledge. Today, for instance, the ubiquity of algorithms in daily life is frequently perceived as an invasive form of control and surveillance. At the same time, algorithms allow us to understand patterns governing natural phenomena and human behaviour. While cloud computing depletes resources and contributes to the climate crisis, it also enables the observation of global warming and the deployment of mitigation strategies. Like a perverse cybernetic system, input produces output, which in turn becomes new input in a recursive loop.

The visual arts have long perceived this tension and transformed it into devices, artifacts, and interactive machines. Unlike architecture—where algorithmic automation is often associated with efficiency, control, and optimization—artists have sensed the expressive potential of computation and begun to experiment with its languages. Mathematics became a shared language, though understood in a structuralist sense. Linguistic research initiated by Ferdinand de Saussure and developed by Noam Chomsky in the 1960s fundamentally influenced computational programming.

Sol LeWitt's "mathematical" works, articulated through instructions written in natural language, developed processes, variations, and systems

⁴ G. Pask, "The Architectural Relevance of Cybernetics," *Architectural Design*, 39, 9, 1969, pp. 494–496.

that anticipated generative and computational art. For LeWitt, it was the idea—not the artist’s hand—that defined the artwork. In *Paragraphs on Conceptual Art* (1967), he famously stated that “the idea becomes a machine that makes the art.”⁵

In Italy, the artistic avant-garde and an emerging technological industry created a fertile cultural climate. In 1962, Adriano Olivetti sponsored the exhibition *Arte programmata: arte cinetica, opere moltiplicate, opera aperta* in Milan. As Umberto Eco observed, art and beauty were no longer “something immobile, waiting to be looked at, but something in progress.”⁶ Programmed mechanisms celebrated movement and produced unpredictable patterns—open works in the fullest sense.

Movement and temporality were joined by geometry. The Hungarian artist Vera Molnar, a pioneer of generative art, experimented with geometric abstraction in which algorithms and randomness defined “the rules of the game.” By parametrizing simple figures—often letters of the alphabet—Molnar generated permutations that were always different yet always referable to the same underlying matrix. Language, understood as a system of signs, became a container of meaning subject to continuous manipulation.

Through procedural randomness, Molnar sought to remove human choice from artistic practice. Paradoxically, by focusing on freedom through its absence, her work foregrounded a condition central to human creativity. Deterministic rigor was counterbalanced by moments of formal surprise. Similarly, in the 1970s, the British artist Harold Cohen developed *AARON*, a software designed to simulate drawing and generate unpredictable results. Through feedback and archival memory, the system learned from past actions, recalling them to inform future ones according to predefined instructions.

TOWARDS A DIGITAL MANNERISM

The machine thus becomes a database capable of capturing gestures and reproducing styles—what Renaissance theory would have called *maniera*. Generative AI systems merely automate stylistic imitation, reminding us of a fundamental truth of creativity: nothing is created from nothing; everything is generated from what already exists. There is no generative process without a dataset, and therefore no creation without imitation.

⁵ S. LeWitt, “Paragraphs on Conceptual Art”, *Artforum*, 5, 10, 1967, p. 79.

⁶ U. Eco, *Arte programmata*, exh. cat., ed. B. Munari, G. Soavi, Olivetti, Milano, 1962, s. p.

In fifteenth-century literature, Renaissance scholars used *maniera* as a synonym for style, emphasizing that artistic creation depends on precedents. Vasari refined this idea by framing Renaissance art as the emulation of the ancients, whose aesthetic, technical, and moral principles were considered exemplary. For classical and artisanal traditions alike, creativity developed through imitation: selecting, assimilating, transforming, and occasionally contesting inherited forms.

Art as *mimesis* entered a crisis with modernity. For ideological and practical reasons, imitation was banished from art and architecture schools in favour of originality as an absolute value. Today, this supremacy of the “new for the sake of newness” is so deeply ingrained that, as Mario Carpo observes, “no one today knows what imitation is, what it does, or how it works.”⁷ Ironically, it is precisely generative machines that are reactivating imitation after a century of repression.

Jean Baudrillard’s concept of the *simulacrum* further complicates this return. In postmodernity, images no longer refer to an external reality but circulate within a closed system of signs—*hyperreality*. Generative AI aligns with this logic. Its outputs are not copies of originals but syntheses of prior images, producing what Baudrillard would call “copies without originals.”⁸

Refik Anadol’s recent work exemplifies this condition. By training a large language model on the metadata of MoMA’s collection, Anadol generated constantly transforming visual fields oscillating between recognition and estrangement. Rather than classifying or representing the world, his AI explores dreams, hallucinations, and speculative memory—an archaeology of possible pasts.

FROM TACIT KNOWLEDGE TO THE BLACK BOX

In *Introduction to Mathematics*, Alfred North Whitehead writes: “Civilization advances by extending the number of important operations which we can perform without thinking about them.”⁹ This provocation captures a paradox at the heart of technical culture: progress coincides not with heightened awareness but with the delegation of cognition.

⁷ M. Carpo, “Generative AI, Imitation Learning, and the Automation of Tacit Knowledge,” *A+U*, 646, 2024, p. 104–109.

⁸ J. Baudrillard, *Simulacra and Simulation*, trans. University of Michigan Press, Ann Arbor, 1994.

⁹ A. N. Whitehead, *An Introduction to Mathematics*, Williams and Norgate, London, 1911, p. 61.

Architecture exemplifies this logic. Architectural design is a problem-solving methodology grounded in what Michael Polanyi described as tacit knowledge.¹⁰ For centuries, this knowledge was transmitted through apprenticeship and imitation. Yona Friedman described architecture as a vocation learned through practice rather than instruction.¹¹

References—projects, typologies, strategies—have always guided architectural thinking, yet their relevance remains implicit and difficult to formalize. Architectural problems cannot be fully specified in advance, nor can their solutions be conclusively validated. This reliance on tacit judgment long shielded architecture from automation.

Digitization altered this balance. Quantification migrated from structure to geometry, topology, orientations, and networks.¹² Therefore, the adoption of an increasingly quantitative approach—modularity, proportional relationships, and environmental considerations—has led to a greater understanding of architectural design and, at the same time, a departure from that empiricism¹³ which, in the absence of objective criteria, characterized architecture before computation. In fact, what computational tools did was to translate aspects of tacit knowledge into software. Expertise became transferable, but epistemic transparency was lost.

This shift culminates in the black box, “word (that) is used by cyberneticians whenever a machine or a set of commands is too complex. Instead, they draw a little box that they don’t need to know about other than its inputs and outputs.”¹⁴ As Wiener and later Bruno Latour observed, black boxes function precisely because their inner workings are no longer interrogated. In AI-driven design, references are no longer cited; they are absorbed. Interpretation gives way to interpolation.

The distinction between automation and autonomy becomes critical. Kantian autonomy implies self-legislation; Luhmann’s autopoiesis describes self-reproducing but operationally closed systems. AI systems are neither autonomous agents nor simple machines. They are recursive,

¹⁰ M. Polanyi, *The Tacit Dimension*, Doubleday & Company, Garden City, New York, 1966, p. 4.

¹¹ Y. Friedman, M. Orazi, *Yona Friedman. The Dilution of Architecture*, Park Books, Zürich, 2015, p. 29.

¹² C. Alexander *et al.*, *A Pattern Language: Towns, Buildings, Construction*, Oxford University Press, New York, 1977.

¹³ B. Zevi, “Cervelli elettronici? No macchine calcolatrici,” *L’architettura: cronaca e storia*, 6, 62, 1960, pp. 508–509.

¹⁴ B. Latour, *Science in Action: How to Follow Scientists and Engineers Through Society*, Harvard University Press, Cambridge Mass., 1987, p. 2.

adaptive, and opaque—quasi-autopoietic systems whose apparent autonomy is simulated rather than enacted. AI does not replace human autonomy by becoming autonomous. It displaces autonomy by making deliberation unnecessary.

TOWARD A SPECULATIVE AUTONOMY

If automation absorbs optimization and black boxes encapsulate tacit knowledge, human agency relocates: from execution to interruption, from control to play. Surrealist automatism sought to bypass rational intention; contemporary generative systems invite engagement with hallucination, error, and deviation. Machines become performative partners—unruly, opaque, generative. Creativity emerges through constraints, in the ludic space where rules are bent, subverted, or reimaged.

Generative and algorithmic art frees artists from idiosyncratic expression while introducing new choices. Recursive algorithms present halting problems: when to stop the process, when conditions are satisfied. Authorship remains an act of conscious and rational decision—extreme self-determination. Art and computation chase each other, intersecting to produce new rules, exceptions, and transgressions. From Turing’s Imitation Game to Baudrillard’s simulacra, AI simulates, deceives, and allows us to act *as if*. In this interplay, meaning is evoked, not merely created. Cohen reminds us: “The real power, the real magic [...] lies not in the creation of images, but in the evocation of meaning.”¹⁵

In the age of *αὐτόματα*—self-moving systems that simulate autonomy without possessing it—the challenge is to redefine human agency: not as authorship *ex nihilo*, but as the capacity to frame, select, and speculate within systems that perform, subvert, and dream. When machines no longer obey but dream, the task is to learn how to dream with them.

¹⁵ H. Cohen, “What is an Image,” <https://aaronshome.com/aaron/publications/whatis-animage.pdf#:~:text=The%20conclusions%20offered%20bear%20upon,intentionality.%20They%20propose%20also%20that>, (accessed 20 December 2025).

REFERENCES

- Alexander, Christopher *et al.* (1977), *A Pattern Language: Towns, Buildings, Construction*, New York: Oxford University Press.
- Baudrillard, Jean (1994), *Simulacra and Simulation*, trans. Sheila Faria Glaser, Ann Arbor: University of Michigan Press.
- Carmo, Mario (2024), "Generative AI, Imitation Learning, and the Automation of Tacit Knowledge," *A+U*, 646, pp. 104–109.
- Cohen, Harold (s. a.) "What is an Image", <https://aaronshome.com/aaron/publications/whatisanimage.pdf#:~:text=The%20conclusions%20offered%20bear%20upon,intentionality.%20They%20propose%20also%20that> (accessed 20 December 2025).
- Eco, Umberto (1962), *Arte programmata*, exh. cat., ed. Bruno Munari, Giorgio Soavi, Milano: Olivetti, s. p.
- Friedman, Yona, Matteo Orazi (2015), *The Dilution of Architecture*, ed. Nader Sejar, Zürich: Park Books.
- Latour, Bruno (1987), *Science in Action: How to Follow Scientists and Engineers Through Society*, Cambridge, Mass.: Harvard University Press.
- LeWitt, Sol (1967), "Paragraphs on Conceptual Art," *Artforum*, 5, 10, pp. 79–83.
- Pask, Gordon (1969), "The Architectural Relevance of Cybernetics," *Architectural Design*, 39, 9, pp. 494–496.
- Polanyi, Michael (1966), *The Tacit Dimension*, Garden City, NY: Doubleday & Company.
- Reichardt, Jasia (1968), "Introduction," in Jasia Reichardt (ed.), *Cybernetic Serendipity: The Computer and the Arts*, London: Institute of Contemporary Arts, pp. 5–10.
- von Foerster, Heinz (2003), "Ethics and Second-Order Cybernetics," in: *Understanding Understanding: Essays on Cybernetics and Cognition*, New York: Springer, pp. 287–304.
- Whitehead, Alfred North (1911), *An Introduction to Mathematics*, Oxford: Oxford University Press.
- Wiener, Norbert (1948), *Cybernetics: Or Control and Communication in the Animal and the Machine*, Cambridge, Mass.: MIT Press.
- Zevi, Bruno (1960), "Cervelli elettronici? No macchine calcolatrici," *L'architettura: cronaca e storia*, 6, 62, pp. 508–509.

MIMESIS, PRECEDENT, AND ARTIFICIAL INTELLIGENCE: A CONVERSATION WITH MARIO CARPO

ANDREJA PROKOPIJEVIĆ: Early digital theorists often suggested that architects would gradually move from designing objects to designing processes, rules, and systems. Today, design increasingly unfolds through computational systems capable of generating spatial proposals from large bodies of precedents and data. Where would you situate these developments within the longer shift you have described, from the authorial model of design towards more distributed or statistical forms of authorship?

MARIO CARPO: The short answer is that we should look at the history of computation in its entire duration, from when it started to this day, to see how the different parameters we introduced were ebbing and flowing over time. So, when it all started in 1956, when the idea of artificial intelligence was invented, the first methodologies of AI—which came by Marvin Minsky in 1960–61—defined artificial intelligence as a general problem-solving machine. It’s a thinking machine. They called such machines electronic brains.

This first age of artificial intelligence boomed in the 1960s and was still influential in the 1970s when architects started to experiment with it—think of Nicholas Negroponte and a bit earlier Cedric Price—this was the idea of artificial intelligence as a thinking machine, which for architects is a machine that does design, makes design choices, which was Negroponte’s idea in *The Architectural Machine*. This machine replaces the architect. It makes design choices. The idea was: with a bigger computer, more computing power, we will make it work. But as we know, in the 1970s, it didn’t work. This dream of the 1960s and 1970s was abandoned. In the history of computer science, this is called the winter of artificial intelligence, when many came to the conclusion that not only it did not work, but that the conceptual framework was wrong from the start. They concluded it would never work. So, this story starts in the

1950s, booms in the 1960s, declines in the 1970s, disappears in the 1980s. Now, when we come to what I call the digital turn in architecture, a story which starts in the early 1990s with Peter Eisenman, Greg Lynn, Bernard Cache, etc., the computer we used at that time was not a thinking machine, but a drawing machine. It did not make design choices. It was a tool that makes drawings—spline modelling, etc.—what later became parametricism. But it was not a thinking machine.

So, phase one: computers are thinking machines—extinguished in the 1970s–80s. In the 1990s: computers are drawing machines. And this was computation in architecture until 10 years ago, when artificial intelligence came back. But the second coming of artificial intelligence—machine learning—is what made it work. And unlike in the 1960s, this time it works. Today we are back to where we started 60 years ago: dealing with intelligent machines that can make design choices. But this is a novelty, because everything we knew as the digital turn was not about artificial intelligence, it was about the automation of drawing. Now we are back to artificial intelligence as such: a thinking machine, an intelligent machine, a machine to which we can delegate design choices. It is brand new. We don't know how to deal with it. We don't know how it will affect design agency. Because before, we were still in charge. Now it's a new story. And I don't know where it came from and I have no idea where it's going. We'll see.

AP: As generative AI produces new forms through the recombination and transformation of existing precedents, parallels have often been drawn with the logic of Mannerism. You describe Mannerism not simply as a continuation of tradition but as a critical engagement with it. How might this notion of “licensed deviation,” as you once noted, help us understand contemporary algorithmic design practices?

MC: The way I define it, generative AI is an imitation machine. The term is controversial because we do not all mean the same thing by imitation. But at any rate, an imitation machine is a machine that generates new images which are similar to the images on which it was trained. Generative AI is a machine that makes new stuff out of old stuff. Think of it as a grinder, like a food processor. What you put in determines what you take out. If you train the model with a dataset made exclusively of white dogs with a black nose, what you get out is going to be white dogs with a black nose. If you train the machine with a dataset made of yellow cats with a pink nose, you will get yellow cats with a pink nose. It is similarity.

It is about style transfer, or what I call the imitation machine—the automation of visual imitation.

We shall come back to that, because imitation is not a neutral term. It is in fact a very problematic one. But if you want to go into the mannerism discussion, there is an article I wrote in *Log* a couple of years ago where I discussed this topic with regard to Peter Eisenman and mannerism. The idea that mannerism critiques precedent is similar, to some extent, to what generative AI is doing, because generative AI transforms precedent. But there is a difference, to which Peter Eisenman is very much alert. When a human being—someone like him, or any mannerist in time—refers to precedent, you recycle precedent, you critique it, you take a position. You take a stance *vis-à-vis* precedent: you may like it, you may not like it, or you may only like it to a certain extent.

Mannerism is, let's put it that way, calculated rule-breaking, or regulated rule-breaking, as in the Saussurean dialectic of *langue* and *parole* in structural linguistics. There is a code, you learn the rules of the code, and then you break some, because that is how languages work. Now, generative AI appears to be doing something like that. But one thing that generative AI cannot do is take a critical position *vis-à-vis* precedent, because it does not have a critical position. Generative AI cannot do calculated rule-breaking because it is not a rule-based system. It does not know rules, does not use rules, does not make rules, and does not break rules. So, there is one thing that generative AI cannot be: it cannot be mannerist. Because in order to be mannerist, you must know what the rules are and which rules you want to break. Generative AI imitates, but imitates in a way that a child does before going to school, without knowing the rules. We only learn the rules when we go to school. We learn grammar, we learn to write. As a kid, before going to school, one can speak by imitating voices, pure imitation, without knowing any rules. The rules we learn later. Then, once we know the rules, we may break them—because we know them, or because we choose to ignore them. But that happens after learning.

Generative AI is in the state of a child before going to school. It imitates, but it does not know the rules. So, it can imitate mannerist works. It can be mannerist in style, but it cannot be mannerist in spirit. Because a mannerist in spirit is someone who knows the rule he is breaking. And generative AI cannot do that—not today, and probably not in the near future, although who knows? Things change fast. But for the time being, generative AI cannot break rules because it cannot make rules. It does not work with rules. There are no rules in generative AI.

We had rules in earlier computational systems—shape grammars, for example—but this is something we have abandoned. Generative AI does not work with rules. So, there cannot be any form of mannerism as a critical stance, as a critique of precedent. Generative AI can imitate precedent in the way a child would before going to school. So, it is very good at learning, but it does not know grammar. That is both the advantage and the disadvantage of generative AI. It is at the mental age of a child of about five years, before entering school. That is where it stands. It may improve over time, but for the time being, it does not.

AP: If we now have forms that are generated through a logic of precedent and variation, it affects the way we perceive them as images. Mechanical printing once created a new kind of visual trust by guaranteeing the identical reproducibility of images. In contrast, today's digital images, especially those generated by AI, can be perfectly realistic yet entirely synthetic. Do you think contemporary visual culture is entering a new phase in which the authority of images is once again becoming unstable?

MC: Yes, that is happening, but this is not new. This is more of the same. It all started with the rise of digital images in the 1980s or early 1990s. The idea of pixelated images—images made of pixels—has all been explained in a fantastic book by William Mitchell, *The Reconfigured Eye*, published in the early 1990s. At the time nobody had a digital camera, but he already knew how it worked. All the explanation is there. In a nutshell, the images we knew—photographic images—are an indexical trace of reality because they are materially created by the impact of electrons inside the camera. So, you have a traditional camera, not a digital one. It is a box with a lens and a photosensitive film. You take a snapshot, and what actually happens is a mechanical operation: light goes through the lens, it is inverted because it is a camera obscura, and particles hit the photographic film. When they hit, they leave an imprint. That is what we call an indexical trace. If something is recorded on that film, it is because something physically happened. That is why such images could function as evidence. They could be used in court. They recorded something that actually took place—unless someone was cheating. Of course, manipulation always existed. For example, in the Soviet Union, when someone in the Politburo fell out of favour, they would reprint photographs and remove that person. But that required manual, chemical manipulation, and often you could detect it. So, the indexical image had the value of proof.

If the film inside the camera records the imprint of a picture where you see that there was someone behind you when that picture was taken, it means that person was there. The image records an event that happened.

Then digital cameras came. Inside the camera there is no film, but a raster of photosensitive sensors. They convert incoming light into numerical data. So, the image becomes a number. Once it is a mathematical notation, the indexical link between image and reality is broken. The event is translated into a sequence of numbers. And once it is a number, anything can be modified. For example, digital cameras automatically remove red-eye. But that is not a recording of what happened; it is an algorithmic manipulation. William Mitchell already pointed this out in the early 1990s. He said that previously, if a newspaper received a photograph, they assumed that the event had occurred. But with digital imaging, that assumption could no longer be taken for granted. The indexical relation was already broken at that time—30 or 40 years ago.

Today it is entirely broken. What we are experiencing now is not a completely new phenomenon, but an intensification of a process that started decades ago. The idea that images are reliable traces of reality has already been fundamentally undermined.

ŽELJKO RADINKOVIĆ: What do you mean by parametric spaces of possibility? If one assumes that parameters can generate infinitely many variants, does this not risk losing specifically human or existential possibility in favour of a kind of indifferent arbitrariness? Can we therefore still meaningfully speak of so-called mass individualisation?

MC: Mass customisation is the term that I use. Mass individualisation means the same. Yes, everything that is digital is, by definition, mass customisable. It does not mean that it is customised, but in principle it can be. Everything that is parametric is, by definition, mass customisable. Like the parametric notation we studied in high school in calculus: you write the equation of a parabola with x and y , and then parameters a , b , and c . That is not the notation of a single parabola; it is the notation of a family of parabolas. All different, all similar. You change the values of the parameters and you get different parabolas—larger, smaller, wider—but they are all parabolas because the x - y notation is in the same format.

The principle of parametricism, which is the core principle of all digital notation, is that you never notate an object; you notate a family of objects. This is what Bernard Cache, inspired by Gilles Deleuze, called

the objectile—not the object, but a generic object. You notate a family of objects, not an individual one. Then, once you have the family, you assign values to the parameters and you individualise it. You obtain a specific instance. In Aristotelian terms, you notate a genus, within which you have species. Or in mathematical terms, a set. So, the marketing term is customisation; the philosophical term is individuation. That is how digital fabrication works.

This idea is not entirely new. In scholastic philosophy, there was no science of the individual, only a science of species or genera. It is an old idea, but parametricism and digitality have translated it into the way we fabricate objects, not just how we conceptualise them.

ŽR: Would you describe your understanding of architecture as a step toward post-human design logic?

MC: Post-human? I would say pre-humanist. If I were an Aristotelian philosopher in the 14th century, I would describe the world in these terms. I would not have the machine to implement it, of course. So, I would not say post-human. I would say pre-mechanical. The idea of the human we have in mind is the human of the industrial age. We are getting out of that. We are moving away from the human subject of the industrial, mechanical environment—the world of identical reproduction, Benjamin’s aura, etc. The idea of endless variability in digital tools is closer to the variability of artisanal craft before industrial standardization. In the artisanal world, everything is different. In the digital world, everything can be different again—unless we deliberately choose to make identical copies. Industrial production made us accustomed to sameness. But that was an exception, in historical terms.

ŽR: How would you situate your conception of architecture in relation to Walter Benjamin? In Benjamin’s analysis, when infinitely many copies exist, the original loses its privileged status, its aura. In parametric design, there is often no original in the traditional sense, yet a series of individual forms emerges. Do you see these individuations as carriers of a new kind of auratic determination?

MC: Walter Benjamin was absolutely right in conceptualising the technical logic of a machine-made environment. If we live in a world of machine-made objects—think of printing—you have one matrix, one

mould, and you produce identical copies. The original has an aura; the copies do not. This logic extends into industrial production. The original design belongs to the author, and each identical reproduction generates royalties. That is the logic of industrial modernity. But this logic does not apply to craft, and it does not apply to digital making. In craft, every object is different. An artisan makes each object individually, adapting to materials, circumstances, or even personal preference. Digital making works in a similar way. The digital maker is today's artisan. The concept of aura does not apply in the same way. We only need identical reproduction in certain cases—for example, a ballpoint pen. The one I'm holding right now is good for all. Anyone can use it. But in other cases, such as, say, dental crowns, variation is necessary because each tooth is different. Today we have the technology to produce variation efficiently. This is the novelty. It is not an artisan making each object individually—it is a machine mass-producing variations.

AP: When printed images became the dominant medium for transmitting architectural knowledge, handmade drawings gradually lost their status as factual documents and became more autonomous artistic expressions. Today, images can be generated, modified, and circulated almost instantaneously. Does the distinction between drawing, model, and image still carry epistemological meaning for architecture?

MC: We just mentioned Walter Benjamin and identical reproduction. But the first person who built a cultural theory of identical copies was five centuries earlier: Leon Battista Alberti. He had no machines. When Alberti came up with his theory of making identical copies, print did not exist—or rather, it existed, but he did not know it. He was writing *De re aedificatoria* around 1452, exactly at the time when Gutenberg was inventing printing with movable type. But Alberti did not know it. He assumed that texts would continue to be reproduced through scribal copying, as they always had been. He knew, as a humanist scholar, that copying introduces errors. Every copyist makes mistakes. That is why Renaissance scholars, when rediscovering classical texts—Vitruvius, Cicero, Quintilian—found multiple versions, all different. This led to the invention of philology: reconstructing the original text by comparing copies. So, Alberti understood that textual transmission is unstable. For this reason, he refused to use images in his treatises. He preferred *verbis solis*—words alone—because he believed that the drift of textual errors

could be controlled more easily than the drift of visual ones. A drawing copied by another artist could deviate significantly from the original. Words, even if corrupted, could be more easily corrected through comparison. Thus, Alberti translated visual ideas into textual descriptions, in order to preserve authorial control. That was the condition of scribal culture. Then Gutenberg's printing changed everything. Within a generation, printing spread across Europe—not only because of accuracy, but because it was cheaper. Identical reproduction became both technically and economically dominant.

Now, Alberti also defines architecture in a very specific way. At the beginning of *De re aedificatoria*, he writes that architecture is not a building, but an idea conceived in the mind, expressed through drawings and models, and then executed without change.

There is:

- first, an idea
- then a drawing (an identical copy of the idea)
- then a building (an identical copy of the drawing)

This only works if identical copies are possible. Alberti's theory depends on reproducibility—even though the technology did not yet exist. That is why he is so important. He anticipates a technological condition that would only emerge later. He never saw his book printed. He died before its publication. But he had heard that in Germany there was a machine capable of producing more copies in one hour than a scriptorium could in a week. He understood that something was changing, even if he did not fully experience it. In that sense, the logic of identical reproduction is not Benjamin's invention—it begins with Alberti.

AP: Recent developments in machine learning, particularly generative models, have reintroduced the notions of style and imitation (*mimesis*) into technological practice. If such systems operate by recombining precedents within datasets, what does this suggest about the role of imitation in artistic creation? And how does this relate to the parametric “space of possibilities,” where potentially infinite variations can be produced and where the idea of “mass individualisation” first emerged?

MC: Mass customisation we have already discussed. But imitation is the core topic here. We think, as modern people, that imitation is a bad

thing—something we should avoid. If a student in an architecture school is told, “you are imitating,” it means failure. It means you will not become a good architect. We have this prejudice: imitation is a crime, something to be hidden, something shameful. But this is a modern idea. In the classical tradition, imitation was the foundation of most art, science, and technology. It was called *mimesis* in Greek, *imitatio* or *emulatio* in Latin. The ancients did not always define these terms precisely; instead, they used stories. For example, the painter Zeuxis. One story: he painted grapes so realistically that birds came and tried to eat them. That is imitation as perfect copying—realism. But there is another story. He was asked to paint the most beautiful woman—an ideal figure, someone like Helen of Troy. He searched for a perfect model but could not find one. So, he selected five different women and combined their features. The final image did not copy any one of them, but incorporated the best of each. People could recognize elements, but could not say exactly which part came from which model. He produced something better than any individual instance of reality.

There are two kinds of imitation, then:

- imitation as copying reality (the grapes)
- imitation as improving upon reality (the ideal figure)

The second is what we may call creative imitation. You do not imitate one model—you imitate many, select, transform, and produce something new: similar to all, identical to none. This is how imitation was understood in the classical tradition and in the Renaissance. This idea was applied not only to images, but also to writing: how to imitate Cicero or Virgil without becoming a mere copyist. The bad imitator is like a mirror or a monkey—someone who reproduces the original without transformation. The good imitator creates something new by selecting and transforming. This model of imitation remained dominant until the Industrial Revolution. With photography, people realized that machines could reproduce reality better than humans. Imitation lost its prestige, and modern art moved away from it. Now, with generative AI, imitation returns. But it does not return as copying—it returns as transformation. Generative AI does what Zeuxis did, but with millions of models instead of five. It processes vast datasets and produces outputs that are derived from them—similar to all, identical to none. This is not copying. It is inspiration, assimilation, transformation. That is why our traditional legal

and conceptual frameworks—copyright, originality—struggle to apply. Copyright is based on identical reproduction. You pay royalties for copying. But there is no “right of inspiration.” If you are inspired by Leonardo, you do not pay royalties. Only if you reproduce his work identically. Generative AI operates in this domain of inspiration, not copying. It works more like the human mind than like a mechanical reproduction system. And this is precisely why we struggle to understand it: for the past two centuries, we have modelled creativity on machines, forgetting that human creativity does not work that way.

Making identical copies is difficult for humans—we are not photocopy machines. But being inspired, transforming what we see, producing something new—that is what we do all the time. That is how the human mind works. If you want identical copies, use a machine. But inspiration and transformation—that is human.

MARCO VANUCCI: Michael Polanyi famously argued that we “know more than we can tell.” Today, AI systems seem to reverse this condition: machines produce results that even their creators cannot fully explain. What does this opacity mean for architectural knowledge?

MC: Yes, this is a very important point. Polanyi’s idea—that we know more than we can tell—refers to tacit knowledge, to forms of understanding that cannot be fully articulated in explicit terms. Now, with AI, we have a different situation. Machines produce results that even their creators cannot fully explain. This is what we call opacity. But this is not entirely unprecedented. In many fields, especially in craft and artistic practice, we already operate with forms of knowledge that are not fully explicit. A skilled artisan may not be able to explain exactly how they achieve a certain result. Opacity is not entirely new. What is new is that it is now embedded in a technical system. With rule-based systems, we could always reconstruct the logic. With machine learning, we cannot. So, we move from a system of explicit knowledge to a system where results emerge from processes that are not fully transparent. This raises important questions for architecture, because architecture has traditionally relied on forms of knowledge that could be communicated, transmitted, and justified. Now we have systems that produce results without fully explainable reasoning. The question becomes: how do we validate, how do we trust, how do we use such knowledge? This is still an open question.

MV: Architecture has historically used computation as a tool for optimisation and control. Yet artists have often embraced technical systems for their unpredictability. Do you think architecture has underestimated the creative potential of computational indeterminacy?

MC: Yes, I think that is true. Architecture has traditionally used computation as a tool for control—for optimisation, for efficiency, for precision. But computational systems can also produce unexpected results. Artists have often explored this dimension—using systems not only to control outcomes, but to generate variation, unpredictability, even error. In architecture, this potential has not always been fully embraced. Partly because architecture is a discipline that must deal with constraints—technical, economic, regulatory. But also because of a cultural preference for control. Now, with generative systems, we are beginning to see a shift. Indeterminacy is no longer simply a limitation; it can be a resource. But again, this requires a change in mindset.

MV: Universities once functioned both as institutions that selected social elites and acted as gatekeepers of knowledge. Today knowledge is widely accessible and higher education has become increasingly democratised. In this context, what role should educational institutions play when professional practice no longer depends on mastering a fixed body of knowledge, but on the ability to think critically and adapt continuously?

MC: Yes, this is a major transformation. Universities were historically institutions that controlled access to knowledge. Knowledge was scarce and education was the way to access it. Today, knowledge is widely available. However, the role of education is changing. It is no longer primarily about transmitting a fixed body of knowledge. It is about developing the ability to think, to interpret, to adapt. In architecture, this is particularly important, because the tools, the technologies, the conditions of practice are constantly changing. Education must focus less on stable content and more on critical thinking. This does not mean that knowledge is no longer important, but that its role is different. The challenge is to redefine what it means to be educated in a context where knowledge is no longer scarce.

MV: Could what Jean Baudrillard described as “copies without originals” help us understand what we are witnessing with generative AI today?

MC: Copies with too many originals, which is the same thing, but originals that are no longer traceable. We can put one million originals in the dataset, but we have no idea what generative AI does to transform them. We call it a black box.

MV: If we look at generative and algorithmic art, particularly in the 1960s, artists often embraced unpredictability as a productive dimension of working with machines. Could this offer a way for architects to reestablish agency today—by engaging with indeterminacy rather than control?

MC: That is the degree of unpredictability. The etymology of cybernetics was originally κυβερνήτης, the pilot of the boat. One could say that the real avatar of cybernetics today is the self-driving car, as Norbert Wiener imagined it in 1948. And what it has to do is not to be rule-based, because you don't know when you're driving if that cat is going to cross the road or not. You must have a way of reacting to what that cat does. You see the cat, but you don't know if it's going to cross one way or another. This is the unpredictability to which the human is capable of giving a response. A traditional rule-based machine cannot react to that. But the principle of cybernetics—feedback—is what today the self-driving car assumes and is capable of doing in the way we do it. They train the self-driving car not with rules but with a dataset of incidents. They have a dataset of one million cars meeting one million cats. Out of this dataset, they extrapolate statistically what is the best thing to do. It is a statistical probability, which no human can calculate. If you are a human driving that car, you can pull from your experience of something similar you have seen. But a dataset with millions of examples of black cats about to cross a road is likely to know more about cat behaviour than we do. So, it would probably avoid more cats than we would.

MV: Therefore, to quote Polanyi again, “we know more than we can tell,” which can be suitably applied to what an architect traditionally goes through in the education process. We tend to know things that we cannot fully explain because there is a certain sensitivity that we develop. The late Yona Friedman argued that architecture is best learned through apprenticeship rather than purely formal education. This perspective aligns with Michael Polanyi's notion that “we know more than we can tell.” In architectural education, much of what is learned cannot be fully articulated, as it involves developing a certain sensitivity through practice.

Therefore, the experiential nature of apprenticeship becomes essential. How does AI interrogate the way we learn?

MC: Anthropologists have a name for that—they call it tacit knowledge. It was thought to be the escape door of the idiots—those who cannot explain what they are doing. I remember so many teachers in design studios who, instead of coming with a theory or rules, would say: I don't have a theory. Just look at me and do what I do.

But this is imitation learning, which is the traditional attitude of an artisan. Think of a bricklayer, or a stonemason building a dry stone wall. There is no handbook that explains the most effective way of doing that, because every stone is different. You cannot learn that from books. You learn by following a master—watching, imitating. The master does not give you a book and say, “read this.” He says, follow me, watch what I do, and one day you will do the same. Imitation learning. Now, oddly, generative AI works that way. If I have a dataset with one million bricklayers laying bricks, the machine will imitate that without knowing any rules. That's the problem: monkey see, monkey do. Monkeys can imitate, but they cannot explain what they are doing. This is tacit knowledge. We are training monkeys. Computers are today's automated monkey. I don't know if it's a good thing.

Perhaps with a good recipe book, I could learn to make good risotto in one week and not in twelve years. But it's all about learning by rules or learning by observation. We can learn a lot of things without knowing any rules. When we go to school at the age of six, we already know how to speak, even though we cannot yet read. By simply imitating the sounds we hear—our parents—we learn to communicate. We have an extraordinary capacity to imitate, which is specific to primates—monkeys, apes, and us. Dogs cannot imitate; cats cannot imitate. They learn by trial and error. So having a machine that can now imitate in this way is somewhat disconcerting.

MV: In the past, you would go to university because there was a library. Now knowledge is everywhere. Institutions are no longer selecting the ruling class. There is democratization of knowledge and education. What should we do?

MC: What you describe is the modernist dream—a social democratic dream where universities are responsible for transmitting and protecting

knowledge. Does it still work that way? Nobody gives a fig about knowledge today. Universities are universally despised as places where people replicate useless knowledge. We live in a world where violence and action are more important than competence and expertise. There is a hypothetical which I always use in this case. Imagine living in an Alpine valley. At some point you realize that your pastures do not grow enough wheat to feed your population because of climate change, or there is no rain anymore, or the wheat is not strong. People are starving. There used to be food and now there isn't. What do you do? There are two paths. Number one is science: you become an engineer. You improve the yield of the pastures and the wheat, the quality of the grains and the tools you are using in agriculture. With technology, you improve the performance of the land you live on. You feed your people again with better technology. Number two: you cross the watershed, you go to the next valley, you kill everyone and steal their wheat. The choice is between war or science. During the second half of the twentieth century, we thought that science was the good solution. And now again it's war, as we see. It's easier to go and steal. Because if you're an engineer, you don't need war. You don't need to steal your neighbour's wheat if you can get better wheat from your pastures. Perhaps that's one thing universities can still do: train our future engineers, hoping we shall still need them.

Interview conducted by Andreja Prokopijević, Željko Radinković, and Marco Vanucci.

ARCHITECTURE AS INTERACTION: A CONVERSATION WITH THEODORE SPYROPOULOS

KHÖREIN: In your work with students at the Design Research Lab at the AA School, architecture seems to be approached less through predefined methods or rules and more through conditions in which control is partially suspended. Would you describe your pedagogy as consciously non-didactic, and what kind of knowledge emerges when learning cannot be fully articulated in advance?

THEODORE SPYROPOULOS: Control from this point of view is an illusion and, one could say, it limits and prescribes the nature of enquiry. Thomas Kuhn in *The Structure of Scientific Revolutions* spoke to this when he critically suggested that within the domain of research the questions that we ask inform the answers that we find. This is both obvious and profound when explored within science and the pursuit of knowledge. For me the issues around design research remain an active search through thematic subjects which by necessity need to be approached in an open and non-didactic manner. Part of my role as the Director now for over two decades is to foster a culture that is curious and experimental as a means to challenge the orthodoxies of response. To explore ideas through what the subjects demand rather than a methodology that projects a style or predefined solution space. This means moving beyond a prescriptive response into an active form of learning that is engaged with the capacity to work collectively on the complexities that the research necessitates. This is not experimentation for the sake of experimentation but the means to examine critical questions that challenge the means we work and apply this research. This is for all who participate in this experiment we call the DRL. A culture where staff and students are exploring complex ideas as a collective, a team. No hierarchies but leveraging our experiences and work in a way that assists us to work through a great deal of challenges: conceptual, social, and technical together. Where the

outputs of research are examined from multiple perspectives and where risk is encouraged to push the boundaries.

Knowledge from this perspective is constructed through active engagement that approaches subjects as spaces of enquiry, while attempting to build a literacy that is shared and examined in, for example, simulation environments, observational experiments, historical case studies, and within many bodies of knowledge that are usually outside of the discipline of architecture. This manner of working is in constant dialogue with the material and each other and I believe gives agency to our students and staff bringing a vitality to the collective search. It is a mutually supportive environment that is working with a belief that design can offer people more. Creating a meaningful engagement that challenges all forms of conceptualisation.

KH: Many projects developed under your mentorship generate form through distributed interactions rather than individual intention. How do you understand authorship under these conditions, and what changes when the architect is no longer the primary source of form?

TS: This approach explores both human and non-human forms of interaction. In attempting to humanise technologies and respond to some of today's challenges, the work moves beyond the authoring of buildings to explore architecture as infrastructure, phenomenon as technology, material systems as self-aware and self-assembled systems, and urbanism as adaptive ecologies. My approach challenges individual authorship by recognising that some questions require forms of inquiry that can only be addressed collectively, due to the scale or complexity of the problem. Within this process, the role of the individual within the team shifts as the research develops. It is a co-authorship model that is diverse in its approach and plural in its inputs, both within the team and in response to the technologies or subjects being explored.

Historically, the role of the architect has always involved interfacing with others and sustaining an active dialogue in the realisation of ideas. Yet the privileging of individual authorship remains an orthodoxy within educational structures and in how the architect is projected in society. In the introduction to my PhD titled *Constructing Participatory Environments: A Behavioural Model for Design*, I attempted to address this idea by describing architecture as an ecology of interacting systems. It looked to people like Gordon Pask that sought to think of interaction models

through what he spoke about as Conversation Theory. This could be seen in the development of projects that embraced this open framework in a project like the Fun Palace with Cedric Price and Joan Littlewood as an example.¹

KH: Having worked in proximity to figures such as Peter Eisenman and Patrik Schumacher, both of whom are associated with strong and highly articulated theoretical positions, how did these experiences influence your own approach, particularly your move toward open-ended, experimental and system-based forms of practice?

TS: I was fortunate early in my career to be exposed to environments that never created an artificial divide between theory and practice. This resonated with me, and in particular in Peter's office, where this was clearly manifested within architecture. In preparation for his lectures for his classes at Yale or Princeton, he would share and work through these theories and speculations with us during our lunchtime in the office. I enjoyed this engagement intellectually and found it very important for me in engaging with and contributing to the work at the time in the office. During this period, aside from working on competitions with him, we had projects such as the City of Culture in Galicia, Spain, the Berlin Memorial, and the Arizona football stadium being worked on. As the office was small and the work complex, it was more the setup of a European football team, where we all contributed in the ways the projects necessitated. It was through Peter and this period in the office that I understood what a project in architecture could be. It was a continued education for me as I joined post my studies at the Architectural Association.

¹ "This thesis proposes the design of cybernetic frameworks that attempt to explore architecture as ecology of interacting systems that move beyond the fixed and finite tendencies of the past towards spatial environments that are adaptive, emotive and behavioural. Environments within this framework are attempts to construct interaction scenarios that enable agency, curiosity and play, forging intimate exchanges that are participatory and evolving over time. Interaction understood as the evolving relationships between things allows a generative and time-based framework to explore space as a model of interfacing that shifts the tendencies of passive occupancy towards an active ecology of interacting agents. The work argued here moves away from known models that reinforce habitual responses within architecture, towards an understanding of adaptive systems that are active agents for communication and exploration. Architecture within the context of this thesis is explored as a medium for spatial interfacing. Design is thus considered as durational, real-time and anticipatory exploring human human, human machine, and machine machine communication. The challenge posed is how designers can construct environments that are shared, enable curiosity, evolve and allow for complex interactions to arise through human and non-human agency."

In the work of Patrik and Zaha, this was different. In Zaha, you had an intuitive force that had a vision and sensibility that was formally expressive and elegantly delivered through various media. Patrik theorised this and evolved it through models of organisation, translating this vision to the scale that the work was radically evolving towards. Their collaboration shaped a practice that created a framework to allow an atelier-like office to scale from 5 to 500 and maintain a sensibility. Patrik's theoretical position, which he articulated in *Autopoiesis of Architecture* (2010) and through the term of Parametricism that he coined in 2008 at the Venice Architecture Biennale, was always in direct dialogue with the DRL and his work within the lab. In particular, this emerged when we set up a research agenda in 2005 with colleagues titled Parametric Urbanism. Here, Patrik incubated many of the emerging aspects of his design theory in dialogue with the work at the office on masterplans, such as One North in Singapore (2001), which were a sign of things to come.

My theories drew from art movements of the 1970s New York art scene, quantum physics, computation, adaptive systems, and second-order cybernetics. That manifested in the work that I have pursued through Minimaforms, my two decades of directing the DRL, publications like *Adaptive Ecologies*, and my PhD, where I wrote on relationality and behaviour and approaches towards design that folded uncertainty and latency as important factors in the nature of how and what we worked on.

In contrast to Patrik's parametric approach, I developed a theoretical framework for a model of urbanism that was not guided by top-down formalism, but by generative and behavioural systems that were emergent and relational. This was a position from my side that was against master planning and sought examining poly-scalar relationships that foregrounded computational frameworks that allow for change. Through an active dialogue, this project evolved into a discourse that was applied in proposals and created meaningful project discussions that have remained spirited within the DRL. Though there are at times radically different points of view between our work within the DRL, our collaboration has offered a dynamic and multifaceted point of view that is plural. Nurturing an environment where differences in the development of research are welcomed and important. This for me affords a context to prototype ideas and engage with collaborators who offer a deep engagement with the project of architecture. This within a collaborative design research environment the likes of the DRL allow an explorative model for design.

In my work, if the issues could be explored spatially, then I believed they were architectural and our knowledge could have an important contribution. I have allowed that understanding to lead me, and my efforts to be shaped through open experimentation. The shift from representational forms of communication to active prototyping, and from human to non-human engagement, allowed me to move intellectually and in the realisation of the work in a dynamic manner. Rather than applying a language, I allowed for languages to be born through active and collective research and experimentation.

KH: Collapse, deviation, and failure appear frequently in experimental work at the DRL, though not as mistakes in a conventional sense. Has error become a productive condition of architectural thinking, and how do you distinguish it from a genuine loss of disciplinary criteria?

TS: Failure here is not failure. Rather, it is an attempt to examine and construct an approach that moves beyond forms of representation toward the active prototyping of ideas. By constructing proofs of concept, the work enables more informed questions and allows the research to generate an active feedback loop. As Samuel Beckett expressed beautifully: “Ever tried. Ever failed. No matter. Try again. Fail again. Fail better.” I would argue that this does not represent a loss of disciplinary criteria, but rather a critical enquiry into the ways we explore design. It purposefully opens up habitual responses to questions, allowing us to engage with them in more meaningful ways.

KH: In your work, where architecture unfolds over time rather than stabilising into a fixed form, how does this affect the way space is remembered or internalised? Does temporality itself become an active agent, participating in the shaping of spatial experience rather than simply framing it?

TS: Our work is conceptually about formation rather than form. Form here is understood as a moment in time. This draws from our understandings in the sciences and has served to influence the name of our practice, Minimaforms. The name recognises that form is arrested in time, and that this moment of understanding is the minima—when form comes into being.

With Stephen, my brother, we wanted to explore work that was emotive, behaviour-based, that gave agency and formed a collective memory.

In works from *Memory Cloud* to more recent projects such as *Of and In the World*, observers are active participants rather than passive observers. The approach is motivated by enabling agency and constructing frameworks that unfold over time.

KH: Your projects often unfold over time rather than arriving at a final state. How does this temporal openness affect how architectural work is evaluated, documented or remembered?

TS: In my work with *Minimaforms*, and with my students and staff at the DRL, there is a conceptual shift from the fixed and finite toward the dynamic and evolving. There is a deliberate questioning of the limits of representation, accompanied by a move toward the active prototyping of ideas.

The shift from form to formation was an early driver of the work, recognising our world as something that unfolds in time. It was an attempt to examine agency and to develop models that consider intelligence as something not unique to humans. This explores other forms of intelligence, ecologies, and environments as active participants in our architectural and artistic developments.

The work seeks to develop design systems that remain open frameworks for spatial interfacing, while recognising that many issues may remain latent or uncertain. This uncertainty as the only certainty which necessitates an approach to consider adaptive systems, systems of learning and behaviour. The aim from my perspective is to consider relationality as would be described in quantum physics as means to understand space time. To consider intelligence not as an attribute to a thing itself but a demonstration between things. This brings the world of quantum physics and second order cybernetics together for me.

KH: Your work engages many ideas historically associated with cybernetics—interaction, feedback, behaviour. Do you see this as a continuation of that lineage, or has contemporary machine learning shifted these questions into a fundamentally different terrain?

TS: The work of second-order cybernetics has been an important reference in developing systems and frameworks, particularly those exploring models of the brain and systems of learning. This includes cyberneticians such as Gordon Pask, Gregory Bateson, and W. Grey Walter. Pask is perhaps most well-known through his work in architecture with Cedric

Price, John and Julia Frazer, and MIT's Architecture Machine Group, as well as his exhibitions of cybernetic sculptures in seminal exhibitions such as *Cybernetic Serendipity* at the ICA in London.

What I find important in this work is that much of it was conceived as thought experiments, without direct access to the technologies that now make up our contemporary condition. From AI and machine learning to the concept of the black box, it was conceptually vast in scope and offered a means to examine complexity. Pask, in his paper "The Architectural Relevance of Cybernetics," spoke to the fact that, for him, architecture did not yet have a theory or means to deal with complexity. His interest in architecture was grounded in the belief that cybernetics could become that theory.

In our work, in particular that which explores interaction and systems design, we examine Pask's concept of conversation theory as a model to explore more participatory means of engaging the everyday. His work explored learning and novelty, while our work explores collective intelligence and emotive response. In a time where our AI systems are designing systems, we once again find ourselves in a condition of black box understandings—algorithms designing algorithms that offer solutions but no means to understand them within the knowledge structures we have been educated in.

This necessitates new forms of understanding and participation with these systems. So yes, it has evolved into something new, though that period was faced with similar engagements that were not disciplinary-specific, and I believe their questions are ever more important in understanding how we arrived here. My attempts are to demystify and humanise technology.

KH: In projects involving adaptive systems, you have explored installations that use machine learning to develop emotive forms of interaction, communicating through movement, sound, and other sensory cues. What would it mean to live with such systems architecturally, and how do they complicate distinctions between automation, autonomy, and affect?

TS: An example that illustrates this approach is Minimaforms' project *Petting Zoo*. The intention of this work is to respond to a robotic and automated future that many of my colleagues embrace as a dominant model of production. *Petting Zoo* explored the more complex question of how we might live with these systems. It marks a purposeful shift away from

the drivers of production, assembly, and fabrication toward emotive and behavioural forms of interfacing. The project is a speculative, life-like robotic environment that raises questions about how future environments might actively enable new forms of communication with the everyday. The artificially intelligent creatures are designed with the capacity to learn and explore behaviours through interaction with participants.

The installation exhibited life-like attributes through conversational forms of interaction, establishing communication with users that is both emotive and sensorial. Conceived as an immersive installation environment, social and synthetic systems of interaction allow the pets to engage with participants and evolve their behaviours over time. The pets stimulate participation through animate behaviours communicated through movement, sound, and illumination. These behaviours evolve through continued interaction, enabling each pet to develop distinct personalities. Interactions are stimulated through engagement with human users as well as with other pets within the population. Intimacy and curiosity act as enabling agents, externalising personal experience through direct visual, haptic, and aural forms of communication.

These experiments explore artificial intelligence as a means of prompting us to consider how we might co-evolve and inhabit future human-machine environments. Moving beyond robotics understood primarily as tools of production, *Petting Zoo* examines the emotive and behavioural dimensions of our engagement with these systems and with each other. Over time, and in response to habit, the pets can be understood to develop a form of “personality.” As the spatial orientation and environmental conditions of the installation generate diverse interaction scenarios for each pet, they gradually develop a unique capacity to exhibit varied behaviours. These personalities evolve through a process of systemic self-observation, opening new territories for exploring how we might design systems that evolve alongside us.

It speaks to the capacity of humans to see life like attributes in the synthetic systems. To humanise technologies and our empathy towards them when they are designed with features that engage us and stimulate curiosity.

KH: If, as Michael Polanyi argues, we know more than we can explicitly articulate, how should architecture relate to the opaque systems it increasingly relies on? Can the black box function as a meaningful architectural condition, rather than merely a technical limitation?

TS: I agree with the argument made by Polanyi. Quantum physics reminds us that how we see the world and how the world actually is can be very different. The “black box” therefore remains an important conceptual apparatus. From early cybernetic thinking, beginning with the Macy Conferences, to the present, it has served as a way of framing thought experiments that allow us to move both conceptually and technically in ways that are powerful because of their generative capacity.

Today this issue re-emerges in the need to transform the black box condition into a meaningful framework for engaging with our ever-evolving technological present. For instance, we now design processes that themselves design other processes, as seen in many developments in AI. Algorithms are increasingly designing other algorithms, and we often have little meaningful access to track or fully understand how their outputs are produced.

As we move toward more agentic technological futures, this condition will become increasingly present in everyday life. It therefore requires new forms of thinking and learning to emerge. In my work, I believe that behaviour and conversation design theory will play an important role in addressing this challenge.

KH: Architecture has long projected expectations about use and behaviour into spatial form. In responsive systems based on feedback and continuous adjustment, this anticipatory role seems to give way to ongoing modulation. How do you understand this shift, and what does it imply for architecture’s political role today?

TS: This shift has extended into all aspects of life as the communication revolution has transformed both the ways we communicate with each other and the systems that communicate with us. Moving toward an anticipatory model of architecture requires recognising that architecture often remains a form of resistance to these changes. Part of this resistance lies in the implicit time required to realise buildings, but a more significant factor is the habitual way we relate to the built environment; as one of the few legible and consistent artefacts in an otherwise rapidly changing world. However, the conservatism that often shapes our approaches to architecture and urbanism can be dangerous. It risks perpetuating models of development that repeat past mistakes rather than examining meaningful alternatives. Engaging with change, therefore, becomes a necessary and important strategy, one that responds to the accelerated

demands placed on architects and to the volatility of the forces that shape these decisions, whether market-driven, social, or political.

Allowing systems themselves to adapt and evolve could lead to more informed decision-making and create opportunities for more participatory forms of development among diverse stakeholders. Some of the work at the Architectural Association's Design Research Lab (DRL), for example, explores gamified environments that allow people to better understand their surroundings and potentially co-design the spaces they inhabit. Such approaches aim to amplify voices and create learning environments that can influence urban decision-making and design.

An important reference for me in thinking about anticipatory frameworks is Buckminster Fuller's World Game, developed for Expo 67 in response to the exhibition's theme, *Man and His World*. Fuller approached the planet as a unitary system, seeking to make visible the conditions required to "make the world work." His goal was to help humanity design a sustainable future through a conscious exploration of cause and effect, using gameplay to address issues such as energy, poverty, and resource distribution. In doing so, Fuller attempted to bypass what he saw as the limitations of politics and economics constrained by nation-state thinking, which he believed had become an obstacle to solving planetary-scale problems. The ambition of the World Game was to explore how the world might function for 100% of humanity through spontaneous cooperation.

This motivated approach challenged conventional forms of representation. Fuller's efforts included developing a new map projection that minimised territorial distortion, while also constructing one of the earliest interactive interfaces and gamified environments. It was a model that framed a conceptual problem in which design itself played a central role in reimagining how the world might be organised.

KH: When architectural processes are shaped less by discrete decisions than by the calibration of interacting systems, the capacity to act can no longer be grounded in authorship or full foresight. Under these conditions, what distinguishes architecture from engineering, environmental systems, or infrastructural management? Is this distinction a matter of technique, judgement or a particular way of framing problems rather than solving them?

TS: Rather than seeing this as a means to continue treating systemic decisions as products of specialisation, I see it as an opportunity to pursue

a greater synthesis. This tendency can already be seen in emerging sciences that seek broader integration, from genetics and biology to studies of the ecologies in which they operate. To return to some of your earlier questions, it offers a way of seeing the world not as a collection of objects but as a network of relations. This perspective appears in many forms, from the work of physicist Carlo Rovelli to Jack Burnham's writings on kinetic and systems art in the 1970s.

Ultimately, it is a way of framing our understanding of the world. We exist both outside and inside this understanding, acting simultaneously as observers and participants within it.

KH: If adaptive spatial systems shift architecture from a discipline that stabilises knowledge into one that continuously responds to behaviour, feedback and change, how does this transformation affect the production of architectural knowledge itself? Can knowledge still be understood as cumulative, or does it instead emerge as something situational, distributed and contingent?

TS: Intelligence, from my perspective, is not an attribute of a thing. It is relational and demonstrated between things. Knowledge today is ever more important to be understood, and how this will evolve with forms of intelligence that aggregate human knowledge. From my perspective, treating these AIs as conversational partners is an important start, allowing the plasticity of our own knowledge structures to engage in new forms of learning and understanding. A co-intelligent system recognises that what we speak of regarding these systems today is modelled on us and our data. If we recognise that all things have intelligence and leverage these models in a complementary way, we can find the means to address the immediacy of issues, both social, political, and ecological. Only when we shed this notion of explicit control and actively explore meaningful and I will restate, human pursuits that are responsible, will we be able to understand our relationality to the world as necessitating situational and distributed forms of knowledge structures that are context-aware and coherent to the issues that we are facing.

Interview conducted by Andreja Prokopijević, Sara Dragišić, and Snežana Vesnić.

TECHNO-POLITICS: A CONVERSATION WITH BERNARD CACHE

Bernard Cache is among those who have most rigorously examined the promises—and limitations—of digital technologies in architecture. Yet, despite his credentials, it is difficult to imagine a more ironic beginning to our conversation about enabling technologies than watching them fail. We spent nearly half an hour attempting, and failing, to establish a video connection. The cause of this technical glitch remains unclear. Nevertheless, it is a fitting prelude.

Trained across architecture, economics, and philosophy—most notably under the guidance of Gilles Deleuze—Cache has consistently woven together theory, computation, and non-standard modes of production. He anticipated the architectural consequences of CAD/CAM technologies and articulated the concept of the objectile: an architectural object understood not as a fixed form, but as a “continuum through variation,” a field of potential transformations: an idea that continues to shape contemporary digital practice.

Cache asks that we speak only of Desargues, Pascal, and projective geometry—historical figures and conceptual frameworks to which we will indeed return. Still, I cannot resist asking what he is currently working on, and what occupies his thinking today.

BERNARD CACHE: I am still working on topics related to projective geometry, and I am currently supervising two PhD students whose research overlaps with my own and focuses on the use of open-source software—specifically Speckle, developed by the Romanian architect Dimitri Stefanescu. Speckle’s initial aim is to develop an application that enables real-time exchange within architectural practice.

MARCO VANUCCI: The *open source* is, first of all, a political issue.

BC: Yes, I emphasize this because I believe open source is a central issue today. Not only in an ideological sense, but in a very concrete political one: it directly affects the everyday life of architects and architectural offices. If, for any reason, a studio cannot afford commercial software licenses, it may not even be able to view its own work. This is extremely serious. It means that you do not truly own what you produce. We need to respond to this state of affairs.

MV: In your most recent writings, a very critical stance emerges toward the way in which digital technology has permeated our lives. Yet, the enthusiasm with which you and other pioneers of parametric design embraced computer-aided design and manufacturing was grounded in the context of a new world order characterized by open markets and the free movement of goods, people, and information. A world in which the neoliberal project seemed to have delivered the “end of history” and, with it, thanks to the development of digital technology, new extraordinary possibilities to design a different society, free from the limitations of one-size-fit-all ideologies.

Today, that project is in crisis: the world is marked by protectionism, sovereigntism, and autocracies: digital technologies, instead of uniting us, have contributed to dividing society. How do you experience this change?

BC: My position today is one of retreat within the folds of history in order to change the future. More precisely, I am looking at things that were about to appear in the past but did not. The idea is that, with a broader past, by actualizing things that remained virtual, we might have a better future. Otherwise, I am deeply pessimistic about the future: I believe we are heading toward a stalemate. I would recommend reading *Sans transition* by Jean-Baptiste Fressoz,¹ which shows that there has never been a true “energy transition,” since technological systems have always been deeply intertwined. It is an illusion to believe that we can simply abandon unsustainable or obsolete energy sources.

Reversing course is extremely difficult. When I worked on digital technologies, I operated within a structured professional environment grounded in mathematics, rules, and syntax. I could not have imagined that speaking to one’s children or paying taxes would one day require the use of extremely expensive—and often poorly designed—digital tools.

¹ F. M. Clarke, D. E. Smith, “‘Essay pour les Coniques’ of Blaise Pascal,” *Isis*, 10, 1, 1928, pp. 16–20.

Today, it is almost impossible to participate fully in social life without owning a device that costs between 500 and 1,000 euros. These technologies exclude the poor, the elderly, people with disabilities, and those who are digitally illiterate. We are creating a new form of illiteracy. One day, I would like to work on a public digital terminal—standardized, affordable, and based on a minimal shared syntax. The first step would be to introduce mandatory public standards for the user interfaces of all services of public interest—for example, by requiring a single, universal command that allows users to return to a neutral point² in any application. The second step would be to build consensus around a basic, low-cost, sustainable terminal for every citizen, since we can no longer function as human beings without such devices. This terminal should then be produced locally, at least on a European scale, while also involving collaboration with developing countries.

MV: The first issue, then, is not only technological but syntactic.

BC: Exactly. I insist on the word *syntax*. The problem of false truths is not only a matter of content; it is also a matter of syntax. We opened digital technology to the world without providing the structures needed to make it intelligible. From Tamagotchi to Facebook, entry into the digital realm has been widespread but unstructured. Today, no one wants to write code anymore; no one wants to think about graphical syntax. And yet, it is syntax that makes truth possible. This problem is in fact deeply rooted in our history. Even in mathematics, geometric figures have always been considered secondary relative to the text, which means, for instance, that a figure could be drawn symmetrically even though the theorem relied on an arbitrary configuration, not restricted to a symmetrical one. As Mario Carpo³ has emphasized very well, images were always neglected in classical culture, and then, were treated as if they were no graphical syntax.

MV: You are referring to the pervasiveness of digital technology in everyday life.

BC: Yes. Even more than a political issue, this pervasiveness has become an anthropological one, just as in geology we now have to speak of the

² A neutral point means a state in which the software is not engaged in any command or action.

³ M. Carpo, *Architecture in the Age of Printing*, The MIT Press, Cambridge, MA, 2001.

Anthropocene due to the pervasiveness of our technological impact on nature. This anthropological pervasiveness of our minds and bodies is such that, for example, when I go to the mountains, I find myself among the few who hike without wearing a watch that tracks their heart rate. It is astonishing. Today, in Switzerland, there are people who cycle through the Canton of Vaud while tracing the borders of the entire country on their GPS, all within the same region. They call it a work of art. But the artwork is the route, not the digital experience. What truly concerns them is the digital object.

I never imagined that digital technology would penetrate every aspect of life so deeply.

MV: This concern about the “tyranny of the algorithm” is not new. Architect Luigi Moretti, the father of parametric design, already in the 1970s, shortly before his death, warned against extending algorithmic logic outside controlled disciplinary contexts. He said it could not be indiscriminately scaled to all aspects of life.

BC: That is true, and we can see it clearly today. In France, for example, there is debate about the use of mobile phones in schools. A device that scrambles signals would be enough, and could be switched on and off at will, according to an appropriate time schedule: students would still have their phones in the classroom, but could not use them when it is inappropriate. In case of an emergency, they could use a landline, just as in the past.

MV: So, school as a protected space, a place free from the invasiveness of digital devices.

BC: Exactly. It is troubling that we can no longer imagine society as having technology-free spaces—places where there is no digital communication, only communication between people, or silence. School should be one of those spaces. This is not a reactionary argument; it concerns the very possibility of autonomy. How can a child become autonomous if they are constantly tracked, reachable, and controlled?

MV: Ultimately, the broader outcome of the digital revolution has been precisely this possibility of redesigning an atomized, granular society, made possible by economies without scale. This has led to extreme individualism. Everyone has their own device, their own “scale.”

BC: Yes, individualism and isolation. The beginning of this conversation (with the problem with the camera) was perfect, in the most perverse sense. What surprises me is that Europe does not respond with its own tools. Why do we not have European open-source software for communicating among ourselves? Why did the LiMux⁴ project, which proposed open-source software for the public administration of Munich and a few other cities, fail? Why did Paris not join the project? Why do public administration not provide their employees with Fairphones⁵ for professional use. Why does the French army grant contracts to Microsoft without making serious bids? Why has the French school administration just renewed its contract with Microsoft while the government claims to be looking for independent solutions? European servility amazes me. Europe could form new alliances, for example with the Maghreb and the Mashreq. We have enormous intellectual resources in Algeria, Egypt, and other neighboring countries. But you need to establish the rules of the game. Without syntax, nothing can be built.

MV: Back to syntax. What exactly do you mean?

BC: The fact that there are public, shared, slowly evolving rules around which consensus can be created. Syntax is a stable reference. Take Wikipedia. It is one of the few large successful projects in the digital world. I use it a lot and support the project financially. But recently, when I tried to donate, I could not find “Switzerland” in its place in the alphabetical list of countries. There are also elementary lexical considerations: we should not use words that are not in the dictionary unless there are sound reasons to do so.

MV: So syntax can be understood as the foundation of truth.

BC: Exactly. Without syntax, there is no truth. Truth is even more a question of syntax than of content. And when syntax has to evolve, the evolution should be done in continuity. No open-source strategy can succeed without institutional continuity and support. Young, brilliant software developers can invent whatever they want, but some institutions must support them: universities, the state. Otherwise, everything is lost when the developers abandon their task. This is a fundamental political issue.

⁴ See: “LiMux,” Wikipedia, <https://en.wikipedia.org/wiki/LiMux>, (accessed 2 March 2026).

⁵ Fairphone, <https://shop.fairphone.com/fr>, (accessed 2 March 2026).

Actually, this is the support that universities such as Stanford have provided to computer and software companies since the beginning of the digital age.

MV: Your path over the years has led you to study Vitruvius, Dürer, Pascal, Leibniz, and Desargues. Why this attention to the past?

BC: Because I've heard people, even well-intentioned ones, like Paul Virilio, say that we live in a "non-Euclidean" world. These are violent slogans. All rationality comes from Euclid. Even non-Euclidean geometries are variants of Euclidean geometry. When Descartes introduced algebra, he was still thinking geometrically, imagining shapes when writing equations. There was continuity.⁶ People who cut the links with fundamental fields such as Euclidean geometry are just shrinking the past and endangering our present and future.

MV: And today, this historical awareness is missing.

BC: Yes. In many engineering schools there are no longer consistent courses in the history of mathematics. Yet the tradition continues. Andrea Del Centina, a great Italian historian of geometry, wrote a fundamental book: *From Here to Infinity: Tracing the Origin and Development of Projective Geometry*.⁷ It is my daily bible. I keep it beside me every day.

MV: So geometry, formal logic as a gateway to other forms of knowledge. Yet, today we can say many things about architecture—from parametric to artificial intelligence—but it is clear that the formal expression of design has lost centrality. Is that not so? And why?

BC: Yes, and rightly so. Because the problem—apart from the far right, immigration, and digital technology—is climate change. People do not realize the urgency. Here in Switzerland, we are already more than 2°C above the pre-industrial average temperature. Let us say we have thirty years left, after which it is the end. Some discussions are truly urgent. For instance, whatever happens in Iran nowadays, this country will remain with summers reaching 50°C, and two thirds of the population lacking water. And guess who is responsible for that?

⁶ H. J. M. Bos, *Redefining Geometrical Exactness: Descartes' Transformation of the Early Modern Concept of Construction*, Springer, New York, 2001.

⁷ A. Del Centina, A. Gimigliano, *From Here to Infinity: Tracing the Origin and Development of Projective Geometry*, Springer, Cham, 2025.

We are restituting, in the guise of heath, the oil that the British had stolen to the Iranians. This oil, reappropriated by Mosaddegh in 1951, was stolen a second time in 1953 by the americans with the complacency of the Shah. And now they are considering his son as a convenient solution to the problems they created.

This is why the architectural problem is no longer fancy design: the issue is data management and sustainable processes of production, construction, maintenance and, of course, habitability. It is so complex that perhaps it could be a challenge for AI.

So, my research is not anymore focused on formal expression, as it was naively in the 1990s. Instead, I am trying to reconnect links that were never articulated.

MV: Yet, you insist on studying the axioms and constructions of projective geometry, a field in which the formal logic of mathematics encodes epistemological and philosophical assumptions. You have remarked that, since the mid-eighteenth century—since Kant, “the first philosopher who had nothing to say about mathematics”—a “harmful legacy has persisted, separating philosophy from mathematics.” Is it this unity between the sciences and the humanities that, in some ways, you are trying to reconnect?

BC: Projective geometry is a good example of things in the past that did not happen. To be sure, it developed very well in the field of mathematics, but it did not fulfill its promise of delivering common methods to several practices such as: draughting, stonecutting, or solar clock construction. I will give you a very important example with the concept of the infinite. The common idea is that a line has 2 infinities: one at each of its ends. For a number of reasons, Desargues came up with the idea that a line has only one single point at infinity, which links its two ends. The image of this unique point at infinity is the vanishing point attached to the direction of the line. This uniqueness of the infinite point transforms the line into a cyclic element. It is a key notion of projective geometry, which remains to be fully implemented in CAD software, since it would imply being able to transform translation into rotation, and vice versa.

Nobody knows to what extent an ordinary user of that software would be ready to take that into account, considering that there has been at least one user of projective geometry who understood that perfectly, namely Pascal. But strangely enough, Pascal never tried to implement this uniqueness of the infinite in his philosophical writings.

The same happens with Leibniz, when he says that a monad is a “perspective” on the world. Leibniz stays with the common idea that each individual is a point of view on the universe, without taking advantage of the notion of involution, so central in Desargues’ geometry. My guess is that projective geometry is full of concepts that we can still take advantage of, and interestingly enough, this geometry emerged within the field of architecture.

It is strange and interesting because all this changes our idea of the classical as something boring, already organized, with nothing more to say. I am astonished by the contemporaneity one encounters in the folds of history. Just think of the fact that Desargues only wrote drafts of projects⁸ (*Brouillon de projet*)⁹ or examples.¹⁰ In his *Manière universelle*,¹¹ Abraham Bosse inserted some writings of Desargues, presented as “a notebook like a lost child.”¹²

MV: You seem to be searching, in the folds of history, for traces of the future. History is indeed crucial. It constitutes a fundamental question of our time: on the one hand, the uninterrupted flow of information

⁸ For English readers, a nice introduction to Desargues’ geometry would be: J. V. Field, J. J. Gray, *The Geometrical Work of Girard Desargues*, Springer-Verlag, New York, Berlin, Heidelberg, London, Paris & Tokyo, 1987.

⁹ In geometry, see G. Desargues, *Brouillon project d’une atteinte aux événements des rencontres du cône avec un plan*, Gallica / Bibliothèque nationale de France, <https://gallica.bnf.fr/ark:/12148/bpt6k105071b.image>, (accessed 2 March 2026). In stereotomy, see G. Desargues, *Brouillon Project d’Exemple d’une manière universelle du S. G. D. L. touchant la pratique du trait à preuves pour la coupe des pierres en l’Architecture: et de l’éclaircissement d’une manière de réduire au petit pied en Perspective comme en Géométral, & de tracer tous Quadrans plats d’heures égales au Soleil*, Paris, 1640. In gnomonics, see G. Desargues, *Brouillon Project du S.G.D.L. touchant une manière universelle de poser le style et tracer les lignes d’un Quadrant aux rayons du Soleil, en quelqu’onque endroit possible, avec la Règle, le Compas, l’équerre et le plomb*, 1640; for the recovered edition, see A. J. Turner, “Another Lost Work by Girard Desargues Recovered,” *Archives Internationales d’Histoire des Sciences*, 34, 1984, pp. 61–67.

¹⁰ In perspective, see G. Desargues, *L’exemple de l’une des manières universelles du Sieur Girard Desargues Lyonnais touchant la pratique de la perspective sans employer aucun tiers point, de distance ny d’autre nature, qui soit hors du champ de l’ouvrage*, Jacques Dugast, Paris, 1636.

¹¹ A. Bosse, *Manière universelle de Mr Desargues, pour pratiquer la perspective par petit-pied, comme le géométral. Ensemble les places et proportions des fortes et foibles touches, teintes ou couleurs*, Pierre Des-Hayes, Paris, 1648, p. 311.

¹² The beginning of the notebook inserted as a lost child: A. Bosse, *Manière universelle de Mr Desargues, pour pratiquer la perspective par petit-pied, comme le géométral. Ensemble les places et proportions des fortes et foibles touches, teintes ou couleurs*, Pierre Des-Hayes, Paris, 1648, p. 335.

encourages historical amnesia; on the other, probabilistic systems linked to big data assume that what happened and has been recorded will recur. Mario Carpo argues that Large Language Models, in fact, automate “creative imitation,” a process central to Western artistic experience, grounded in the idea of art as *mimesis* of reality.

BC: I understand what Mario means, but I would not put it that way. As you say, artificial intelligence is a matter of probability, of common opinion. It takes what exists, makes a probability calculation, and provides the most likely answers. It is what you do when you have no knowledge of a problem. Statistically, it may work many times.

However, for example, one of the greatest experts in the field, Luc Julia (French engineer who invented Apple Siri and is now Chief Technology Officer at Renault) clearly states that the fully autonomous car will very likely never exist. When you take a driverless taxi in San Francisco, there is actually someone in a control room who intervenes when a problem arises. So, this is not knowledge. It is not syntax: it is probability. The question of syntax is fundamental. It is not because we have a lot of information that we have knowledge. People are lost because this information is unstructured.

MV: Somehow, this is also Noam Chomsky’s position: he argues that AI is deficient because it processes language statistically rather than through an innate grammar that’s similar to that of humans. It lacks true understanding and critical thinking, remaining morally indifferent despite ethical safeguards.

BC: Yes, it is precisely a question of grammar, of syntax, and of lexicon. It is the same reason why, in projective geometry—especially in applying perspective—I realized that fundamental elements were not properly named. This is one of the reasons why this geometry never reached the public. *The more information you get, the more you need to organize it. Just remember that one disinformation technique is to submerge your enemy with irrelevant, unstructured information.*

Naming things is extremely important, and the fact that Desargues did not do so is worth discussing, because he was actually very attentive to language. For instance, he did not name the plane parallel to the picture plane passing by the observer. This observer plane intersects the object plane in a line made of points that have as much importance as

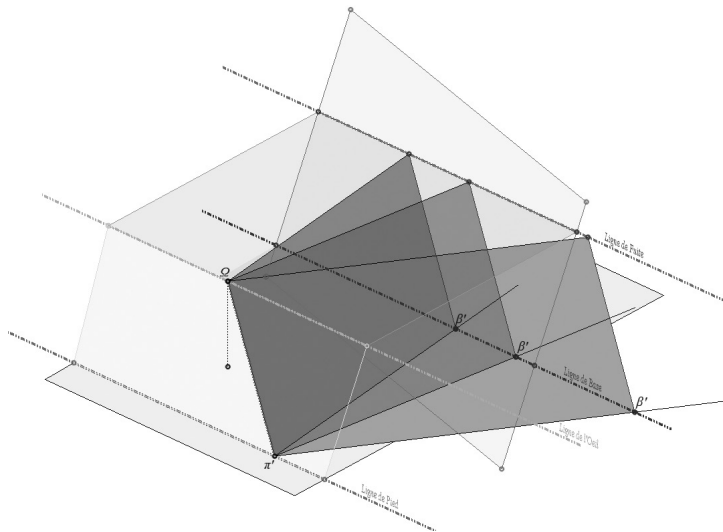


Figure 2. Foot point on the Foot line: the point of convergence of object lines having parallel images. Illustration by Bernard Cache.

the vanishing points on the vanishing line. But Desargues did not give them a special name and, as such, those points—which I would call “foot points”—remained unnoticed.

My guess is that there are at least three layers in Desargues’ thought, and that he only disseminated the first because he thought of the others as out of reach of the average draughtsman, stonemason, or builder of solar clocks.

For example, in his first writing on perspective, *L’exemple* (1636),¹³ he dedicates nine pages out of ten to explain the first layer, which amounts to building a coordinate system in the object plane that transforms into another system into the picture plane, together with a reduction scale. Then, in the last page, he writes: “since I have a bit of space left, I will say something for the contemplatives.” And in this single spare page, he explains very important things that amount to making the foot point an equivalent of the vanishing point. Desargues will never come back to this second layer although it provides constructive methods equivalent to those of Apollonius, but extended to general oblique cones rather than limited to circular ones. Actually, this second layer is not so complex, but Desargues was already engaged in the development of his third layer, based on the rather sophisticated concept of involution.¹⁴ Not only does this powerful concept require some deep mathematical understanding but it requires calculations that would repel most practitioners. As a result, Pascal would develop alternative methods based on the theorem which bears his name. Pascal then wrote a whole treatise on conics¹⁵ which, for a number of reasons, was never published. Hence, seventeenth-century projective geometry remains a kind of phantom theory

¹³ G. Desargues, “L’Exemple de l’une des manières universelles,” in *Œuvres de Desargues réunies et analysées*, Leiber, Paris, 1864.

¹⁴ For mathematicians, involution is the ancestor of the 19th century cross ratio. In perspective, all rectangles object * image have constant area, provided you measure them appropriately: a) the measure of the object is its distance to the foot point; b) the measure of the image is its distance to the vanishing point. Since all those rectangles are equal, their corner opposed to the observer’s eye lies on a hyperbola. Desargues starts his major *Brouillon Project ... des Rencontres d’un Cône avec un Plan* by involuting this planar hyperbola into a linear relation on the vanishing radius. The problem is that Desargues does not refer to perspective as his starting consideration. Not only he does not mention that the *stump* of his involution is actually the observer’s eye, but he quickly discards it from his rather complex relations of involution. Hence, his theory of involution is totally abstracted from the perspectival scene. See B. Cache, *The Perspective Block*, forthcoming.

¹⁵ Blaise Pascal, *Conicorum Opus Completum et Conica Apollonii et alia innumera unica fere propositione amplexens*, 1654, which is only known through the notes taken by Leibniz, in particular on the 1st chapter “Generatio conisectionum tangentium et secantium;

that was never fully formulated, although it engages fundamental notions such as representation. Ignoring the involution and the complex topology of the projective plane, art historians such as Erwin Panofsky¹⁶ can discuss at length upon perspective without entering into its intricacies.

Once again, it is a matter of syntax. In mathematics, there are phases of invention, but they must be followed by phases of formalization and organization. Syntax seems to me the most important point, without which we could not fight the far right.

The far right gains traction when there is no order. This is why Trump attacks universities: because universities are the places where knowledge is built, given order and meaning.

MV: I fear, however, that despite good intentions, these are topics that risk being somewhat impenetrable. In architecture, for example, computational design has ended up alienating many architects who, in reaction, retreated into the status quo or the tradition.

BC: We must not go backward. We must organize. And we must take autonomy. My argument is not about returning to candles.

MV: On the contrary, your position sounds more like a call to arms: that is, a call to shift the discussion to a political plane through the development of open-source software and platforms. However, it could be seen as an extension beyond disciplinary boundaries.

BC: I think open source is needed and it requires institutional support: a European digital autonomy that would collaborate with North Africa, where there are excellent engineers and intellectual resources. And then organization, and also having spaces where there is no more digital: setting limits, a sense of boundary that goes together with syntax.

Another project to work on is the digitalization of public administration: every citizen should have access to a basic digital terminal to perform essential state services (taxes, healthcare, banking, etc.). A service that only allows checking bank balance, doing postal services, and normal tasks of a modern citizen.

seu projectio peripheriae, tangentium et secantium circuli, in quibuscumque oculi, plani ac tabellae positionibus.”

¹⁶ E. Panofsky, *Perspective as Symbolic Form* [1924], trans. C. S. Wood, Zone Books, New York, 1991.

Then there is the issue of mobility. The electric car makes no sense for someone like me, who lives in a city with the metro and good public transport. But I was born in the countryside, and I know perfectly well that, when you are in the fields, you need a car for the last mile. I think of intelligent and public cars. They could be public vehicles, not privately owned, for which the state grants concessions fostering competition: these fleets of cars should operate permanently and should be locally controlled. This would be a leftist approach to electric cars, as opposed to the electric private SUV. What I mean is that there is room for action: digital technologies have become what they are today because boundaries were not set, and large American corporations were allowed to develop them out of control.

MV: For years, the focus was on experimenting with new possibilities offered by new numerically controlled production tools: CAD/CAM, file-to-production, parametric design, etc. These technologies afforded the possibility to design complex forms, and so complex forms were designed just because it was finally possible to do so. Today, however, it is urgent to understand what kind of society we want to build. What you call syntax, in other words, the problem of understanding the meaning of our technological world order.

BC: Today, we must “tame the beast.” I repeat: the fact that parents oppose a device that blocks children’s communications in schools is a very important symptom. Yet those same parents agree to surveillance that checks all students’ backpacks at the school entrance: they prefer actual police-style control (!) rather than a radical, clear, and simple solution: signal scramblers switched on and off at appropriate time. This is an architectural issue: defining limits, walls within which communication occurs only when it makes sense. We must think of a digital architecture with disconnected enclosures.

MV: Speaking of architecture, I recall a wonderful lecture you gave at the Architectural Association, in which, speaking of the Tower of the Winds by Andronikos of Cyrrhus in Athens, you discussed how, at the origins of Western architecture, we do not find a refuge, but a machine to measure (time, sun, wind). And again: your interest in Vitruvius lies in his work on the design of war machines. How does the concept of the *machine de guerre* inform your architectural thinking and your view of architecture

as a calculating tool? Do you still think of architecture as *machinatio*, of being a tool for measuring and computing reality?

BC: Today, in architecture, I am mainly focused on the software aspect. Beside projective geometry, which still has to be fully implemented within CAD software, I am supporting the work of two of my PhD students: Ahmed Waël and Raphael Vouilloz, who are developing a Speckle viewer enabling a better understanding of geometrical construction. Speckle is a startup that is developing an application allowing one not to be trapped in a single software but, on the contrary, to connect different software platforms. This is fundamental, and it is what underlies open source.

As for the city, I think the fundamental problem is climate change. I remember a friend in Rome, five or six years ago, telling me that it was already 30°C at the end of May. Now in France, we are almost there. The speed of this phenomenon is incredible.

So yes, definitely, we should build with timber, with CNC machines, etc. But I have never seen a precise analysis of the climatic implications of these systems.

The book I mentioned earlier, *Sans transition*, shows how energies are interdependent and that you cannot easily switch from one system to another: when coal arrived, people used less wood for heating, but to go into the mines, wooden beams were needed, so wood consumption actually increased. Therefore, without wood you cannot have coal, and without coal, you cannot have steel or cement, etc. Beyond the slogans of some politicians, implementing an “ecological transition” is much more difficult than people think. Perhaps artificial intelligence can help, because the problem is very complex.

MV: So do you really think it is only a matter of software? What about hardware? What role does built architecture play?

BC: In the end, we still need to produce wooden houses. I am still involved in this project through collaboration with TopSolid: not yet at the scale of the entire house, but in furniture and soon in carpentry. Prefabrication and the organization of production are making progress. It is important not to abandon them. Just as we need oil to build solar panels. But understanding whether we are moving toward good or bad outcomes is very difficult.

MV: So we need a new mathematical, technological humanism.

BC: Politically, yes.

MV: Forgive me for the digression, but I think it somehow relates to what we are discussing: on one hand, technology frightens us; on the other, it promises salvation, omniscience, total computation—there is, in short, a mystical aura surrounding the salvific power of technology. Even Pascal, in the *mystical hexagram* linked to the study of conics, glimpsed the possibility of a transcendent meaning.

BC: Hm, what was called the mystical hexagram is actually the unique theorem¹⁷ upon which Pascal built the 400 propositions of his full treatise on conics, precisely in order to bypass the complexity of Desargues' involution. Yet I am very suspicious of everything called *mystical*. Religion is certainly what prevented Pascal to take full advantage of the unicity of the infinite within his philosophical writings although it could have been a nice theological concept.

But yes, today it is up to us to save ourselves. We have to develop technology and guide it. If we use it without political control—which is what is happening now—it will be our catastrophe.

We must find a way to organize technology by engaging in the solution of practical problems. For example: how to organize digital communication in schools?

When I say we must prevent any communication via cell phones, it does not mean there should be no computers to learn coding or to organize one's digital space, and even to teach smart uses of cell phones. Let us be clear: my position is not anti-technology.

You began by saying we could not have imagined the situation we are in, and that is true. But we should take the syntax we had in a highly organized professional field and bring it to the political field of common digital life.

MV: It is very clear: a plea for a new digital sovereignty.

¹⁷ Pascal's theorem has it that any hexagon inscribed into a conic has the three intersection points of its three pairs of opposed sides aligned on a straight line. In the particular case where the hexagon becomes regular, those three intersection points are the unique point at infinity of the three directions of its sides, lying on the unique line at infinity of the plane.

BC: Yes. I think that, in a few years, between Trump, Putin, Xi Jinping, and the climate crisis, if we do nothing, our human species will just disappear.

MV: In architecture, there is this tendency toward technocratic “solutionism.” The risk is developing research that is unguided and generic. Take the last Venice Biennale: there are good projects, animated by good intentions, but too often without a political or humanistic vision.

BC: And above all, without planning. We need to return to planning.

MV: Planning is another major theme. As architects, we often talk about the city, but we have stopped taking care of it. At most, we describe it, but too often we are strangers to the dynamics that determine its form and function.

BC: Yes, but for example, the Paris metro works well and it was planned. For me, the Paris metro is a paradise: people of all colors, all social levels, compressed in a closed space, coexist well. It gives me hope every time I take it.

MV: I would say that is an excellent conclusion. Thank you.

Interview conducted by Marco Vanucci.

Roberto Bottazzi*

THEODORA VARDOULI, *GRAPH VISION: DIGITAL ARCHITECTURE'S SKELETONS*, THE MIT PRESS, CAMBRIDGE, MASS., 2024.

Graph Vision provides a historical overview of the introduction of graphs in architecture and urban design. Schematically, graphs are mathematical tools that pertain to the branch of finite mathematics. Through a simple language of nodes and edges, graphs can chart out and count the relations between objects. They do not describe the visual properties of objects, rather how they relate and structure. As abstract mathematical tools, their application to architecture does not concern the actual building, but rather its representation; in other words, graphs can only be used in the design process. Applied to plans, graphs could represent rooms as nodes and connections between spaces as edges. The implications of this simple move were however profound: rather than representing the geometrical qualities of buildings, graphs captured their structural relations such as the circulation layout and the overall arrangement of rooms. They attended to the desire to apply mathematical methods to architecture and, due to the potentially large amount of calculations involved in manipulating plans through graphs, mainframe computers became the ideal instruments to support such project. As Vardouli reminds us, graphs provided the foundations for what would eventually become known as digital architecture.

A large part of the book is dedicated to the pioneers of this line of research: Lionel March, Philip Steadman, and Philip Tabor at the Centre for Land Use and Built Form Studies (LUBFS) at Cambridge University, Yona Friedman, operating between France and the USA, Christopher Alexander, Nicholas Negroponte, and William Mitchell. The survey operated by Vardouli concludes at the end of the 1970s and therefore does not assess the work of Bill Hillier and his UCL-based group Space Syntax.

* Roberto Bottazzi: The Bartlett School of Architecture, University College London; roberto.bottazzi@ucl.ac.uk.

Around these central figures, Vardouli reconstructs a carefully and lesser known network of collaborators, artists, philosophers, and mathematicians that contributed in different ways to the development of a graph-based architecture. The vast majority of the research on graphs took place in the Anglo-Saxon world, with some exceptions such as the work carried out by Yona Friedman in France.

The book, an elegant and meticulously designed object, is structured into three main parts, plus a conclusion: “Images,” “Tools,” “Infrastructure,” and “Skeletons.” The first section traces the attempts—mainly by Lionel March and Philip Steadman—to introduce graphs to develop a scientific approach to design. Vardouli unfolds the different roles played by graphs in this context. They were mobilised to do away with, or at least diminish, the importance of form as well as the visual and geometrical aspects of design, in favour of structural and organisational characters (Lionel March and Ray Matela maintained that graphs were immune to geometrical considerations). Regardless of the initial ambitions, graphs themselves were visual artefacts whose abstract aesthetic was, and still is, very enticing. Minimal representations of plans as graphs or large catalogues of combinations of rooms also fascinated architects. “Tools” surveys the many attempts to implement graphs in architectural design to analyse and automatically generate plans. The discussion uncovers the issues emerging from striking a balance between the whole set of parameters informing a particular layout and the limitations provided by graphs. Finally, “Infrastructure” describes how graphs enhanced participation by either involving end users or providing an exhaustive catalogue of all possible plans based on a single graph for users to select from.

To grasp the novelty ushered by the introduction of graphs in design, it suffices to point out that only a couple of years earlier Le Corbusier’s major works on mathematics and architecture—*Modulor* (1948) and *Modulor 2* (1955)—were still based on geometrical proportions which he had broadly inherited from the renaissance. Tellingly, the symposium held at the Triennale in Milan in 1951 to discuss and validate Le Corbusier’s *Modulor* was titled *La Divina Proporzione* [The Divine Proportion], a direct homage to Luca Pacioli.¹

Through the figure of the graph, Vardouli unfolds a rich aesthetic, social, technological, and political debate that animated the architectural discourse from the late 1950s until the 1980s. Some of the factors that

¹ A. C. Cimoli, F. Irace (eds.), *La divina proporzione: Triennale 1951*, Electa, Milano, 2007.

drew architects to graphs were: the industrialisation of the construction industry prompted by the need for post-war mass housing, the importance of a socially-oriented architecture that paid less attention to form and more to program and organisation, and the possibilities provided by computers to manage complex tasks to develop a science of design. A particular political agenda that graphs elicited was that of participation: in experiments such as Friedman's *FLATWRITER* or Alexander's *Pattern Language*, graphs systematised the production of layouts—mostly for residential buildings—or urban spaces potentially bypassing the need to involve architects in the design process. Participation was perhaps an aspect of an even larger political agenda for design: as Vardouli notices, “graphs structured choice”² and therefore became vehicles for emancipation.

Though Vardouli's book has a clear historical focus, as for any historical survey, it is unavoidable to draw parallels between the ambitions of the experiments on graphs and the more recent or contemporary state of the discourse of digital architecture. By expanding the timeframe considered, graphs appear to occupy a peculiar position. Mario Carpo characterised the periodisation of digital architecture through two significant turns: the first one which took place in the early 1990s when powerful and cheaper personal computers came to fruition allowing architects to model complex geometries, the second which emerged some twenty years later when large datasets, i.e. collections of numbers, became available for processing. If the first digital turn largely concentrated on geometry, the second one returned to mathematics and code to analyse massive datasets. According to this scheme, graphs would belong to a pre-digital era, however, the conceptual problems they raised were closer to that of the second digital turn: logic, mathematics, structuring data not only pertain to recent debate on digital design, but they also confront, again, the tension between the non-visual aspects mathematics and the retinal qualities of architectural representation and design. *Graph Vision* offers a detailed overview of a set of concerns that still animates the contemporary discourse: these are issues related to counting, searching, systematising large sets of objects, as well as the primacy of deductive over inductive reasoning methods to define what intelligence might mean in design (which the book brilliantly captures in the analysis of the differences between

² T. Vardouli, *Graph Vision*, p. 180.

“specification” and “recognition” in the work of, respectively, Friedman and Negroponte-Groisser).

The pre-digital work on graphs also offers other elements of reflections that remind us that some issues in computational design persist regardless of technological change. One such theme is the role of the architect *vis-à-vis* the introduction of computers. Vardouli recalls how Friedman and later Negroponte designed different modes of interaction between architects and machines. Whereas Yona Friedman tasked computers to map out all the possibilities from which end users would choose, Negroponte’s YONA—a computer program inspired by Friedman’s work—allowed users to directly see the graphs, presenting them as “a malleable, visual entity, accessible well outside the cycles of a mathematical elite.”³ Contrary to the recent abundance of processing power, the research on graphs was characterised by technological scarcity. The limited computational capacities of mainframes computers of the 1960s and 1970s demanded a relentless, almost heroic, effort to simplify calculations to avoid so-called combinatorial explosion (the exponential growth of combinations caused by a linear increase of parameters). Graphs offers a highly economic way to represent architecture for a computer; they could recode an architectural plan though a limited set of numbers, disregarding geometrical information. They also framed architectural intelligence as largely a deductive problem. It is interesting to observe the contrast with the paradigm guiding the current generation of Artificial Intelligence which is trained on large datasets and partially grounds its intelligence in inductive operations.

Finally, in the midst of the integration of large language models (LLMs) in architecture, Vardouli reminds us that digital architecture had already flirted with language before. From George Stiny’s *Shape Grammar*—influenced by Chomsky’s generative grammar—to diagrammatic use of graphs by C. S. Peirce, proto-graph methods to analyse language had already been devised. Their abstract qualities responded to the crisis of empirical methods by substituting them with logical methods. The domain of references mobilised to chart the relation between language, architecture, and computation reinforces the impression that graphs were largely an Anglo-Saxon affair. The vast majority of the experiments and discussions took place either in the United States or England. The work of linguists from the Slavic regions seemed to have had little to no

³ T. Vardouli, *Graph Vision*, p. 182.

impact on the formalisation of architecture through graphs. As a result, the pool of references to draw from seemed to be determined earlier on and was never really questioned (perhaps, with the exception of Stiny's *Shape Grammar*). Perhaps as a consequence of this, the perimeter of the architectural issues and concerns attended through graphs also remained unchanged throughout some twenty years of research (plans constituted the main focus of the research).

Theodora Vardouli's *Graph Vision* offers a compelling, painstakingly detailed, and elegant overview of a period in architectural research which, at first glance, would appear far removed from the technological condition we inhabit. Through a closer inspection, however, we observe not only the persistence of recurrent themes—mathematics, philosophy, design's political agency—but also the importance of having a *vision*: graphs were the instruments and the instigators of a larger intellectual project that questioned architecture's role in society. A reminder that digital architecture was never just about technology.

Alberto Fernández González*

DANIEL KOEHLER, *COMPOSITIONAL INTELLIGENCE: ARCHITECTURAL TYPOLOGY THROUGH GENERATIVE AI*, ROUTLEDGE, ABINGDON / NEW YORK, 2026.

The study of automata has long occupied a central position in discussions of automation and technical systems. From early mechanical automata to modern computational models, these systems explore how complex behaviour can emerge from structured rule-based processes. A particularly influential contribution to this field comes from Stephen Wolfram, whose work on cellular automata demonstrated that simple computational rules can generate highly complex and unpredictable patterns. Wolfram argues that even systems governed by deterministic rules may produce behaviours that appear irreducibly complex and difficult to predict.¹

This insight fundamentally reshaped the understanding of automated systems. Rather than viewing automata as rigid machines executing predictable instructions, Wolfram's research revealed that deterministic systems themselves can give rise to emergent phenomena. The implications of this discovery extend beyond computer science into fields such as biology, physics, and architecture, where rule-based systems have been used to model growth, adaptation, and spatial organisation.

Architectural interest in automata emerged particularly in computational design research, where cellular automata and generative algorithms were explored as mechanisms capable of producing complex spatial structures from simple local interactions.² These approaches suggested that architecture could be understood not only as a designed object but also

¹ S. Wolfram, *Cellular Automata and Complexity*, Addison-Wesley, Reading, Mass., 1994.

² W. J. Mitchell, *The Logic of Architecture: Design, Computation, and Cognition*, MIT Press, Cambridge, Mass., 1990; J. Frazer, *An Evolutionary Architecture*, Architectural Association, London, 1995.

* Alberto Fernández González: The Bartlett School of Architecture, University College London; Department of Architecture, University of Chile, Santiago; alberto.fernandez.11@ucl.ac.uk.

as the emergent result of dynamic systems. However, the emergence of generative artificial intelligence introduces a new form of automated systems that differs significantly from the deterministic automata explored by Wolfram. Machine learning models do not operate through explicit rules defined in advance; instead, they derive patterns from large datasets through statistical learning processes. In this sense, contemporary generative systems represent a shift from deterministic automata toward probabilistic autonomous systems.

Daniel Koehler's *Compositional Intelligence: Architectural Typology Through Generative AI* enters precisely at this moment of transition. While the book does not explicitly position itself within automata theory, it can be read as an investigation into how architecture operates within a new generation of automated systems shaped by machine learning and generative AI. The central question therefore becomes: does generative AI represent a continuation of the automata tradition in architecture, or does it mark the emergence of a fundamentally different form of automated intelligence?

AUTOMATA AND THE ORIGINS OF COMPUTATIONAL INTELLIGENCE

The intellectual foundations of automata theory are closely linked to early developments in computer science. In the 1930s, Alan Turing introduced the concept of the universal machine, demonstrating that a single computational system could simulate any other machine through symbolic manipulation.³ Turing's work established the theoretical basis for programmable computation and provided the first formal description of automated reasoning.

Shortly afterward, John von Neumann extended these ideas by developing the concept of self-reproducing automata. Von Neumann explored how machines could replicate themselves through formal rule systems, introducing models in which complex structures emerge from interactions between simple components.⁴ His work laid the groundwork for cellular automata, systems in which discrete elements evolve through local rules applied iteratively across a grid.

³ A. Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society*, 42, 1936, pp. 230–265.

⁴ J. von Neumann, *Theory of Self-Reproducing Automata*, University of Illinois Press, Urbana, 1966.

Automata theory therefore emerged as a framework for understanding how complex behaviours could arise from simple rule-based mechanisms. In these systems, global patterns result from local interactions governed by deterministic procedures. A significant expansion of this idea occurred in the late twentieth century through the work of Stephen Wolfram. Wolfram demonstrated that even extremely simple cellular automata rules could produce behaviours of remarkable complexity and unpredictability. Some cellular automata generate repetitive patterns, while others evolve into chaotic or seemingly random structures. Importantly, these behaviours arise despite the underlying rules remaining entirely deterministic.⁵

Wolfram's research revealed a paradox central to automata theory: deterministic systems can produce outcomes that appear unpredictable. Emergence, in this sense, does not require randomness but can arise from simple recursive interactions. For architecture, these ideas opened new possibilities. Designers began exploring rule-based computational systems as tools for generating spatial structures, allowing form to emerge from dynamic processes rather than predetermined geometries.

ARCHITECTURAL GENERATIVE SYSTEMS AS AN AUTOMATED ENVIRONMENT

The influence of automata theory on architecture has been visible in various strands of computational design. Cellular automata, shape grammars, and parametric systems have all been used to generate architectural configurations through algorithmic procedures.⁶ These approaches share a common principle: architectural form emerges from explicit rule sets. Designers specify the rules governing interactions between elements, and the system iteratively produces spatial structures according to those rules. Within this framework, architectural automation remains fundamentally deterministic. Even when the resulting patterns appear complex or unpredictable, the generative process can be traced back to the underlying rule structure.

Generative artificial intelligence, however, introduces a different model of computation. Instead of operating through predefined rules, machine learning systems derive patterns from data through statistical training processes. In this direction, diffusion models generate images by

⁵ S. Wolfram, *A New Kind of Science*, Wolfram Media, Champaign, 2002.

⁶ K. Terzidis, *Algorithmic Architecture*, Routledge, New York, 2006.

progressively refining noise according to probability distributions learned from large datasets. Large language models generate text by predicting likely sequences of tokens based on prior training. This shift marks a transition from deterministic rule-based automation to probabilistic generative systems.

Koehler begins his book by situating architecture within what he describes as the digital city. Contemporary urban environments are increasingly shaped by algorithmic infrastructures such as financial analytics, environmental monitoring systems, and logistical optimisation platforms. These computational systems operate continuously, influencing how buildings are designed, evaluated, and constructed. Rather than functioning as discrete machines, these automated processes form distributed networks embedded within urban space. Real estate algorithms determine property values, energy optimisation systems regulate building performance, and data-driven planning models influence urban development patterns.

Within this context, architecture appears as a component within a larger automated environment, where the traditional role of the architect as a singular author of form becomes less clear when design decisions are shaped by algorithmic evaluations and computational constraints. Koehler's analysis therefore frames the contemporary city as a kind of distributed automaton, in which spatial outcomes emerge from interconnected technical systems rather than centralised design control.

GENERATIVE SYNTHETIC AND THE DISCRETISATION OF ARCHITECTURE

The emergence of generative AI introduces a new layer to this automated environment. According to Koehler, machine learning models possess the ability to synthesise architectural imagery and spatial patterns from large datasets. These models encode what he describes as synthetic knowledge, derived from the aggregation of collective visual and spatial information. When prompted to generate architectural scenes, diffusion models produce images that resemble plausible urban environments despite not reproducing specific precedents. The resulting compositions often incorporate landscape elements such as vegetation or environmental context even when these features are not explicitly requested.

Koehler interprets this phenomenon as evidence that generative systems internalise latent cultural patterns present in their training data,

being in that sense the architectural knowledge is not programmed into the system through explicit rules but emerges from statistical relationships within datasets. This process differs fundamentally from the rule-based logic of classical automata, instead of executing predefined instructions, generative AI navigates a high-dimensional space of learned correlations.

A central methodological concept in the book is tokenisation (borrowed from natural language processing), a key to understanding how buildings can be represented as data and how this enables new forms of computational analysis and design. Language models represent text as sequences of tokens, discrete units that serve as the basic elements of computation, in that sense Koehler proposes that architecture can be interpreted in a similar way, in where buildings and urban environments can be decomposed into informational components that function as tokens within generative systems (these components may include spatial relationships, architectural elements, or environmental features).

Tokenisation introduces a form of architectural discretisation reminiscent of cellular automata for urban simulation models⁷ explaining how cities can be modelled as complex adaptive systems, where urban patterns emerge from local interactions between spatial units such as parcels, roads, or agents. In both cases (Batty and Koehler), complex structures emerge from interactions between smaller units. However, the mechanisms governing these interactions differ significantly. In cellular automata, the state of each cell evolves according to explicit transition rules applied locally. In generative AI systems, tokens interact through statistical correlations learned during training. The relationships between elements are therefore implicit rather than explicitly defined.

SUPERPOSITION, UNPREDICTABILITY AND CREATIVE PROBABILISTIC SYSTEMS

Another important concept introduced by Koehler is superposition, derived from observations about neural networks. In these systems, multiple functions can coexist within overlapping parameter spaces. A single neural network may encode numerous patterns simultaneously without separating them into discrete modules. Koehler suggests that architecture

⁷ M. Batty, *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*, MIT Press, Cambridge, Mass., 2005.

could adopt a similar logic. Instead of organising buildings through clear hierarchical divisions between structural, environmental, and programmatic systems, architectural elements could participate in multiple functions simultaneously. Superposition challenges the modular logic that has historically structured architectural design. Rather than dividing systems into separate layers, generative approaches encourage overlapping relationships between different spatial and functional components. From the perspective of automata theory, this shift reflects a movement from sequential rule execution toward parallel distributed computation.

A key theme explored by Koehler concerns the role of unpredictability in generative systems. Classical automata are deterministic: given the same initial conditions, they produce identical outcomes. Although complex patterns may emerge, the process remains fully reproducible. Generative AI systems behave differently. Because they rely on probabilistic sampling within latent spaces, their outputs can vary even when prompted with identical inputs. Each generation represents a new exploration of the underlying probability distribution. This unpredictability introduces a new relationship between human designers and automated systems. Rather than specifying every aspect of the design process, architects interact with generative models through prompts and iterative refinement. The machine produces variations that the designer evaluates and curates. In this sense, generative AI does not eliminate human agency but transforms it. The architect becomes less a controller of deterministic processes and more a collaborator navigating a space of possibilities generated by autonomous systems.

AUTOMATA RECONSIDERED/REVISITED/REINVENTED IN THE AGE OF AI

Viewed through the lens of automata theory, Koehler's concept of compositional intelligence represents both a continuation and a departure from earlier computational paradigms. The work of Turing, von Neumann, Conway, Langton, and Wolfram established a tradition of rule-based computational systems in which complex behaviour emerges from simple deterministic interactions. These models emphasised transparency: the mechanisms governing system behaviour could be fully described through formal rules. Generative AI introduces a different form of automation characterised by statistical inference and large-scale pattern synthesis. The resulting systems operate as probabilistic environments

rather than deterministic machines. Koehler's framework therefore suggests the emergence of a new class of automated systems, in which generative behaviour arises from learned correlations rather than explicit rules.

Daniel Koehler's *Compositional Intelligence* provides a thoughtful exploration of how generative AI reshapes architectural design and theory. While the book does not explicitly situate itself within automata theory, its arguments intersect strongly with longstanding debates about automation, emergence, and human agency. By tracing a trajectory from deterministic rule-based systems toward probabilistic generative models, Koehler highlights a fundamental transformation in architectural computation. Contemporary generative AI systems operate less like classical automata executing predefined instructions and more like complex probabilistic environments that produce emergent spatial possibilities.

This shift raises important questions about authorship, control, and the role of unpredictability in design. The architect is no longer the sole generator of form but becomes a participant within a broader ecosystem of computational processes. In this sense, *Compositional Intelligence* can be understood as an exploration of architecture within a new generation of automated systems. Rather than replacing the tradition of automata, generative AI extends it into a domain where emergence arises not only from rules but also from data. The future of architectural automation may ultimately lie in the interaction between these two paradigms: the deterministic structures explored by classical automata theory and the probabilistic generative systems emerging from contemporary artificial intelligence.

CIP - Каталогизacija y publikaciji
Народна библиотека Србије, Београд

1+72

KHÖREIN / : journal for Architecture and Philosophy / editors in chief
Petar Bojanić, Snežana Vesnić. - Vol. 1, no. 1 (2023)- . - Belgrade : Institute for
Philosophy and Social Theory, 2023- (Belgrade : Alta Nova). - 23 cm

Polugodišnje. - Glavni stv. nasl. na grčkom jeziku.
ISSN 2956-1892 = Khörein
COBISS.SR-ID 115067401