Opportunities for Generative AI in Design, Simulation, and Optimisation of Engineering Systems

Dr Andrew Duncan.
*(*Alan Turing Institute & Imperial College*, United Kingdom).*

Abstract

Generative AI is revolutionizing the design, simulation, and optimization of engineering systems by enabling faster, more efficient, and highly innovative solutions. We describe an increasingly common workflow for the use of conditional and unconditional generative models in the context of challenging engineering tasks, such as inverse design and optimal sensor placement. To be effective within the context of complex engineering challenges, several technical challenges must be overcome, including the ability to accurately characterise physical states of the system; being effective at exploring the design space; and supporting rapid exploration of the design space for optimisation subject to constraints.

1. Introduction

The integration of deep generative models into engineering workflows is transforming simulation-based optimization. Traditional approaches rely on physics-based models, numerical solvers, and empirical heuristics, which can be computationally expensive and struggle with high-dimensional design spaces. Many engineering challenges are optimization problems, aiming to determine an optimal design while satisfying constraints that are often complex and best articulated through examples. The objective function typically depends on costly numerical simulations, such as those involving partial differential equations (PDEs) or stochastic PDEs, making iterative design computationally demanding.

To address scalability, surrogate models approximate complex physical processes using machine learning techniques like Gaussian processes, deep neural networks, or polynomial chaos expansion. These models solve regression problems, mapping input parameters (e.g., geometric variables, material properties) to system responses (e.g., stress distributions, flow fields) for faster evaluations. However, they are limited in settings with stochasticity or relationships that are not easily characterised.

Generative models offer an alternative by learning underlying data distributions. They can be categorized as *unconditional models*, which approximate a distribution p_θ from sample data, and *conditional models*, which estimate $p_\theta(\cdot\,|x)$where x incorporates contextual information like boundary

conditions or design constraints. Conditional generative models are particularly valuable for inverse problems—determining an optimal design given a desired system response—and simulation tasks where they synthesize plausible solutions under physical or computational constraints.

Recent advances in generative modelling—including GANs [Goodfellow et al, 2014], Variational Autoencoders [Kingma et al, 2013], energy-based models, diffusion models [Song et al, 2021], and flow-matching—have shown promise in inverse design, topology optimization, and uncertainty quantification. For engineering applications, generative models must ensure physically consistent sample spaces, particularly in PDE-governed or multi-physics settings. Naive generative approaches may miss rare but impactful designs. Finally, rapid optimization is crucial, requiring hybrid approaches that integrate generative models with classical solvers, surrogate models, uncertainty quantification, and active learning techniques.

2. The Generative Modelling Workflow for Engineering Design

The generative AI paradigm lends itself well to a workflow for simulation-based optimisation. Figure 1 illustrates a generative AI-driven workflow for the design, simulation, and optimization of engineering systems. At the core of the workflow is the generative model, which synthesizes candidate designs based on prior knowledge stored in a database. These generated designs serve as initial inputs for evaluation and refinement.

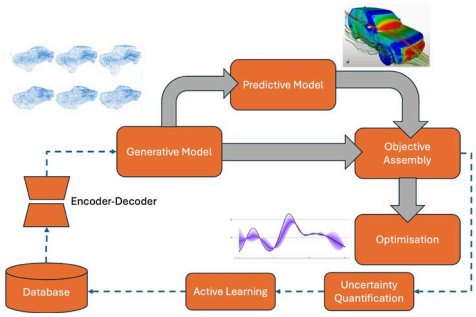

Figure 1: A schematic of a typical workflow for Generative AI accelerated simulation-driven optimisation.

Once candidate designs are generated, they are assessed using a predictive model, a surrogate model for expensive high-fidelity simulations. In simple settings, the predictive model can be readily integrated directly into the generative model. The outputs of the predictive model are aggregated to formulate the optimization problem. This can be solved using a variety of methods, ranging from black box approaches to gradient based methods, if the generative model supports it. The workflow can also incorporate active learning, detecting regions where uncertainty is high due to being out of the

original training regime, allowing the framework to adaptively refine the design space, prioritizing regions that contribute to model improvement.

3. Case Study: Optimal Sensor Placement

We apply the workflow to an optimal sensor placement for a stochastic nonlinear, partial differential equation, for which high-fidelity solutions can be obtained through a numerical solver. We seek to place sensors at positions $x_1, x_2, \ldots, x_N$ such that we minimise the uncertainty around an inhomogeneous field $\kappa = \kappa(x)$, based on noisy observations $y_i = u(x_i) + \xi_i$, where $u = G(\kappa)$, and where ξ_i are i.i.d. mean-zero random variables, for $i = 1, \ldots, N$. Here G is a stochastic, nonlinear operator representing the nonlinear numerical solver. To capture the smoothness properties of fields u, κ, we adopt resolution independent parametrisation, using an Implicit Neural Representation (INR) Autoencoder with SIREN activation functions [Sitzmann et al, 2020], yielding low-dimensional embedding vectors $z = (z_u, z_\kappa)$. When a sample is generated from the model, we can readily transform it into a (u, κ) pair using the INR encoders. For the generative model, we use a joint energy model $p_\theta((z_u, z_\kappa))$ on the pairs of encodings, trained using an energy-discrepancy loss [Schroder et al, 2023],

$$\mathrm{D_q}\,(\mathrm{p_{data}},\,\mathrm{p_\theta}) \coloneqq \int [-log p_\theta(z)] p_{data}(dz) - \int \int q(z'|z)\big[U_q(z)\big] p_{data}(dz),$$

where $U_q(y) \coloneqq -\log \int q(y|x)\, p_\theta(x) \mathrm{d}\, x$, and where $q(\cdot\,|x)$ is a Gaussian distribution with mean x and variance δ^2. This loss is a smoothed version of the score-matching loss, offering more stability at significantly lower cost.

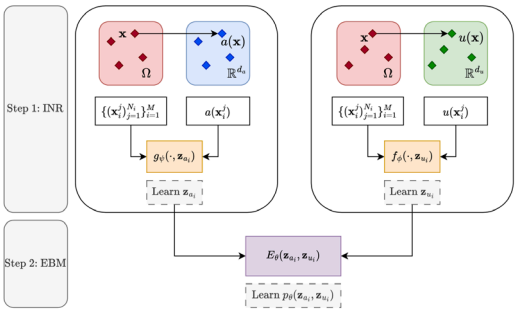

Figure 2: A schematic of the model architecture.

To find an optimal sensor configuration, we maximise expected information gain $\mathrm{U(x)} = \int \mathrm{KL}\big(\mathrm{p}(\mathrm{z_\kappa},\, \mathrm{z_\kappa}|\mathrm{x},\, \mathrm{y}) \big\| \, p_\theta(z_\kappa, z_u)\big)\big] p(y|x) dx$, of a new design point x. We use stochastic gradient ascent to optimise the utility, generating samples from the generative model for each step of the optimisation, without necessitating any further evaluations of the numerical solver.

Dataset	Data generation		Training		Inference
	10%	Rest	INR	EBM	
Boundary Value Problem	1 mins	9 mins	65 mins	34 mins	1.1 mins
Steady-State Diffusion	8 mins	70 mins	73 mins	40 mins	2.5 mins
Navier Stokes	30 mins	4 hours	91 mins	47 mins	3.2 mins
Lotka Volterra Model	3 mins	27 mins	67 mins	38 mins	1.4 mins

Table 1: Run-times for training and inference of the sensor placement workflow.

Table 1 lists the training and inference costs for a class of demonstrative stochastic PDE based models, see [Cordero-Encinar et al, 2025] for more details. We mention in passing that a direct approach using the underlying PDE solver would take days-to-weeks to run.

4. Forward Perspectives

While the above workflow can be repeatedly used across a wide range of different (e.g. boundary and initial) conditions, it remains bounds to a single task. Foundation models, i.e. task-agnostic models which can be subsequently fine-tuned and post-trained for specific tasks, are a potential opportunity for a general-purpose cross-domain tool for simulation-based optimisation. By integrating multiple modalities of engineering data (text, meshes, spatio-temporal data, etc), foundation models can be readily deployed for translation and conversion, inference and optimisation.

5. References

Goodfellow, Ian, et al. (2014) *Generative adversarial nets*: Advances in neural information processing systems 27.

Kingma, Diederik P., and Max Welling (2013). *Auto-encoding variational Bayes:* Int. Conf. on Learning Representations.

Song, Yang, et al. (2020) *Score-Based Generative Modeling through Stochastic Differential Equations*: International Conference on Learning Representations.

Sitzmann, Vincent, et al. (2020) *Implicit neural representations with periodic activation functions:* Advances in neural information processing systems 33: 7462-7473.

Schröder, T., Ou, Z., Lim, J., Li, Y., Vollmer, S., & Duncan, A. (2023). *Energy discrepancies: a score-independent loss for energy-based models*: Advances in Neural Information Processing Systems, *36*, 45300-45338.

Encinar, Paula Cordero, et al. *Deep Optimal Sensor Placement for Black Box Stochastic Simulations:* The 28th International Conference on Artificial Intelligence and Statistics.