Kempff J.I. (2026). ‘Simulated Certainty:
Architectural Knowledge from Models
to Machine - Learning Surrogates’,
University College London, pp. 1–7.

Simulated Certainty: Architectural Knowledge from Models to Machine Learning Surrogates

Kempff J.I., Bartlett School of Architecture, University College London, *jose.kempff.25@ucl.ac.uk*

Simulation has become a central epistemic instrument in architectural design, enabling the anticipation and evaluation of building performance across structural, environmental, and fluid domains. Beyond its predictive function, simulation operates as a mode of knowledge production, mediating between abstract models and empirical phenomena. This literature review examines simulation theory as a framework for understanding the contemporary rise of machine learning surrogate models in architectural practice. Drawing on philosophical, historical, and technical literature, it traces the evolution of simulation from deterministic, physics-based solvers toward data-driven approximations learned from precedent. Particular attention is given to how surrogate models reconfigure longstanding epistemological tensions surrounding abstraction, transparency, and trust. While surrogate models offer significant advantages in speed, interactivity, and accessibility—especially in early-stage design—they also introduce essential *epistemic opacity* and heightened dependence on training domains. The review argues that these models do not constitute an epistemological rupture, but rather intensify existing challenges in simulation-based reasoning. Ultimately, it proposes that architectural simulation is entering a hybrid phase, in which surrogate models and traditional solvers must be critically integrated, demanding new forms of architectural literacy grounded in contextual judgement and *critical trust*.

Keywords: Architectural Simulation, Surrogate Models, *Epistemic Opacity*, *Computational Reliabilism*, *Critical Trust*, Hybrid Workflows, *Epistemic Seduction*, Design Epistemology.

Introduction

Simulation is now woven into the fabric of architectural design, enabling practitioners to anticipate structural behaviour, environmental performance, or fluid dynamics long before construction (Kalay, 2004; Steenson, 2017). From early thermal and structural models to contemporary environmental and multi-physics simulations, architects have relied on computational tools to extend perceptual and cognitive reach beyond what drawings or physical models alone can provide. What distinguishes simulation in architecture is not merely its predictive ambition, but its *epistemic function*: simulation shapes how buildings are understood as systems of forces, flows, and relations rather than static formal compositions.

As Varenne (2021) argues, simulations do not simply represent models; rather, they enact them, creating new epistemic objects that exist between idealised abstractions and empirical phenomena. Framed epistemologically, this positions simulation not as a neutral instrument but as a mode of knowledge production, in which understanding emerges through mediation rather than direct representation. This aligns with broader accounts of scientific and design epistemology that emphasise models and simulations as constructive, interpretive devices operating under conditions of abstraction and uncertainty (Simon, 1969).

Historically, architectural simulation has been grounded in deterministic physics. Techniques such as finite element analysis (FEA), daylight modelling, and computational fluid dynamics (CFD) rely on explicit mathematical formalisms that encode causal assumptions about material behaviour, energy transfer, or fluid flow. Although complex, these assumptions remain open to inspection by expert users. Recently, however, machine learning (ML) approaches—particularly surrogate models—have emerged as faster alternatives. These methods approximate traditional solvers by learning statistical relationships from datasets, offering near real-time feedback.

This literature review examines simulation theory as the foundation for understanding the rise of surrogate models in architectural practice. ML is treated not as an epistemological rupture, but as the latest development within a longer trajectory—one that brings longstanding tensions between abstraction, prediction, trust into sharper relief.

Simulation Theory: Models, Mediation and Abstraction

Simulation theory encompasses a range of philosophical and technical approaches concerned with how simulations function as instruments of knowledge, prediction, and exploration (Varenne, 2021; Durán & Formanek, 2019). Across these accounts, simulations are understood as processes that mediate between abstract models and empirical phenomena, actively shaping what can be known about complex systems.

Crucially, Varenne (2021) distinguishes between models and simulations. Models are static structures describing a system in simplified form, whereas simulations are dynamic processes that make these structures behave over time. By generating trajectories and exposing inconsistencies, simulations produce knowledge unavailable to static representation. This distinction is vital in architecture, where simulation enables the exploration of building performance as an evolving system.

In this review, architectural simulation is understood specifically as the modelling of behavioural or performative systems rather than the geometric or spatial representation of form alone. A three-dimensional digital model, while essential to contemporary practice, does not constitute a simulation unless it encodes assumptions about behaviour and generates evaluative or predictive outcomes. Witt (2021) situates this within a genealogy of mathematical abstraction, characterising simulations as '*epistemic apparatuses*' that restructure how design problems are formulated. This foundation is essential for understanding how surrogate models transform simulation's established epistemic logic.

Simulation in Architectural Practice

Despite its growing prominence, simulation has long been marked by limitations. Steadman (2014) argues that building simulations frequently diverge from real performance due to oversimplified assumptions: steady-state thermal conditions, idealised occupancy, or linear material properties. Crucially, however, Steadman insists that these limitations remain legible: they arise from identifiable choices that can be scrutinised. Trust in traditional simulation is therefore grounded in interpretability rather than accuracy alone.

Picon (2010) extends this by noting that the digital turn fundamentally altered the materiality of architecture yet retained a connection to physical logic through simulation parameters. Architecture has historically tolerated partial opacity in computation—many designers rely on tools they cannot fully audit—but such opacity has typically been layered onto systems governed by explicit causal logics. With the emergence of ML-based simulations, the character of opacity changes. What becomes inaccessible is no longer merely implementation detail, but the relationship between assumptions, mechanisms, and outcomes. This shift marks a critical epistemic threshold that distinguishes surrogate models from earlier forms of architectural simulation.

Machine Learning as a New Regime of Simulation

The recent adoption of machine learning methods in simulation is driven primarily by computational cost. High-resolution CFD or structural analyses can require hours or days of processing, particularly for complex geometries common in contemporary architectural design. ML models promise to reduce this cost by approximating solver behaviour using data rather than equations. Rather than replacing simulation outright, these methods instantiate simulation through an alternative computational logic.

ML simulations can be understood as data-driven behavioural models. To understand this shift, one must contrast the architectural logic of the two systems. In a traditional solver, the workflow is transparent and linear, proceeding from input variables through physics equations to a final output determined by causal logic. If the simulation fails, the error lies in the explicit parameters or the mesh. In a surrogate model, the workflow changes; inputs are filtered through hidden layers of statistical correlations to generate an output. The physics are not calculated but '*hallucinated*' based on probability. The central mechanism—the *black box*—is not a solver of laws but a massive, opaque lookup table of likelihoods. This fundamentally alters the architect's ability to audit the results; one can inspect the data, but not the reasoning.

Durán and Formanek (2019) describe this condition as *essential epistemic opacity*: even with full access to code and parameters, the internal reasoning remains inaccessible to human understanding. Importantly, this opacity differs from mere inaccuracy. While traditional simulations may be wrong due to simplifying assumptions, those assumptions remain available for critique. In contrast, machine learning simulations may be right within their training domain while offering no intelligible account of why.

Shen (2022) argues that trust in such systems should arise not from interpretability, but from stable and reliable interaction over time. Yet it risks conflating visual plausibility with physical validity—particularly in design cultures where simulation imagery functions simultaneously as analytical evidence and representational artefact. Machine learning therefore does not displace simulation theory; it reframes it. By prioritising speed, interactivity, and approximation, it intensifies longstanding tensions between abstraction, prediction, and trust.

Surrogate Models as the Contemporary Development of Simulation

Surrogate models are the most explicit application of machine learning to simulation in architecture and engineering: trained on datasets of inputs and solver outputs, they approximate results in milliseconds and are therefore attractive for early-stage exploration and optimisation. As Tamke et al. (2018) observe, this speed allows for a fundamental shift in design workflows, moving from post-design validation to active, real-time performance feedback.

In fluid dynamics, recent studies demonstrate that deep-learning surrogates can approximate steady-state flow fields (Guo, Li & Iorio, 2016) and complex wake patterns (Thuerey et al., 2020) with high visual fidelity, bypassing iterative solution entirely. More recent work has focused on handling complex and irregular geometries through graph-based representations. Lino, Müller, and Thuerey (2024) trained graph neural networks on large-scale datasets of three-dimensional flow fields, achieving impressive interpolation accuracy across varied configurations.

However, performance degrades sharply outside the training distribution. For architects, these failures are distinct because they are often invisible. While a traditional structural analysis might return a '*non-convergence*' error for an overloaded beam, a surrogate model may confidently predict stability for an impossibly thin member simply because its training data lacked specific failure examples. Similarly, a wind tunnel surrogate might miss a critical vortex shedding event entirely due to a unique geometric feature, returning a 'smooth' visualization that conceals reality. This is consequential in architecture, where novelty of form is a disciplinary expectation.

Comparable patterns emerge in structural and hydrodynamic applications. Alamu, Liu, and Zhu (2025) observe similar behaviour in AI-assisted CFD systems: strong performance under familiar conditions, coupled with abrupt failure at regime boundaries. These findings indicate that surrogate models do not merely approximate solvers; they encode the scope of their validity within the data itself.

Viewed collectively, this literature reveals both the promise and fragility of surrogate models. Because surrogates infer behaviour from examples, they cannot reliably simulate conditions that differ significantly from those they have encountered. In architectural contexts characterised by atypical geometries, hybrid programmes, and speculative forms, this constraint places surrogate models in direct tension with disciplinary ambitions. Surrogates therefore expose a critical epistemic trade-off: they dramatically expand the speed and accessibility of simulation, while simultaneously narrowing the conditions under which simulation can be trusted.

Epistemological Implications for Architectural Design

The introduction of surrogate models intensifies longstanding epistemological tensions between transparency and opacity, causal explanation and pattern recognition, and analytical validity and visual persuasion. Central to these tensions is the issue of trust: Traditional simulations, despite their simplifications and frequent inaccuracies, allow practitioners to engage critically with underlying assumptions. Surrogate models, by contrast, may fail silently. Their opacity means that inaccuracies can remain invisible unless they manifest as dramatic deviations from expected behaviour. This is particularly problematic in architectural culture, where simulation outputs often circulate as images, diagrams, and performance narratives. As Carpo (2017) observes, contemporary design increasingly relies on representational artefacts of performance as instruments of persuasion as much as analysis. In such contexts, visually convincing surrogate outputs may exert epistemic authority disproportionate to their reliability. While del Campo (2022) argues that this 'hallucinatory' quality is a generative feature of neural architecture that liberates design from strict causality, the risk remains that smooth outputs can mask uncertainty—a condition of *epistemic seduction*.

Steadman's (2014) critique thus acquires renewed urgency: the claim that simulations are often wrong becomes more troubling when wrongness is no longer traceable to legible assumptions but embedded within statistical generalisations.

A second implication concerns expertise. As surrogates accelerate exploration, architects risk shifting from interpreters of causal systems to curators of machine-generated possibilities. Designers must develop intuition not only for physical behaviour but for the biases of ML-systems. A likely outcome is hybrid verification workflows, where surrogates guide early-stage exploration while traditional solvers are retained for validation. Such practices align with Durán and Formanek's (2019) *computational reliabilism*, where trust is grounded in demonstrated performance within known bounds

Conclusion

Simulation in architecture has evolved from static modeling toward dynamic mediation, transitioning from deterministic solvers grounded in physical law to data-driven approximations learned from precedent. Across this trajectory, simulation has functioned not merely as a predictive instrument, but as an epistemic practice that structures how architects conceptualize buildings as systems of behavior, interaction, and performance. Machine learning surrogate models represent the latest development in this evolution, extending simulation's reach while exposing its underlying assumptions. By replacing causal computation with statistical inference, they challenge established criteria of transparency and explanation, even as they offer unprecedented speed and accessibility.

While their outputs may be accurate within constrained domains, their inherent 'e*ssential epistemic opacity*' complicates the attribution of trust. What emerges is not a crisis of simulation, but a reconfiguration of the grounds upon which simulation-based knowledge is judged. In this new regime, trust must move beyond the legible causal assumptions of traditional physics and toward a framework of *computational reliabilism*, where validity is grounded in a model's stable and reliable interaction over time. However, for architectural practice—which often prioritizes formal novelty—reliabilism alone is insufficient. The risk of '*epistemic seduction*' remains high, where visually plausible or 'smooth' surrogate outputs can mask physical invalidity or silent failures at regime boundaries.

Consequently, the future of architectural simulation will likely solidify into a coarse-to-fine hybrid workflow. In this model, surrogate models serve as the engine of *divergence*—generating thousands of rapid iterations in early design stages where speed outweighs precision. Traditional solvers are then reclaimed as the engine of *convergence*—deployed to validate selected options and certify performance. This needs that architects cultivate a new form of *biliteracy*: the intuition to guide the 'fast' statistical machine and the rigor to interpret the 'slow' physical verifier. As simulation becomes faster yet more opaque, the task for architects is not to seek absolute certainty, but to cultivate *critical trust*—recognizing simulation as a mediated form of knowledge that both reveals and conceals architectural intelligence.

References

Alamu, O., Liu, Y. & Zhu, J. (2025). 'AI-Augmented Computational Fluid Dynamics: Surrogate Models for Real-Time Simulation', Applied Sciences, 15(3), pp. 2332–2345.

Carpo, M. (2017). 'The Second Digital Turn: Design Beyond Intelligence'. Cambridge, MA: MIT Press.

Del Campo, M. (2022). 'Neural Architecture: Design and Artificial Intelligence'. Novato, CA: ORO Editions.

Durán, J.M. & Formanek, N. (2019). 'Grounds for Trust: Essential Epistemic Opacity and Computational Reliabilism', Minds and Machines, 29(4), pp. 555–578.

Guo, Y., Li, W. & Iorio, F. (2016). 'Convolutional Neural Networks for Steady Flow Approximation', Proceedings of the 22nd ACM SIGKDD Conference, pp. 481–490.

Kalay, Y. (2004). 'Architecture's New Media: Principles, Theories, and Methods of Computer-Aided Design'. Cambridge, MA: MIT Press.

Lino, A., Müller, D. & Thuerey, N. (2024). 'Large-Scale Graph-Machine-Learning Surrogate Models for 3D Flowfield Prediction', Advanced Modeling and Simulation in Engineering Sciences, 11(2), pp. 201–223.

Picon, A. (2010). 'Digital Culture in Architecture: An Introduction for the Design Professions.' Basel: Birkhäuser.

Simon, H.A. (1969). 'The Sciences of the Artificial'. Cambridge, MA: MIT Press.

Shen, M.W. (2022). 'Trust in AI: Interpretability Is Not Necessary or Sufficient, While Black-Box Interaction Is Necessary and Sufficient', arXiv preprint, arXiv:2202.05302.

Steadman, P. (2014). 'Why Are Building Simulations So Often Wrong?', Building Research & Information, 42(4), pp. 491–502.

Steenson, M.W. (2017). 'Architectural Intelligence'. Cambridge, MA: MIT Press.

Tamke, M., Nicholas, P. & Zwierzycki, M. (2018). 'Machine Learning for Architectural Design: Practices and Simulation', International Journal of Architectural Computing, 16(2), pp. 123-138.

Thuerey, N., Holl, P. & Wiewel, S. (2020). 'Deep Learning Methods for Fluid Flow Simulation', Computer Graphics Forum, 39(2), pp. 613–634.

Varenne, F. (2021). F'rom Models to Simulations'. London: Routledge.

Witt, A. (2021). 'Formulations: Architecture, Mathematics, and Algorithms'. Cambridge, MA: MIT Press.