From AI rhetoric to critical digital practice: overlooking the conundrum

Daniele Villa

Politecnico di Milano, Dipartimento di Architettura e Studi Urbani
(daniele.villa@polimi.it)

«Everything became content, then content became nothing.»
Text of a cartoon by Bruce Eric Kaplan, *The New Yorker*, 2025

The first and most comprehensive study commissioned by the Royal Institute of British Architects (RIBA, 2024), examining the penetration and impact of Artificial Intelligence technologies within the professional fields related to architectural and urban design, dates back slightly over a year. This research offers an in-depth analysis, successfully identifying and addressing significant issues despite operating within a context dominated by accelerated technological development. Such rushing is partly genuine, driven by considerable investment concentration, and partly artificial, inflated by a collective narrative that is frequently opaque and oversimplifying.

One year after its initial release, the raw data collected within the British context are unsurprising and clearly indicate an exponential growth in the impact exerted by the extensive array of technologies broadly categorized under the umbrella term *Artificial Intelligence*. These technologies are increasingly reshaping professions associated with architectural and urban design and related disciplinary knowledge.

The British institute surveyed over 500 members, arriving at an initial straightforward conclusion backed by reliable data: at least 41% of British architects regularly use artificial intelligence tools in their everyday professional practice. Within this group, 51% report a basic understanding of these new technologies, 32% claim practical competence, and only 6% declare advanced proficiency.

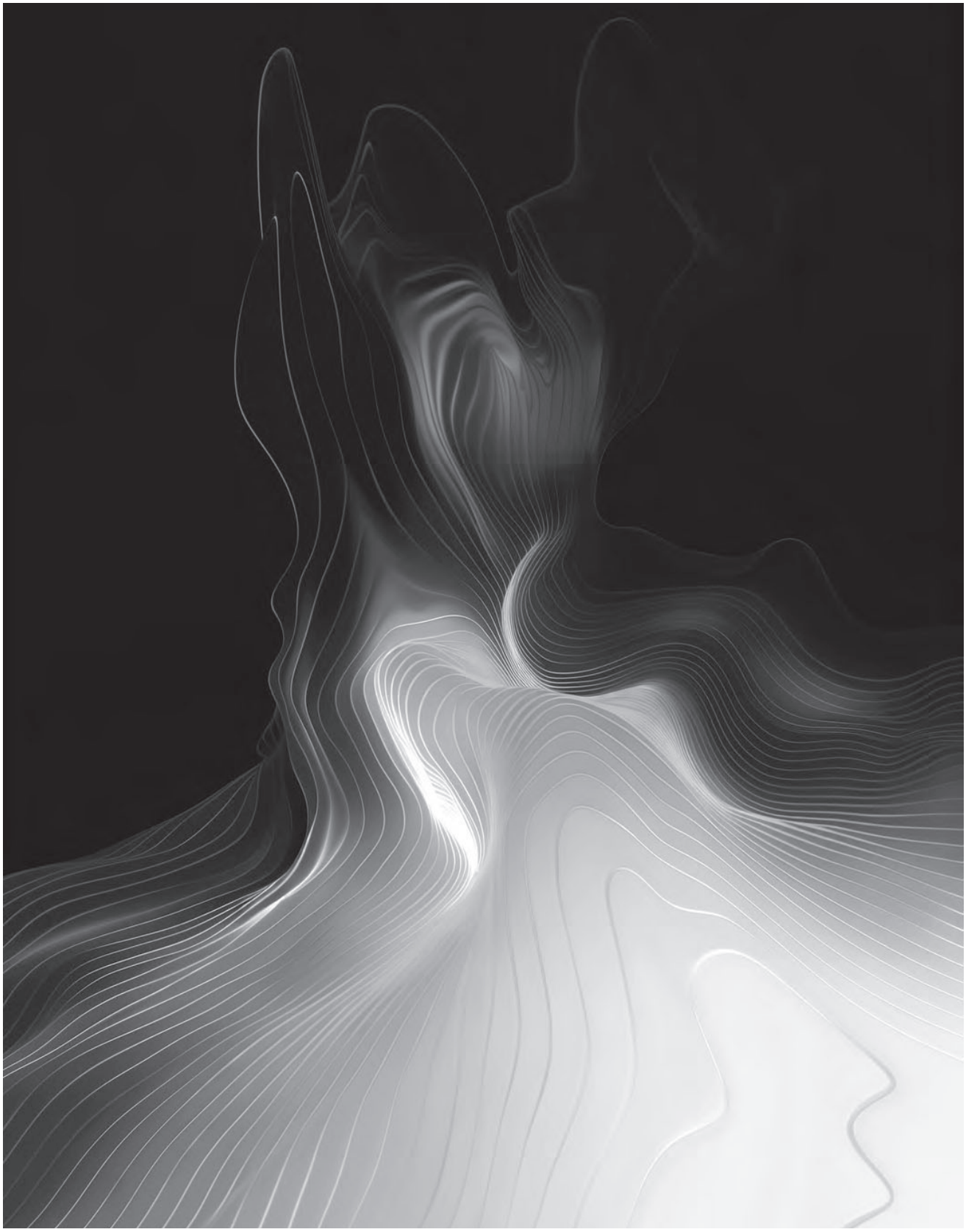
Nearly two years after introducing general-purpose AI-based tools – from machine-learning chatbots to image-generating diffusion algorithms, – we have observed another substantial transformation in the ongoing digital transition. This transformation impacts professional practice, research, and education within architecture and the built environment. As Adrian Malleson, Head of Economic Research and Analysis at RIBA, aptly points out: «The findings suggest that, while a significant number of practices have started to use AI in at least some of their projects (41%), AI adoption in the profession is in its infancy, with many practices not using AI or using it only occasionally. However, the adoption and use of AI are set to increase in the coming years. AI promises a beneficial and rapid evolution of the architectural profession; routine tasks automated, the design process streamlined, carbon reduced, creativity accelerated and expanded, and client and societal outcomes improved» (RIBA, 2024: 12).

Another relevant insight, though broader in nature, is provided by the Higher Education Policy Institute (HEPI). Also based in the United Kingdom, HEPI surveyed 1,041 full-time undergraduate students concerning their use of generative artificial intelligence (GenAI) tools. The extensive scenario outlined by this analysis appears unequivocal: «In 2025, we find that the student use of AI has surged in the last year, with almost all students (92%) now using AI in some form, up from 66% in 2024, and some 88% having used GenAI for assessments, up from 53% in 2024. The main uses of GenAI are explaining concepts, summarising articles and suggesting research ideas, but a significant number of students – 18% – have included AI-generated text directly in their work» (Freeman, 2025: 3).

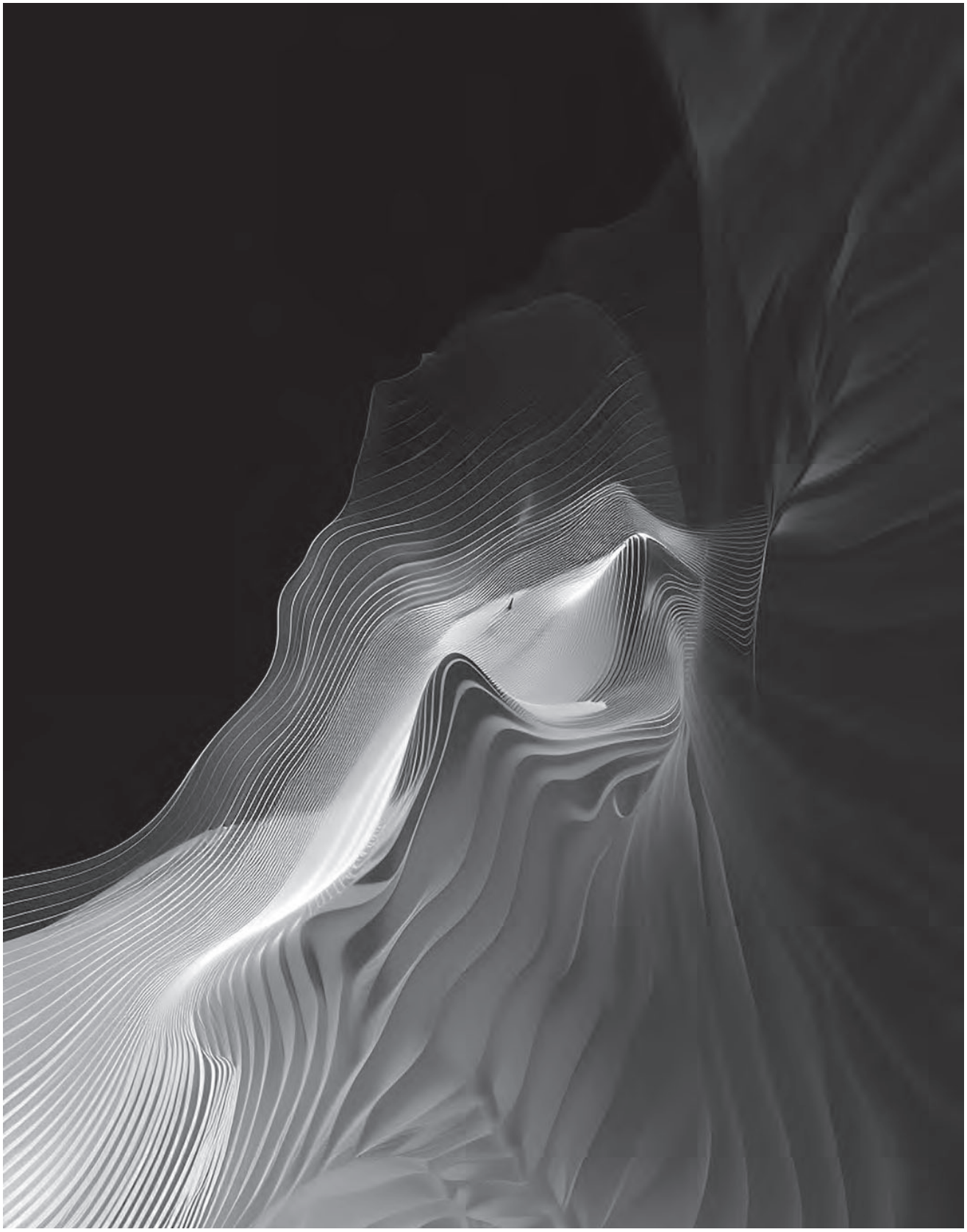
It appears worthwhile to underline some of the less apparent findings from this survey: half of the respondents indicated that they use AI tools to summarize relevant articles, 42% utilize them to structure their thoughts, and as many as 68% consider it essential, within today's professional landscape, to develop new skills for effectively understanding and employing artificial intelligence.

One of the main risks connected to a daily reality in which artificial intelligence appears ubiquitous and encompassing, yet increasingly difficult to characterize in specific terms, is the emergence of an overly emphatic, pervasive, and often uncritical collective narrative. This type of discourse hampers our ability to precisely delineate the unique boundaries of this profound technological, cultural, and epistemological paradigm shift: indeed, we are continuously confronted with several techno-enthusiastic, generalized, or even misleading narratives, which sometimes obscure and create confusion, a phenomenon akin to *AI-washing*.

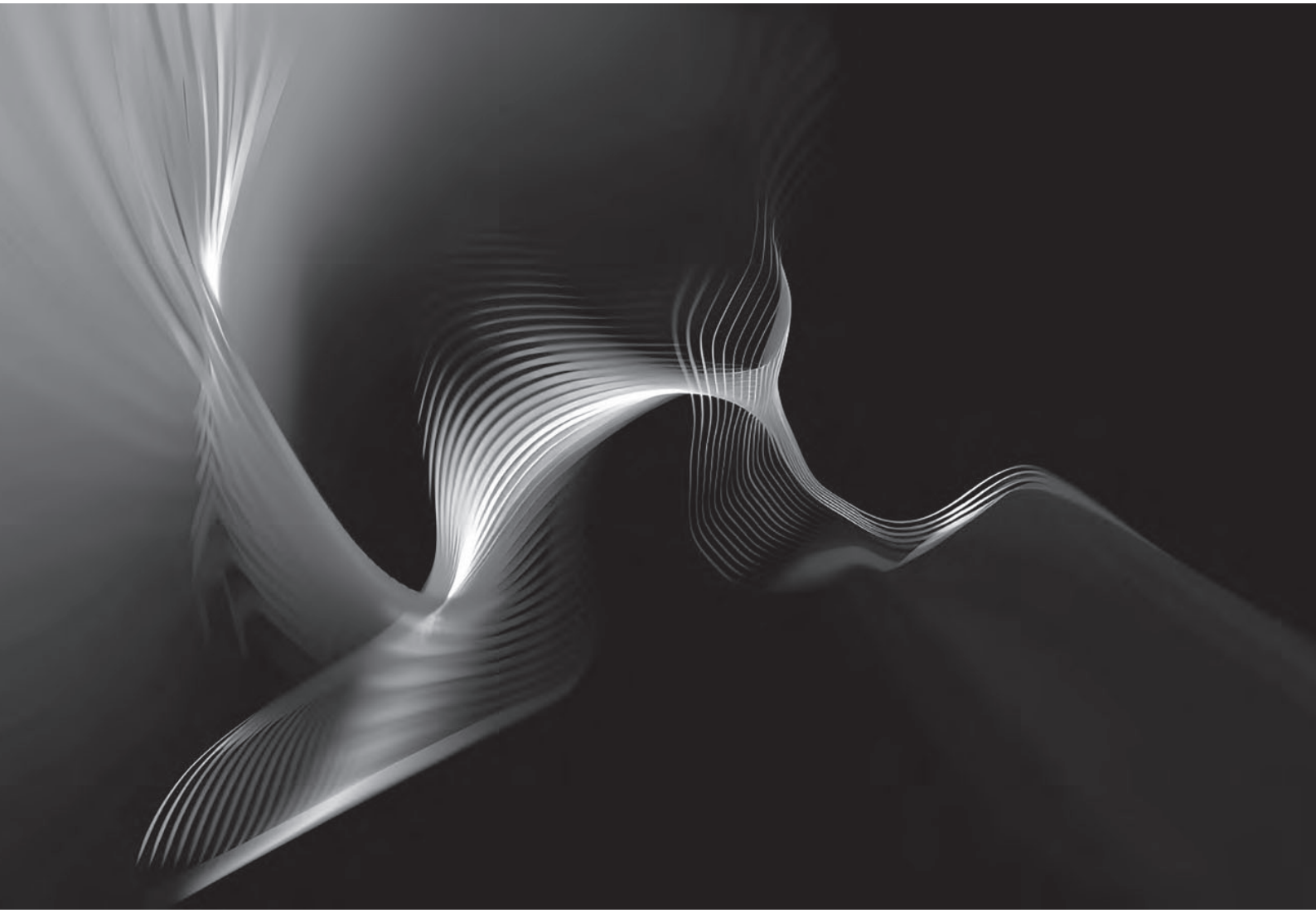
An emergent trend suggests a near future in which none of our cognitive or communicative actions will remain entirely independent from pervasive, automated systems designed to produce texts, images, content, scenarios, and, in the specific context of this discussion, a wide array of undefined spatial interpretations. Instead, it is imperative to clearly define the epistemological, educational, technical, and professional boundaries of this latest major revolution. This requires an accurate assessment of the innovative possibilities and strengths, along with a careful evaluation of the associated risks arising from the diverse characteristics that connect machine learning, the widespread use of large language models (LLMs), and the extensive array of multimodal applications, such as diffusion models, through which billions of synthetic contents are produced daily.



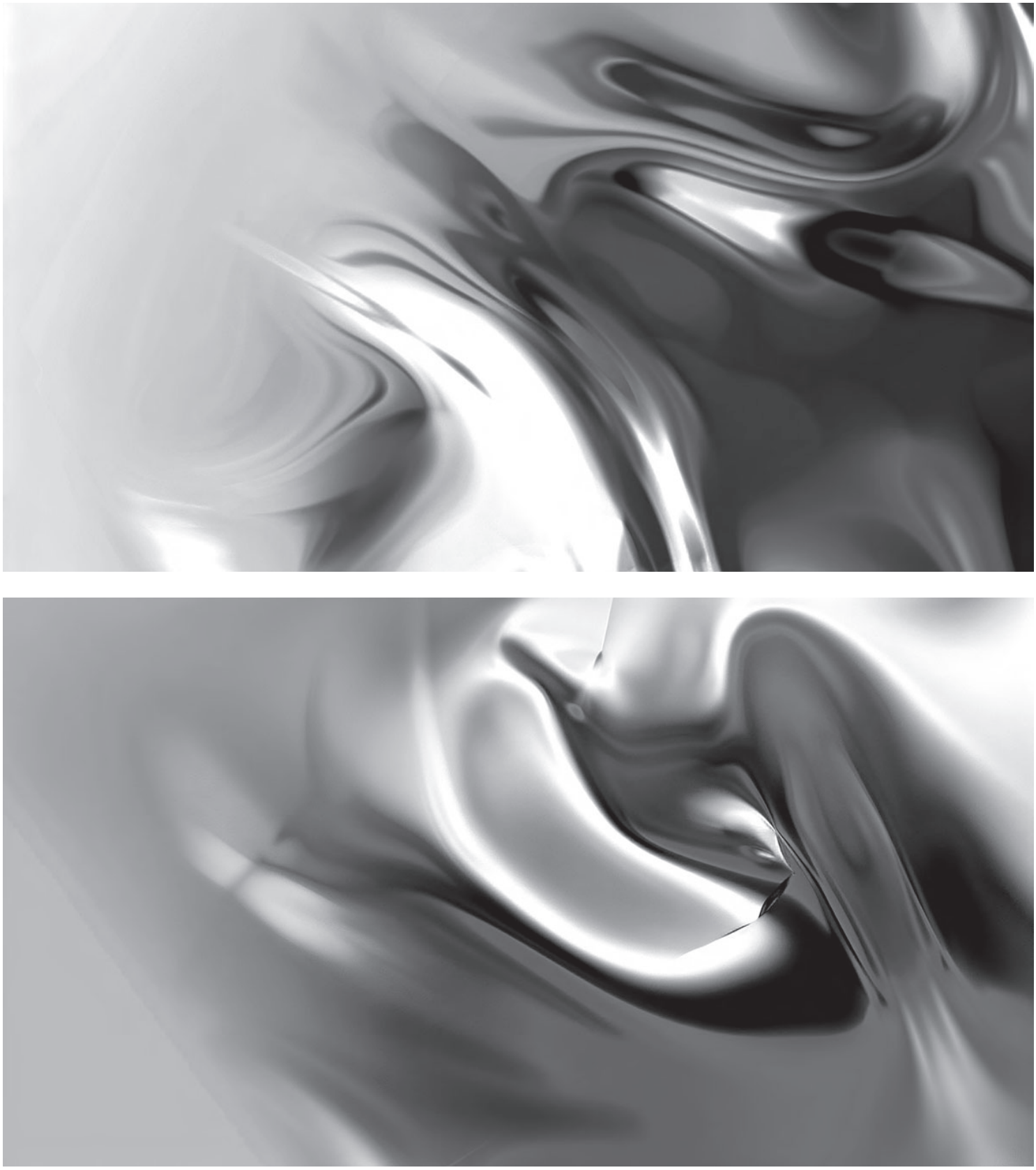
Variation on iridescence.



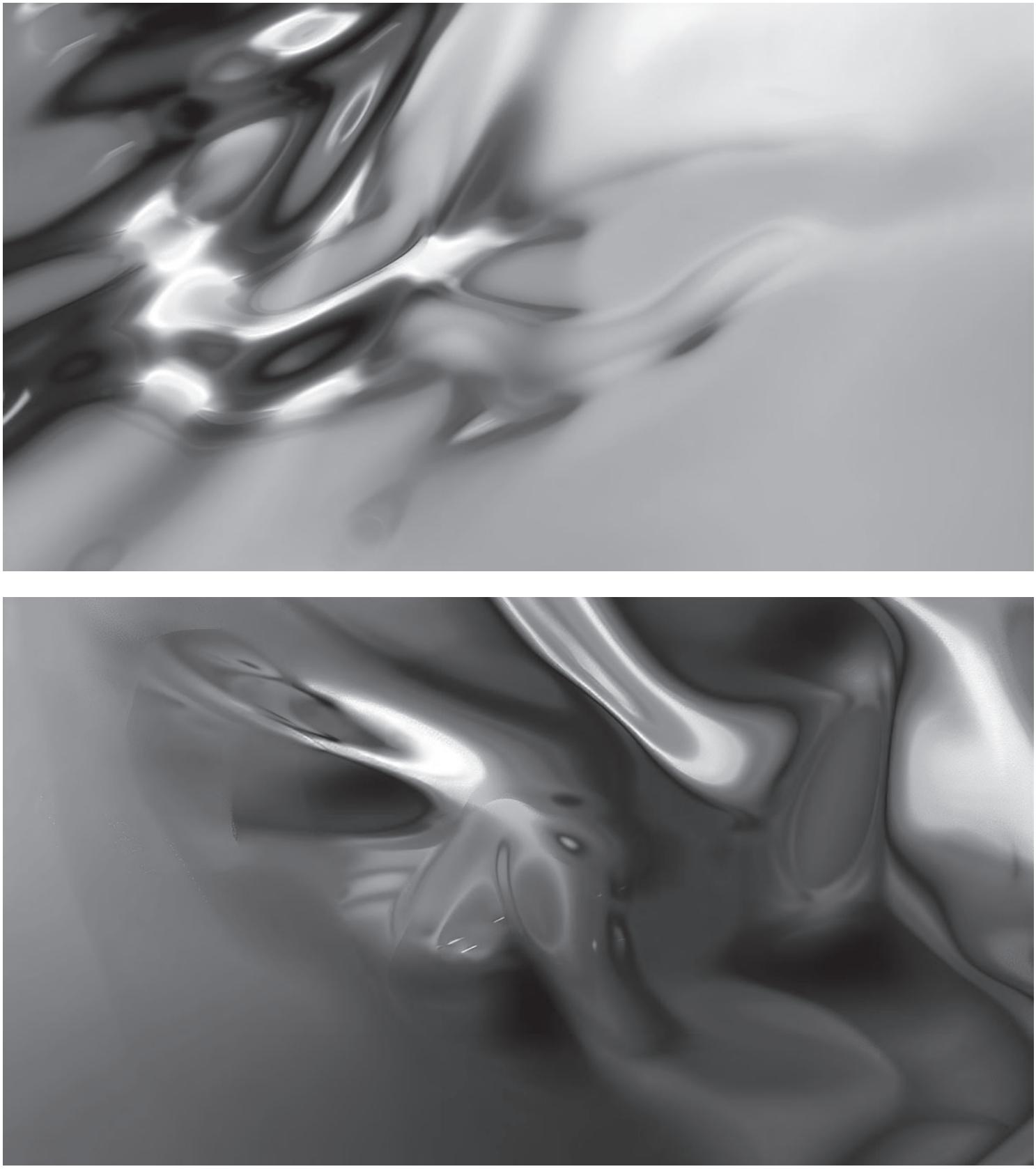
Variation on iridescence 1.



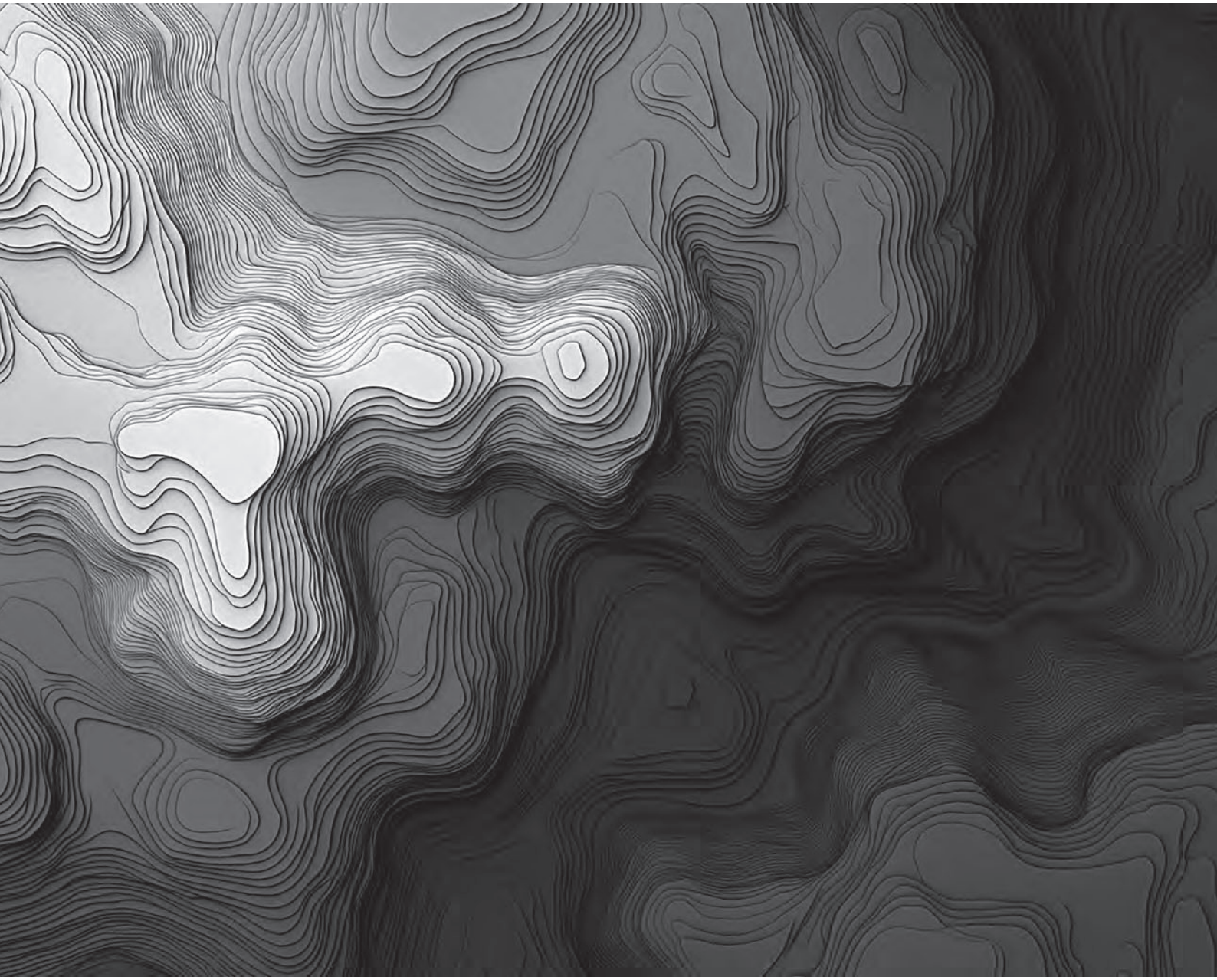
Variation on iridescence 2.
Variation on iridescence 3.



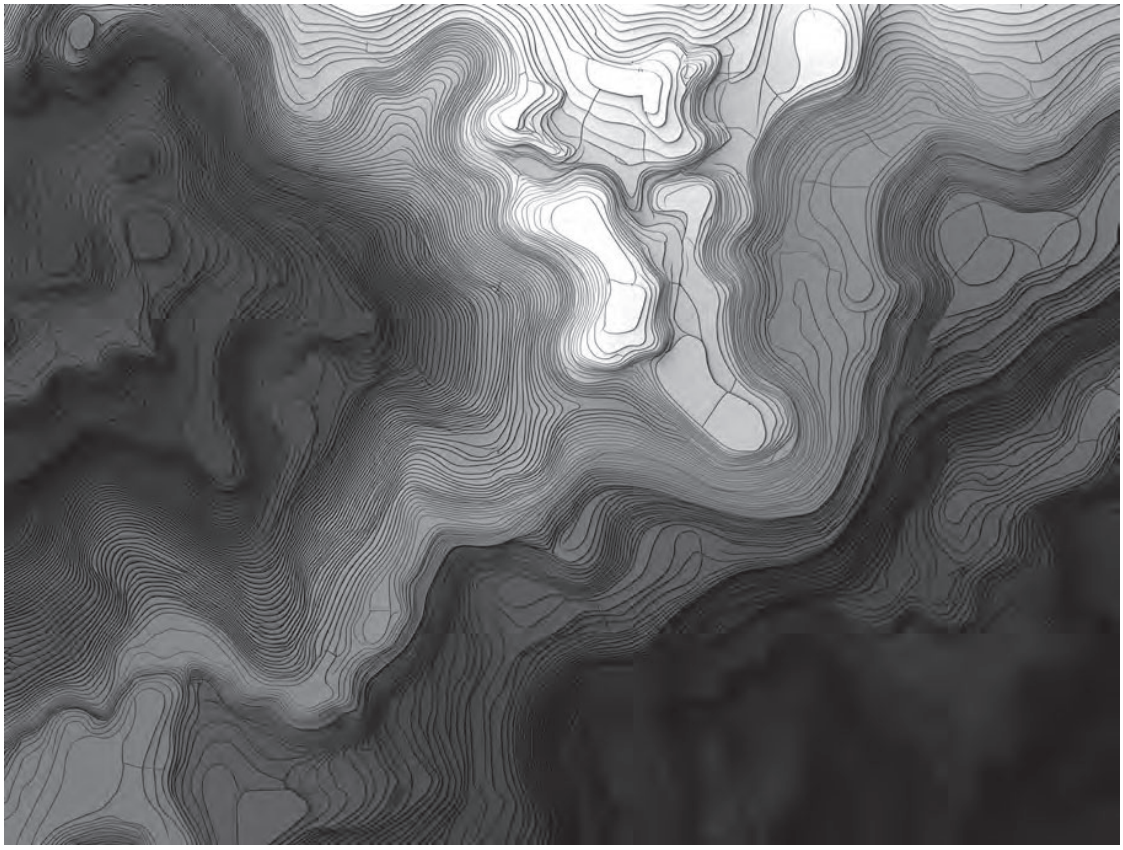
Variation on water iridescence 1-2.

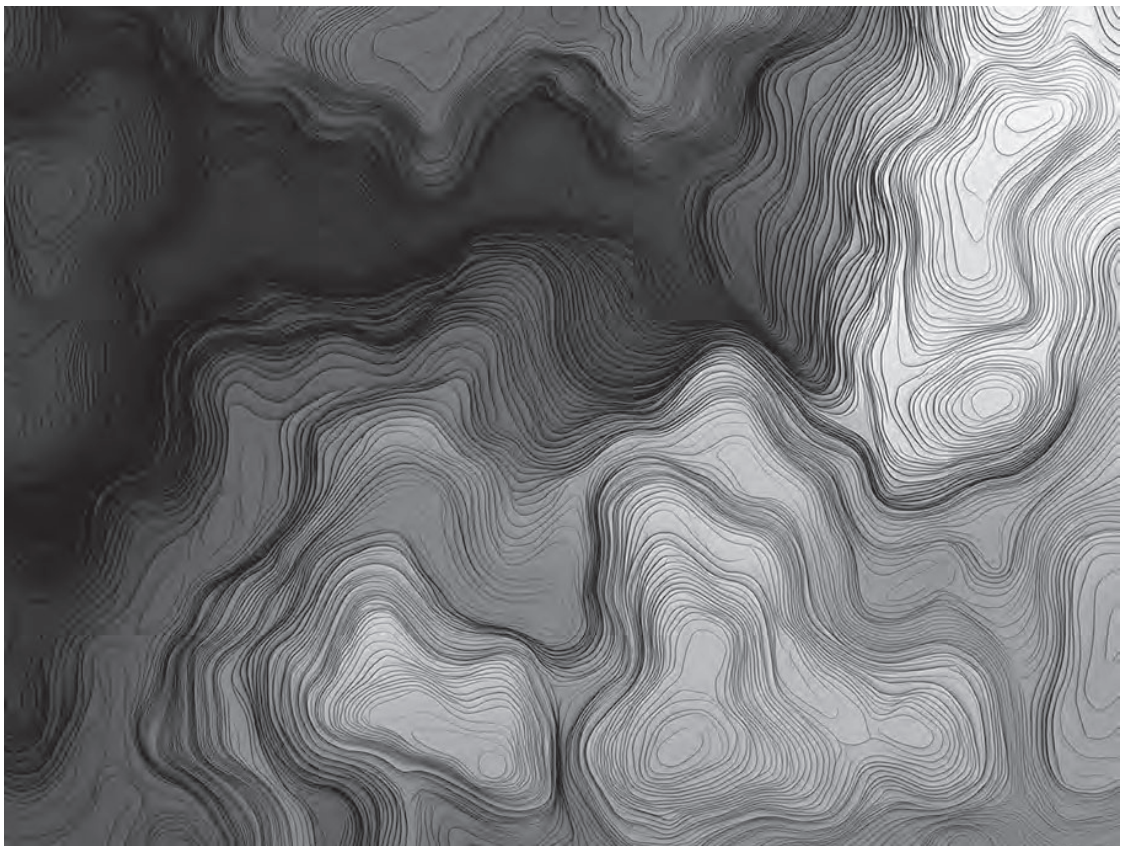


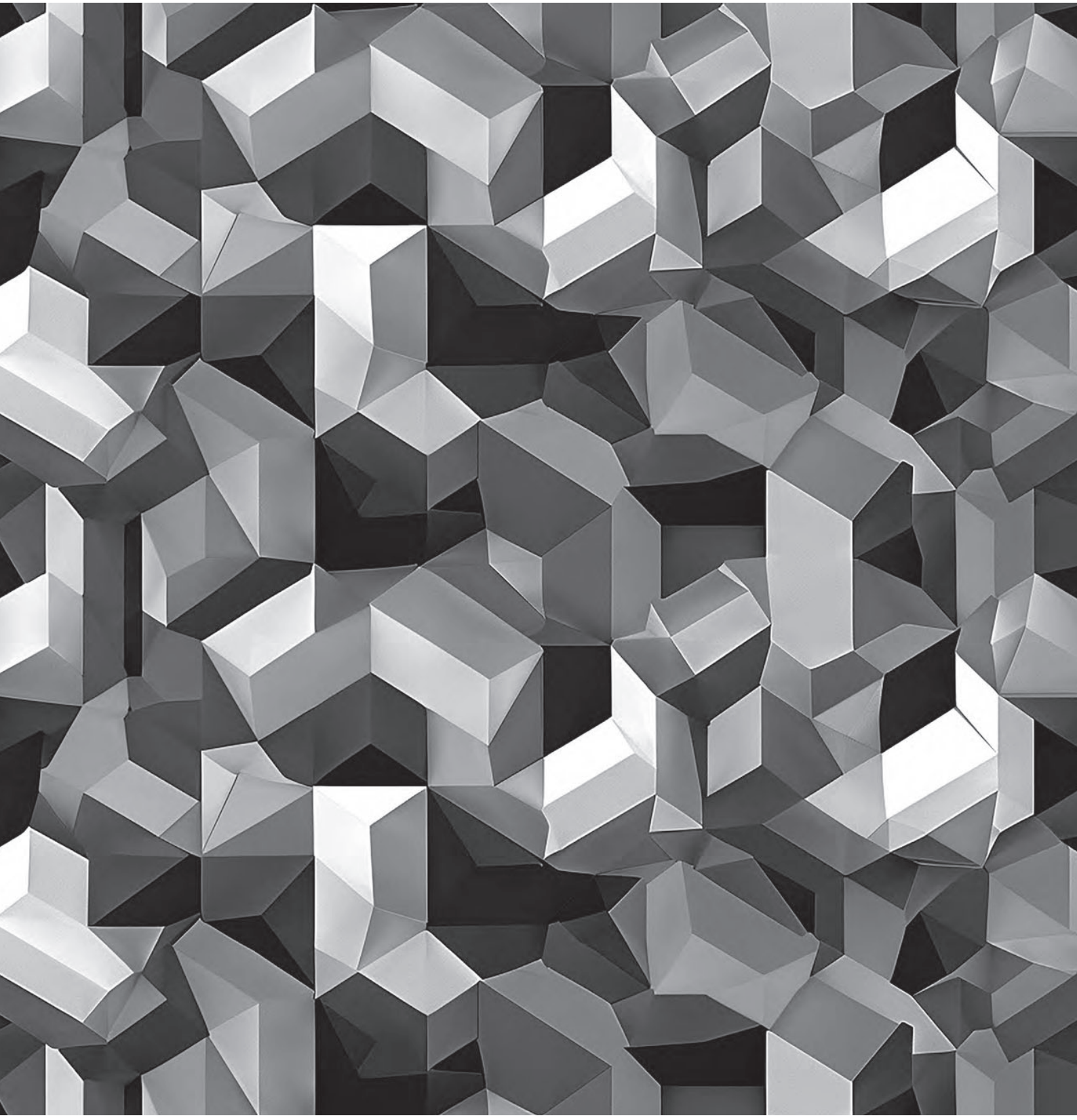
Variation on water iridescence 3-4.



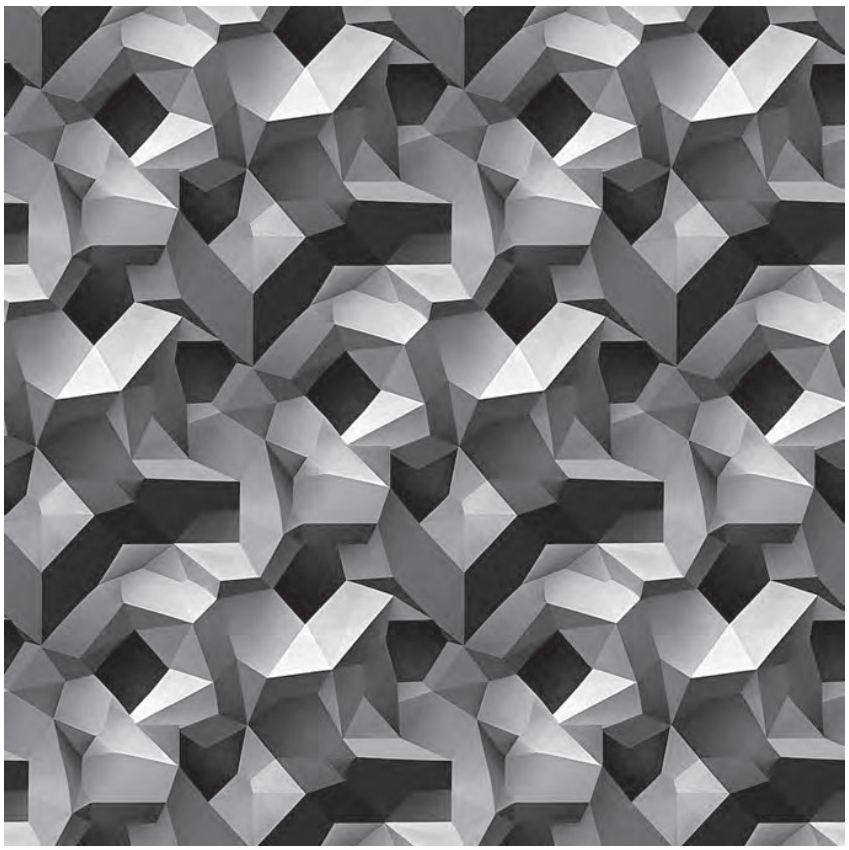
Variation on abstract topography.
Variation on abstract topography 1-2.

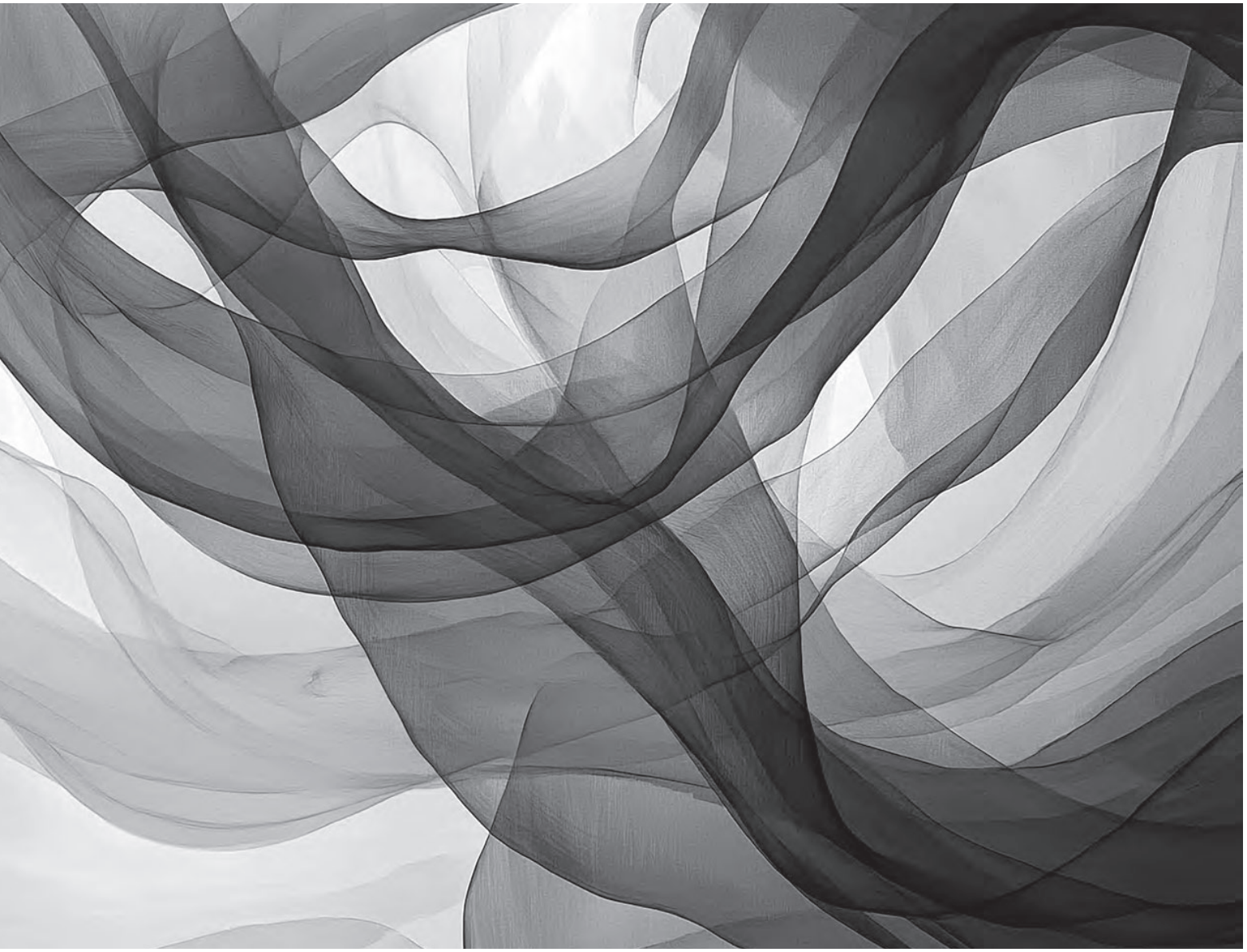






Variation on Pensrose tiling.
Variation on Pensrose tiling 1-2.





Variation on trasparent iridescence.
Variation on trasparent iridescence 1.

Within this framework, the inherently biased rhetoric of constant innovation cultivates a distorted sense of fascination that tends to magnify the transformative potential of novel solutions, severely restricting critical inquiry into their true implications. Consequently, there is a risk of fostering skewed expectations, incapable of adequately reflecting the intricate and evolving relationship between technological advancement and its tangible effects on research and professional practice within architectural and urban design. Overestimating digital technology's disruptive and transformative capabilities, an approach previously seen throughout the 1980s and 1990s with the introduction of early CAD and GIS software, diminishes our capacity to accurately evaluate evolving patterns of human-machine interaction. This overemphasis threatens to obscure the essential transversal skills and interdisciplinary knowledge required for an authentically thoughtful and innovative perspective. Moreover, it disregards the fundamental notion that no technological instrument, irrespective of its sophistication, can function independently of subjective human agency. This agency remains central to processes of selection and experimentation, inherently marked by unpredictability and the integration of distinctly human qualitative reasoning with the predictive or generative methods characteristic of artificial intelligence.

Beyond AI-washing: rethinking data and cracking the black-box

Central to this emerging diffuse storytelling about artificial intelligence is the subtle but problematic phenomenon of *AI-washing*, characterized by the indiscriminate and ubiquitous application of the AI label to nearly every digitized process related to knowledge creation, formalization, and spatial design. This generalized usage severely hampers our capability to establish a coherent semantic and technological framework, which is essential for adequately capturing and articulating the distinctive attributes of this specific form of new digital *epistemic machinery* (Knorr Cetina, 1999). Fundamentally, this narrative creates continuous conceptual and operational ambiguity, ultimately distorting perceptions of the broader field of artificial intelligence: it limits the genuine potential of AI-based tools. It obstructs transparent and unbiased discourse concerning their application's methodological, technical, and ethical implications: «The task is to remain sensitive to the terrain and to watch the shifting and plastic meanings of the term 'artificial intelligence' − like a container into which various things are placed and then removed − because that, too, is part of the story. Simply put, artificial intelligence is now a player in the shaping of knowledge, communication, and power. These reconfigurations are occurring at the level of epistemology, principles of justice, social organization, political expression, culture, understandings of human bodies, subjectivities, and identities: what we are and what we can be» (Crawford, 2021:19).

The matter acquires particular significance and urgency within architectural, urban, and territorial practices, as it redefines the traditionally analogical and deductive methods and expressions of design experimentation while reshaping the essential construction of interdisciplinary connections. Simultaneously, it threatens to obscure the possibilities of these emerging instruments, especially in contexts beyond purely predictive or quantitative uses, potentially overlooking their evolutionary and adaptive flexibility. Occasionally, the actual effort involves moving beyond a form of *enchanted determinism* that often surrounds new technologies, imbued with a sense of rigor, scientific credibility, efficiency, and authority, whose underlying logic remains, in the end, partially elusive to our complete comprehension: «The historian of technology Alex Campolo and I call this *enchanted determinism*: AI systems are seen as enchanted, beyond the known world, yet deterministic in that they discover patterns that can be applied with predictive certainty to everyday life. In discussions of deep learning systems, where machine learning techniques are extended by layering abstract representations of data on top of each other, enchanted determinism acquires an almost theological quality» (Crawford, 2021: 213).

To partially untangle this complexity and better define the contours of certain specific aspects of this conundrum, it might be necessary to begin from the raw material, the fundamental resource upon which the entire *extractive industry* of artificial intelligence rests: data. «Fundamentally, the practices of data accumulation over many years have contributed to a powerful extractive logic, a logic that is now a core feature of how the AI field works. This logic has enriched the tech companies with the largest data pipelines, while the spaces free from data collection have dramatically diminished. How data is understood, captured, classified, and named is fundamentally an act of world-making and containment. It has enormous ramifications for the way artificial intelligence works in the world and which communities are most affected. The myth of data collection as a benevolent practice in computer science has obscured its operations of power, protecting those who profit most while avoiding responsibility for its consequences» (Crawford, 2021: 121).

Indeed, the multifaceted universe of data represents the indispensable raw material from which the identity of various artificial intelligences and their practical applications are inevitably derived. Data are frequently portrayed as the new focal point to leverage, enabling innovative approaches to resolving design-related and strategic challenges. The conception of data as a new, fundamental raw material arises precisely from the preconceived assumption that the vast availability of statistical, numerical, and quantitative sources can offer an increasingly robust and unquestionable foundation for any interpretative, cognitive, or design-related process, even when the object of such transformation is the built environment.

Nevertheless, foundational to every predictive algorithmic paradigm, this methodological stance risks becoming an implicitly reductive *quantitative cage*, wherein design or strategic reasoning is exclusively based on indicators and computational methodologies. Consequently, it struggles to integrate multiple qualitative and interpretative dimensions adequately. The apparent objectivity of data does not inherently guarantee enhanced validity or transparency in decision-making. Still, it can amplify implicit biases, particularly when the limitations of data collection and analysis technologies, possible distortions embedded in datasets, and, more broadly, the political and ethical implications associated with the instrumental manipulation of information are overlooked. The shift from a potential dialogical enrichment toward mere mechanical quantification highlights deep-seated tensions between the tendency to simplify complex processes and the necessity of preserving a nuanced and multifaceted

approach to spatial knowledge and design practices. Without careful reflection on the nature of data, their origins, and how they are processed, decision-making procedures, including those related to spatial matters, risk shifting toward *numerical governance*. Within this framework, the power to define priorities lies predominantly within an artificial machine that remains only partially intelligible and is managed by those who control the technological infrastructures responsible for data collection, organization, and clustering. This risk becomes even more evident when acknowledging that data, by their intrinsic nature, never represent a neutral snapshot of reality; instead, they embody choices made before their collection and encoding and subsequent decisions concerning their storage, accessibility, and structuring.

Thus, the illusion of numerical objectivity poses a methodological challenge that cannot be easily circumvented. Indicators, metrics, and the specific types of data-feeding neural networks, large language models (LLMs), and generative artificial intelligence tools inherently shape subtle constraints that can rigidly and covertly influence outcomes, particularly when forecasting possible future scenarios. To put it in René Thom's words, *prédire n'est pas expliquer* (Thom, 1991): accepting predictive capability as a substitute for explanatory insight carries the risk of mistaking correlation for causation while neglecting the inherent complexity involved in spatial design as a complex decision-making process. This is also tied to our ability to critically reconsider the role of the *qualitative* within processes that appear to be governed solely by the logic of the discrete, reminding us of the operational need for an interplay between the two approaches: «As Enrico Berti has observed, placing emphasis on quantity at the expense of quality arises from a unifying philosophical intention. Explaining the diversity of phenomena through a single generative principle has always been the ambition of religions and metaphysics. In contrast, highlighting the diversity of appearances and the fundamental heterogeneity of qualitative differences characterizes an empiricist approach, which is less ambitious and more cautious. At the same time, in a certain sense, by definition, every theoretical effort inevitably moves toward unification» (Thom, 1980: 474. Author's transl.).

A second crucial aspect relates to the intrinsic opacity of the machine learning systems driving contemporary artificial intelligence applications, marking a fundamental shift from the earlier digital paradigm, where computational processes were governed by machine languages that could be explicitly defined, interpreted, and altered. Unlike previous digital paradigms, the contemporary landscape is shaped by hybrid black-box neural networks that learn autonomously and frequently evolve into systems whose internal workings become blurred, even to the designers who created them. As Kate Crawford aptly observes: «That deep learning approaches are often uninterpretable, even to the engineers who created them, gives these systems an aura of being too complex to regulate and too powerful to refuse. As the social anthropologist F.G. Bailey observed, 'obscuring by mystification' is often employed in public settings to argue for a phenomenon's inevitability. We are told to focus on the innovative nature of the method rather than on what is primary: the purpose of the thing itself» (Crawford, 2021: 214).

It is, therefore, evident that the so-called *black-boxing* becomes crucial for understanding both the cognitive and decision-making processes and the following epistemological implications. While the computational power of neural networks enables increasingly sophisticated analyses and simulations, it simultaneously renders the mechanisms underlying information extraction and interpretation, and the resulting predictive scenarios inevitably obscure. When computational parameters and evaluative criteria remain confined within a *black box* of proprietary codes or protected technologies, verifying whether the output genuinely aligns with a project's objectives or determining the extent to which the initial data conditions the results becomes challenging. This lack of transparency inevitably influences perceptions of professional roles and the emerging skills required in design-related fields. Architects and urban planners risk evolving into a sort of hybrid figure, one part digital-input operator, providing *natural language* prompts to generate scenarios for spatial transformation, and one part digital laborer charged with selecting, systematizing, and verifying the outcomes of a process whose hidden variables only come to light in the final stage of this new *digital assembly line*. This inevitable reliance on increasingly intricate systems presents the challenge of articulating new forms of shared responsibility, encompassing designers, programmers, and researchers within a network of accountability, reexamining the foundational principles of transparency, equity, and inclusion underlying the entire design process, regardless of scale.

Pipeline for AI-based spatial critical practice: the discomfort zone

In the face of ever-evolving digital tools and the often contradictory, lofty rhetoric surrounding them, the debate on the use and impact of artificial intelligence in design and built environment disciplines has reached a turning point. On the one hand, a growing awareness suggests that another redefinition of epistemological and practical foundations across research and professional contexts is inevitable, particularly given the risk of placing near-blind trust in machine learning systems capable of pervasively automating increasingly extensive phases of the design process. Conversely, the non-negotiable value inherent in the distinctly human aptitudes for interpretation, complexity management, fragmentation, qualitative reasoning, unpredictability, and instability is reaffirmed throughout every phase of design work.

This dual development highlights the need to establish hybrid frameworks where AI functions as an integrated, dialogical instrument rather than a mechanism for automated processes aimed at seemingly objective and prepackaged results. One approach to fostering a more aware perspective on the interplay between automation and a kind of *adaptive humanization* emerges from both theoretical and practical experimentation with a range of analog-digital pipelines, fully harnessing the potential of *digital craftsmanship* through continual cycles in which outcomes are fed back into the process and critically reexamined, following digital reworking, AI is transformed into a sort of responsive synthetic partner. A hybrid pipeline that frees AI from the black-box entanglement to which it is currently confined would enable a range of actors, beginning with the specialized expertise of architects and urban planners, to understand the underlying algorithmic models, experiment directly by adjusting specific design parameters, and creatively reimagine the broader

interplay among various AI-based tools, all within a workflow adaptable to the intended outcomes.
Recent advances in *explainable AI* (XAI) confirm the possibility of offering real-time feedback regarding how neural networks generate their responses and proposals, highlighting the variables that most strongly influence specific solutions. Rather than functioning just as a problem-solving tool, machine learning now emerges as part of a broader *cognitive ecosystem*, where algorithms serve as a single component to broaden design horizons and expedite analysis and synthesis without relinquishing the human prerogative to control the final direction.
Moving beyond the hazardous notion of a sweeping *digital delegation*, superficially fueled by AI's pervasive influence, also underscores that spatial transformation and design cannot be restricted to linear frameworks or solutions rooted exclusively in quantitative methodologies. This experimental mode of perpetual negotiation and contradiction, this kind of *analog-digital discomfort zone*, serves as a critical practice, compelling us to reside in a continuously hazardous balance between *apocalyptic* and *integrated* positions, as Umberto Eco (Eco, 1964) would say. In a digital era when both stances appear founded on seemingly monolithic and irrefutable arguments, this *discomfort zone* acts, in essence, as a cognitive and practical mindset that enables the designer to surface new questions and tensions rather than passively accepting predetermined solutions. It offers an invitation to revisit evaluative criteria, to confront contradictions and conflicts rather than avoid them, and, crucially, to remember that digital technologies, now more than ever, are *mass phenomena* whose scope must be examined with that critical spirit that remains our best defense against intellectual complacency and all forms of orthodoxy (Eco, 1964).
Among the tricks most suited to testing this laboratory of critical practice, image production again emerges as one of the most promising. Graphic representation has long served as a medium for dialogue between the human-centered and technical realms, effectively weaving together qualitative factors with more strictly numerical-quantitative issues even before parametric modeling technologies and artificial intelligence tools. This mediating function is more essential than ever in the present day, although it is not free from ambiguities. An image generated or processed by AI-driven tools can rapidly shift from a vehicle for understanding to a tool of all-encompassing seduction and *synthetic authenticity*. While images can also be leveraged to convey ambiguous rhetoric or misleading simplifications, such as ubiquitous photorealistic renderings that systematically obscure critical or unresolved dimensions of a project, it may be prolific to explore the hybrid pipeline and digital craftsmanship through representations deliberately distanced from the hyper-realism of contemporary rendering.

Hybrid visual experiments

The images accompanying this text result from a human-digital negotiation in which both the AI-driven element and a craft-based analog approach are pushed to their extremes. The aim is not so much to produce meaningful or aesthetically appealing visions as to rethink the generative process that makes them. These works are among the test shots comprising hundreds of samples of each type, from which the covers for issues 104, 105, 106, and 107 of *Territorio* were derived. For every final version, dozens of *variations on an abstract theme* were created based on a natural-language textual description used as a generative prompt in one of the most employed AI text-to-image tools: Midjourney. The decision to remain within a strictly abstract framework was motivated by the desire to push this AI tool, primarily designed to produce realistic images, beyond conventional boundaries and toward perhaps more unexpected and compelling outcomes. More broadly, the experiment explores the circularity of the *variation process* originating from a text, whose visual output also becomes part of the AI's multimodal reworking, creating a continuous hybridization between human analog input and synthetic data.
Variation functions not only as a technical-compositional process but also as a conceptual device enabling an investigation of the interplay between stability and change, coherence, and diversification, thereby revealing the nuances that emerge from a single generative core. Taking its cue from Raymond Queneau's Exercises in Style (Queneau, 1947), this approach provocatively engages with language and its reworking through the AI, exploring the modulation of expressive visual registers rather than enhancing the outcome's semantic or iconic dimension, thereby testing the intrinsically fluid and unstable character of literal descriptive text. Digital variation takes on a less transparent, more enigmatic, and cryptic role in this contemporary context. It is inextricably bound to AI models that partially escape human control and, in their inherent black-boxiness, also raise significant critical questions about the purpose and very nature of the new competencies that architects and urban planners must first understand and then develop. These range from the ability to trigger design ideas beginning with written text, employing increasingly effective prompts for interacting with large language models (LLMs), to the more complex skill of modifying and adapting analog-digital tool pipelines for various purposes while paying growing attention to the visual spatialization of phenomena. In this context, it could be intriguing to witness how the many authors from different countries participating in the ongoing Venice Architecture Biennale 2025 are interpreting these topics, given that AI is explicitly highlighted among the principal themes of the Armenian, Bulgarian, Japanese, Pakistani, and Chilean pavilions.

References

Crawford K., 2021, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven: Yale University Press
Eco U., 1964, *Apocalittici e integrati*, Milano: Bompiani
Freeman J., 2025, *Student Generative AI Survey 2025*, London: Higer Education Policy Institute www.hepi.ac.uk/2025/03/10/will-the-use-of-generative-ai-shift-higher-education-from-a-knowledge-first-system-to-a-skills-first-system (access: 2025.03)
Knorr-Cetin, K., 1999, *Epistemic Cultures: How the Sciences Make Knowledge*, Cambridge: Harvard University Press
Queneau, R., 1947, *Exercices de style,* Paris: Gallimard
RIBA AI Report, 2024, London: The Royal Institute of British Architects https://riba-prd-assets.azureedge.net/-/media/GatherContent/Business-Benchmarking/Additional-Documents/RIBA-2024-AI-Report.pdf?rev=7fc210b741f3452b9052b701c243b7dd (access: 2025.03).
Thom R., 1991, *Prédire n'est pas expliquer*, Paris: Champs science
Thom R., 1980, Qaulitatà/Quantità. In: Aa.Vv., *Enciclopedia Einaudi*, Torino: Eianudi.