COMBINING PHYSICS AND AI FOR BETTER AND FASTER MODELS

Engineering challenges increasingly demand models that are both accurate and computationally efficient. Physics-informed surrogate modelling (PISM) combines the strengths of physics and artificial intelligence (AI) to deliver fast, reliable predictions without sacrificing fidelity. By integrating domain knowledge with data-driven techniques, these models enable real-time analysis, design optimisation and a range of advanced engineering applications. This article explores key methodologies, practical considerations, and lessons learned from industrial applications, offering insights for practitioners and stakeholders seeking to harness the full potential of physics-informed AI for their use cases. We conclude with a brief outlook on future developments in this rapidly evolving field.

MARK VAN DEN BROEK, JURIAAN LUCASSEN AND DAAN NIEUWENHUIZEN

Introduction

Engineering design and analysis increasingly demand solutions that balance accuracy with computational efficiency. Traditional high-fidelity simulations, while precise, often impose prohibitive time and resource costs, especially in contexts requiring rapid iteration or real-time decision-making. Surrogate models have emerged as a practical alternative, offering fast approximations of complex systems without sacrificing essential predictive power.

Among these, physics-informed surrogate models (PISMs) represent a significant advancement. By embedding physical laws and constraints into data-driven frameworks, they achieve superior generalisation and reliability compared to purely empirical approaches. This capability is particularly critical in high-tech domains – such as semiconductors, mechatronics, optics, and precision systems – where performance, interpretability and responsiveness are non-negotiable.

At Sioux Technologies, we have observed recurring patterns in successful deployments of PISM: replacing computationally expensive solvers with efficient surrogate models, leveraging hybrid datasets that combine simulations and experiments, and integrating these models into real-time control and generative design workflows. These strategies enable accelerated development cycles, optimised operations, and new possibilities in diagnostics and autonomous design.

This article provides a methodology-focused overview of physics-informed surrogate modelling, with applications spanning real-time digital twins, predictive maintenance, anomaly detection, control systems, inline monitoring and (generative) design optimisation. We will explore the underlying principles, practical challenges, and lessons learned from deploying these models in demanding industrial contexts, offering insights for both practitioners and decision-makers seeking to harness the power of physics-informed AI.

The importance of surrogate models

Modern engineering and scientific design processes increasingly rely on high-fidelity simulations to predict system behaviour. While these simulations are highly accurate, they often come at a steep computational cost. A single evaluation of a complex model, such as a computational fluid dynamics (CFD) analysis of an aircraft wing or a multi-physics simulation of a dynamic system, can take hours or even days. This creates a bottleneck whenever rapid evaluation or large-scale exploration is required.

Surrogate models address this challenge by providing fast, approximate representations of the original system. Once trained or developed, predictions can typically be made orders of magnitude faster, enabling real-time usage. This speed fundamentally changes what is possible: engineers can explore design alternatives interactively, embed predictive models into control systems, or integrate them into digital twins for operational decision-making and predictive maintenance.

AUTHORS' NOTE

The authors are all affiliated with Sioux Technologies in Eindhoven (NL).
Mark van den Broek is technology manager Data Science and Machine Learning, with a broad experience in applying advanced AI techniques across multiple industries to solve complex real-world problems.
Juriaan Lucassen has a background in physics, and in his current role as a senior designer he focuses mostly on high-tech problems that benefit from integrating physics into AI-based solutions.
Daan Nieuwenhuizen has a background in physics and machine learning. He is working as an applied mathematics & AI designer, specialising in applying AI to engineering challenges in high-tech industries.

www.sioux.eu

Beyond real-time speed, surrogate models also unlock the ability to perform advanced algorithmic studies that are otherwise computationally infeasible. For example, design-space explorations aimed at optimisation, sensitivity analysis or uncertainty quantification often require thousands or millions of evaluations, which are infeasible with full-scale simulations. By replacing expensive computations with efficient approximations, surrogates make global optimisation and multi-objective trade-off studies practical within limited timelines. Other advanced techniques such as Bayesian optimisation and Monte Carlo simulations also become feasible when evaluations are nearly instantaneous.

Finally, surrogate models can effectively combine simulated data with experimental data. Experimental data can be used to train surrogate models to be more accurate and realistic, while simulated data can be incorporated to reduce the requirement for many expensive experiments. Pure physics-based models are often also dependent on assumptions and will not be correct when these assumptions are not satisfied. By training surrogate models also on experimental data, we can improve performance in these regions.

In short, surrogate models unlock modelling possibilities with exceptional speed and combine knowledge from simulations and experimental data effectively. At Sioux Technologies, we have repeatedly encountered these motivations across diverse industrial projects. In our experience, it is typically beneficial to go beyond a purely data-driven approach. By incorporating domain and engineering knowledge, hybrid models can be created that adhere to the relevant laws of physics, use less data, and are more interpretable. This branch of surrogate modelling is known as physics-informed surrogate modelling. See also Figure 1 for the positioning of physics-informed machine learning in relation to physics-driven and data driven

Integrating domain knowledge in surrogate models

Surrogate modelling is fundamentally about constructing fast, approximate models that emulate the behaviour of complex physical systems. While traditional surrogate models rely solely on data, PISMs incorporate explicit knowledge of the underlying physics; see Table 1. This integration of physics and data-driven learning is what sets them apart, enabling superior generalisation, robustness and data-efficiency.

A PISM is designed to respect the governing equations, boundary conditions and conservation laws relevant to the system under study. This can be achieved in various ways, such as embedding physics knowledge directly into the model architecture, the loss function or the training process. The result is a model that not only fits the available data but also extrapolates reliably to new scenarios, reducing the risk of unphysical predictions – a critical requirement in high-stakes engineering applications. In contrast, purely data-driven models (such as standard neural networks or regression models) may achieve high accuracy within the training domain but often fail to generalise when confronted with unseen conditions or limited data.

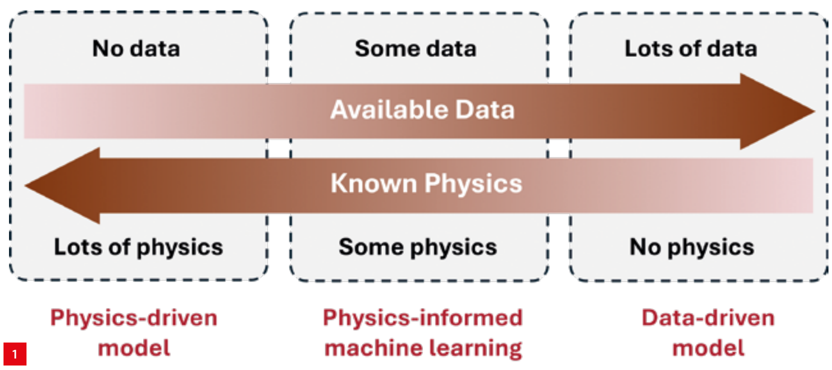
Table 1
Comparison between surrogate models and physics-informed surrogate models (PISMs).

	Surrogate model	PISM
Knowledge source	Purely data-driven	Data and physical laws
Data requirement	High	Lower (physics helps)
Physical consistency	Not guaranteed	Enforced by design
Use case	Black-box optimisation	Scientific/engineering problems with known physics

Physics-informed surrogate models - terminology

Within this field, we find a lot of terms being used interchangeably without any strict definition. The field that combines scientific computing with machine learning is often referred to as SciML (scientific machine learning) or scientific AI, but other terms such as data-driven scientific computing are also prevalent. If you combine machine learning with physical models and insights, you will also find the field being referred to as PIML (physics-informed [or physics-inspired] machine learning). Sometimes you also find the term physics-informed AI, or physics-informed (surrogate) modelling being used. If you combine different data sources (such as experiments and simulations), you often also find the terms hybrid modelling or multi-fidelity modelling.

This article focuses on the explicit integration of physics into the machine-learning process, so we prefer the term physics-informed machine learning, or physics-informed (surrogate) modelling as we feel it aligns most closely with what we are doing. See also Figure 1.



1 *Overview of fully-physics, fully-data-driven and physics-informed machine-learning paradigms. By combining known physics with small amounts of data, benefits from both sides can be combined. (Adapted from Karniadakis, G.E., et al., "Physics-informed machine learning", Nature Reviews Physics, vol. 3, pp. 422-440, 2021).*

Model types and embedding physical knowledge

Physics-informed surrogate modelling encompasses a range of model types; each suited to different problem domains and data regimes. Here, we provide an overview of the three we most often see in practice: Gaussian processes, reduced-order models and neural networks.

Although these techniques seem relatively disparate, it is not uncommon to see several different methods being applied concurrently for optimal performance. For example, we have seen cases where a neural network has been applied as a reduced-order model, before the values were fed into a Gaussian process to do predictions with uncertainty quantification.

- Gaussian processes (GPs):
  GPs are probabilistic models that provide uncertainty quantification and are highly effective in low-data regimes and/or with noisy data. Physics can be incorporated via kernel engineering, where the covariance function reflects domain-specific interactions or symmetries. To keep the computational burden manageable, the input typically needs to be relatively low-dimensional.
- Reduced-order models (ROMs) :
  These models simplify high-dimensional systems by projecting them onto lower-dimensional manifolds, often using techniques like proper orthogonal decomposition (POD) or dynamic mode decomposition (DMD). ROMs are particularly valuable in control and real-time applications, where computational speed is paramount. Not all problems are well suited for ROMs – nonlinear systems are often problematic, as are problems that are difficult to project to a low-dimensional space using linear transformations. In those cases, neural networks are often the most viable alternative.
- Neural networks:
  Neural networks excel at modelling complex, high-dimensional and nonlinear systems, and are particularly effective for inverse problems. Recent advances make it possible to incorporate physical knowledge such as laws, constraints or symmetries directly into the model through specialised architectures, physics-informed loss functions, and domain-specific regularisation. This makes them powerful tools for surrogate modelling in scientific and engineering applications.

Case A
Combining 3D FEM simulations with lumped-system modelling using ROMs

Context and challenge

Lumped-system (0D, zero-dimensional) models allow quick scenario analysis for complex engineering systems, helping verify their KPIs (key performance indicators) after design changes. However, due to simplified nature they can lack accuracy, requiring detailed 3D FEM/CFD simulations for certain components – significantly increasing simulation and development time.

Objective

Develop fast, accurate parametric ROMs of thermal FEM (finite-element) models and integrate them into the system model. This is difficult because commercial solvers typically restrict access to the full parametric model.

Approach

Using available solver data – such as inputs, outputs, and fixed-parameter matrices – the team reconstructed the parametric full-order model. Classical ROM techniques like proper orthogonal decomposition (POD) were then applied to build a parametric reduced-order model, following a 'grey-box' approach that combines partial physics knowledge with data-driven methods. Finally, the ROM was integrated into the system-level model to enable parameter optimisation.

Impact

ROM-based optimisation achieved orders-of-magnitude speed-up versus full-order models while preserving accuracy. Projection-based methods ensured physics fidelity, reducing development cycles and simplifying parameter changes within the system model instead of complex FEM software.

Broader significance

This approach enables detailed physics in system models, making virtual prototyping, optimisation, and digital-twin deployment far more efficient.

Data

Although data is not always a requirement – particularly for approaches like PINNs (physics-informed neural networks), which can leverage embedded physics to achieve good performance with much less data than purely data-driven models – it is often beneficial to help train your physics-informed surrogate model. It is important to think carefully about the data you require and what is available. Data typically originates from the following two sources:

1. Simulation data:
   High-fidelity simulations provide rich datasets for training but may be limited in coverage due to computational cost. Lower-fidelity simulations may also be available, but of limited use because of the approximations involved.
2. Experimental data:
   Although often noisy or sparse, they both validate model predictions and correct systematic biases and errors in your simulation data. This does not necessarily have to be ground-truth data but can also be the result of indirect sensor measurements that nonetheless provide information about the underlying system.

If we utilise both simulation data and experimental data, we often speak of hybrid modelling. More generally, we speak of multi-fidelity modelling when we use multiple data sources with differing levels of accuracy. Combining simulated and experimental data enhances robustness, especially in regimes where simulations are imperfect or data is scarce. Sioux Technologies has implemented hybrid models in various industry cases, fusing synthetic and measured data to improve generalisation.

Concluding remarks

As you might have gathered from these basics, physics-informed surrogate modelling is not a one-size-fits-all solution, but a flexible methodology that adapts to the unique challenges of each engineering domain. By combining the strengths of data-driven learning and physical insight, these models are redefining what is possible in real-time decision-making, generative design, and autonomous system control.

Finally, it is important to note that the field of physics-informed machine learning (PIML) extends well beyond surrogate modelling. In addition to accelerating forward simulations, it can support the discovery of governing equations, reduce complex physics models to interpretable analytical forms, and enable inverse modelling or parameter identification. Beyond these, PIML is increasingly applied in areas such as uncertainty quantification, sensor fusion, as well as physics-informed control and optimisation, making it a versatile toolkit for bridging physics and data-driven approaches.

Neural networks for physics-informed modelling

Design

The choice of neural-network design depends on the problem, but recent advances show strong performance for differential equations by approximating underlying operators. Deep operator networks (DeepONets) learn mappings between function spaces, enabling rapid evaluation of solution operators for complex systems. Fourier neural operators (FNOs) capture global spatial correlations efficiently by learning operators in the frequency domain, making them ideal for spatio-temporal phenomena.

Loss functions and regularisation

Physics-informed neural networks (PINNs) embed governing equations and boundary conditions directly into the loss function, ensuring predictions respect physical laws. PINNs often require little or no data and are widely used for solving partial differential equations or inverse problems. Alternatively, physics-based penalties – such as conservation of energy, mass or momentum – can guide models towards physically plausible solutions. Domain-specific regularisation prevents overfitting and enforces symmetries or invariances through architecture choices, loss terms, or pre/post-processing. A practical approach is adding a lightweight physical model in post-processing to reduce the neural network's predictive burden.

Case B
Accelerating thermal deformation modelling with physics-informed surrogates

Context and challenge

In precision systems, thermal effects can cause subtle yet critical component deformations. Predicting these traditionally requires iterative physics-based simulations – computationally expensive and too slow for real-time or design-iteration workflows.

Objective

Replace heavy numerical solvers with fast, accurate surrogate models that capture complex thermo-mechanical interactions without losing physical fidelity.

Approach

Using physics-informed machine learning (PIML), the team built convolutional neural network (CNN) surrogates that were over 100x faster than conventional solvers while maintaining less than 1% error. Physical constraints were embedded directly into the network architecture to ensure consistency with governing equations, and physics-based post-processing was applied to further improve accuracy. A rigorous validation framework guided development from initial feasibility studies to high-resolution temporal predictions.

Impact

Models achieved orders-of-magnitude speed-up while maintaining tight accuracy, enabling near real-time predictions and advanced optimisation or control strategies.

Broader significance

This case shows how PIML bridges first-principles and data-driven approaches, delivering speed without sacrificing accuracy – informing similar initiatives across domains.

Tools and frameworks for physics-informed surrogate modelling

Physics-informed surrogate modelling relies on a strong ecosystem of frameworks. Nvidia's PhysicsNeMo is a leading open-source solution optimised for GPU acceleration, offering end-to-end pipelines for geometry ingestion, governing equation definition, and real-time predictions. It supports advanced architectures such as PINNs, neural operators, and graph neural networks, and is widely used in digital-twin, CFD, structural mechanics, and electromagnetics applications.

General-purpose frameworks like PyTorch and TensorFlow remain essential for custom architectures, including PINNs, DeepONets, and Fourier neural operators, while GPyTorch enables scalable Gaussian process regression for low-data regimes and uncertainty quantification. Specialised libraries such as DeepXDE and Julia's NeuralPDE.jl target differential equation solvers.

Commercial platforms like Comsol and Ansys increasingly integrate AI-driven surrogate modelling and interoperate with frameworks like PhysicsNeMo and PyTorch. Web-based tools such as Siml.ai further simplify adoption through symbolic equation input and rapid deployment.

When selecting a framework, prioritise advanced architectures, GPU acceleration for scalability, and integration with existing simulation environments. End-to-end pipelines for training, validation and deployment streamline workflows, while strong community support ensures extensibility and problem-solving.

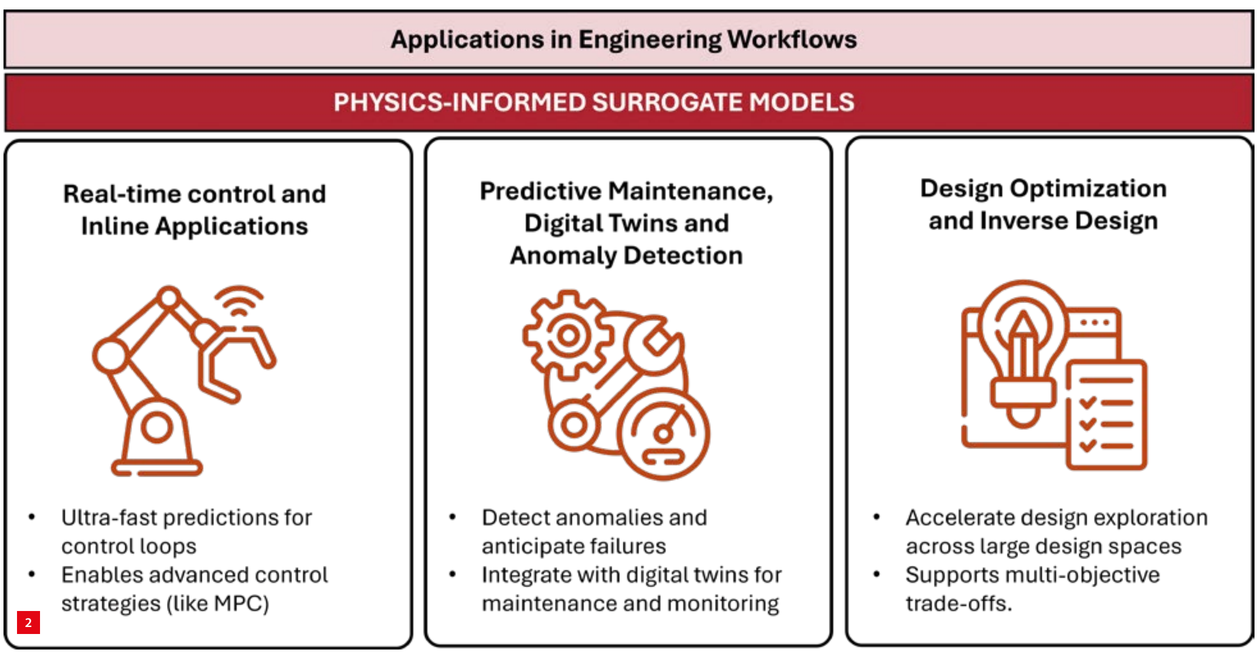
Applications in engineering workflows

Physics-informed surrogate models (PISMs) are reshaping engineering workflows by enabling fast, reliable predictions and optimisations across a spectrum of real-world scenarios. Their adoption is accelerating in various high-tech domains where Sioux Technologies has seen significant impact. We now describe three types of applications; see Figure 2.

Real-time control and inline applications

In advanced manufacturing and precision engineering, control loops demand rapid, trustworthy feedback to maintain optimal system performance. Traditional physics-based models, while accurate, are often too slow for real-time use, especially when dealing with nonlinear dynamics or high-dimensional systems. PISMs, trained to emulate these models, deliver predictions in milliseconds.

Often, in high-tech applications very high precision is required, meaning that the system needs to correct for all effects induced by physical interactions, such as vibrations, temperature changes, and chemical and electrical effects. Full-scale physical models to evaluate these effects take far too long to be used in a real-time control loop. By incorporating domain knowledge and training the surrogate models with a loss based on the physical laws of these processes, it is ensured that the surrogate models generalise well and the amount of data that is necessary for training the models is significantly reduced. The surrogate models increase processing speeds by multiple orders of magnitude, meaning that we can accurately predict the effects and correct for them in real-time.



2

Overview of main application areas for physics-informed surrogate models.

Having prediction models that run in real-time, or even faster, also allows for more complex control algorithms. In model predictive control (MPC), control decisions are based on a prediction of the future state of the system. This allows the control systems to adjust settings proactively. Of course, this is only possible when these predictions can be made fast enough.

Predictive maintenance, digital twins and anomaly detection
Predictive maintenance focuses on anticipating failures before they occur, reducing downtime and optimising asset lifecycles. PISMs offer a new and powerful solution by combining physics and data. For predictive maintenance, this is crucial because failure mechanisms often depend on complex interactions such as thermal stresses, fatigue, and fluid dynamics: phenomena that are hard to capture reliably by black-box models alone. Also, real failure data is typically rare and expensive to model or simulate.

PISMs can be integrated into digital twins, creating predictive platforms that simulate future states of equipment under different usage patterns. This allows maintenance strategies to shift from reactive or scheduled interventions to proactive, risk-based decision making.

Beyond prediction, these models also enhance anomaly detection. By embedding physical constraints and system dynamics into the surrogate, deviations from expected behaviour can be identified more accurately and earlier than with purely statistical or machine-learning approaches. This is particularly valuable for detecting subtle anomalies that precede critical failures, such as abnormal vibration signatures or unexpected thermal gradients. Combining anomaly detection with predictive simulation creates a comprehensive health-monitoring framework, allowing operators to not only forecast failures but also intervene at the earliest signs of abnormality and as such maximising reliability and safety across the asset lifecycle.

Case C
Optimising material formulation with physics-informed surrogates

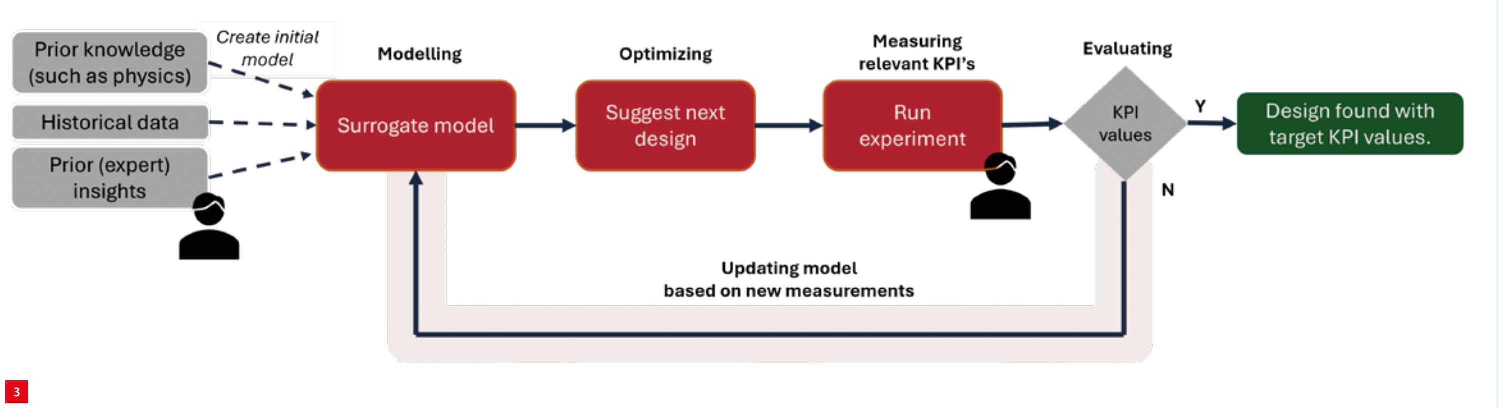
Context and challenge
In advanced material design, achieving precise colour properties in complex formulations is critical. Traditional development relies on iterative physical testing – costly and time-consuming, often requiring numerous experiments.

Objective
Build a modelling framework to predict and optimise formulation outcomes accurately while minimising physical iterations. The challenge was capturing nonlinear interactions between multiple components and process parameters without losing physical plausibility.

Approach
The team developed a surrogate modelling framework using Gaussian process regression (GPR) with automated kernel selection. Physics-informed kernels reflected pigment dispersion and optical behaviour, while prior knowledge (e.g., reflectance spectra) improved generalisation in low-data regimes. Bayesian optimisation guided exploration, using uncertainty estimates to prioritise experiments and reduce trial & error.

Impact
Development cycles dropped from 10-15 iterations to 4-8, cutting material waste, costs, and time-to-market. Surrogates provided actionable insights for recipe optimisation, enabling rapid convergence on target properties.

Broader significance
This approach shows how PISMs merge first-principles with data-driven efficiency. It evolved into a generic recipe-builder framework (see Figure 3), combining Bayesian optimisation and surrogates to suggest next-best experiments – applicable to material design, machine configuration, or any domain requiring fast recipe discovery.



3
Sioux recipe-builder framework that combines (physics-informed) surrogate models and optimisation to optimise 'recipes' such as designs, machine settings, and process settings.

Design optimisation and inverse design
Generative design optimisation explores vast design spaces to identify innovative solutions that meet multiple objectives and constraints. This process often involves thousands or even millions of design evaluations, which is computationally prohibitive when relying solely on high-fidelity simulations such as CFD or FEM. Surrogate models provide an efficient alternative by approximating these expensive simulations, enabling rapid evaluation of candidate designs and accelerating the optimisation process.

In generative design workflows, surrogate models act as the backbone for iterative exploration. Once trained on representative simulation data, they predict performance metrics with high speed and accuracy, enabling optimisation algorithms – such as genetic algorithms, Bayesian optimisation, or gradient-based methods – to evaluate large design populations without costly full-scale simulations.

Two main strategies exist: forward surrogates combined with optimisers, which approximate the physics solver and then search the design space, and direct inverse models, which learn the mapping from performance targets directly to design parameters. Neural-network-based surrogates naturally support gradient tracking, making gradient-based optimisation practical, which is something rarely feasible with full CFD or FEM solvers. Together, these approaches allow engineers to explore unconventional geometries and complex multi-objective trade-offs within feasible timelines.

A critical enhancement in this context is also multi-fidelity modelling. Instead of relying on a single surrogate model trained from high-fidelity data, multi-fidelity approaches combine information from models of varying accuracy and cost. For example, low-fidelity simulations can provide broad coverage of the design space, while high-fidelity simulations refine predictions in critical regions. This hierarchical approach significantly reduces the number of expensive evaluations required while maintaining confidence in the optimisation results.

Lessons learned and best practices

The development and deployment of physics-informed surrogate models (PISMs) in industrial contexts have yielded various practical insights. While each application presents unique challenges, several recurring themes and best practices have emerged from Sioux Technologies' experience and broader industry trends.

Iterative development
In early feasibility studies, focus on simplified models. Proceed by gradually increasing complexity as confidence grows. It is tempting to jump to the final problem immediately, but we often find most technical challenges and physics integration can already be tackled early on.

PISMs can achieve enormous speed-ups, while preserving accuracy
Surrogate models have replaced computationally intensive solvers, often delivering speed-ups of 100x or more while maintaining sub-percent error rates.

Add experimental to simulated data
Hybrid modelling approaches, combining simulation and experimental data, have improved model robustness and reduced the need for costly physical iterations.

Importance of domain knowledge and physics priors
A key lesson is the critical role of domain expertise in model development. PISMs are most effective when they incorporate relevant physical laws, boundary conditions, and expert knowledge. Collaborating closely with domain specialists ensures that models are grounded in reality and tailored to the specific nuances of the application. This approach not only improves accuracy and generalisation but also enhances trust and interpretability. This is essential for adoption in safety-critical and high-value environments.

Example
Understanding the thermal expansion behaviour of materials (case B) under varying exposure conditions was crucial to defining the right boundary conditions and selecting the appropriate surrogate architecture. This kind of case-specific engineering requires a flexible modelling toolbox.

Trade-offs: fidelity, speed and interpretability
Balancing model fidelity, computational speed, and interpretability is a central challenge. High-fidelity models may offer superior accuracy but can be slower and harder to deploy in real-time workflows. Conversely, simpler models may sacrifice detail for speed and ease of integration. Hybrid approaches – so combining physics-based constraints with data-driven flexibility – often provide the best compromise, delivering robust performance without excessive computational overhead. Usually, the specific constraints and requirements of the case at hand inform us of the optimal trade-off and best modelling approach to take.

Example
Think of a neural network surrogate trained on simulated optical profiles. This can be enhanced with embedded physical priors to preserve wavelength-dependent behaviour, while pruning and quantisation techniques can be applied to meet latency constraints for inline usage. The optimal trade-off depends on the specific operational context and requires engineering judgement.

Common pitfalls in model training and deployment
Several pitfalls can hinder the success of surrogate modelling projects:

- Overfitting simulation data:
  Models trained exclusively on synthetic data may fail to generalise to real-world scenarios. Incorporating experimental and sensor data helps correct systematic biases and improves robustness.

Example
Think of a hybrid modelling approach where surrogate models are initially trained on large volumes of simulated data, but only achieve the required accuracy after being supplemented with a limited set of high-quality experimental measurements.

- Ignoring operational variability:
  Real-world systems are subject to noise, drift, and changing conditions. Validating models with operational sensor data and designing for adaptability is crucial.

Example
Early surrogate models suffered from specific errors that became apparent when validating with operational data. From here we were able to improve robustness by expanding the training set with various other samples and explicitly modelling additional effects. These steps reduced the mean relative error substantially.

- Underestimating integration complexity:
  Embedding surrogate models into existing engineering tool chains requires careful planning, standardised interfaces, and ongoing collaboration between modelling and software teams.

Deployment recommendations
Successful deployment of PISMs depends on several practical considerations:

- Hardware and latency:
  Assess computational requirements early on. For real-time applications, ensure that surrogate models can deliver predictions within required latency constraints, leveraging GPU acceleration or edge computing where appropriate.

Example
For latency-critical applications - such as those with millisecond response budgets - models often need to run on embedded platforms like Nvidia Jetson. Neural networks trained in Python or TensorFlow can be converted to optimised formats (e.g., ONNX or TensorRT) to minimise inference time and memory use. In several projects, deploying models directly on edge devices near the machine has eliminated cloud round trips and ensured deterministic response times for control and monitoring tasks.

- Tool-chain integration:
  Select frameworks and platforms that support seamless integration with existing simulation environments, control systems, and data pipelines. Open standards and APIs facilitate interoperability and future scalability.

Example
Physics-informed models can be embedded into simulation tools like Comsol using Python scripting or exported as functional mock-up units (FMUs) for co-simulation.

- Maintainability and lifecycle management:
  Establish processes for model versioning, retraining, and validation. Monitor model performance over time and update as system behaviour or requirements evolve. Sioux Technologies emphasises the value of continuous validation to sustain long-term success.

Example
In practice, this means tracking model versions alongside simulation baselines, automating retraining when new sensor or operational data becomes available, and validating against both synthetic and real-world benchmarks. To support this, MLOps tools such as MLflow or DVC can be used to manage model artifacts, monitor drift, and trigger updates when performance thresholds are breached. Embedding these practices into the engineering workflow ensures that surrogate models remain reliable, interpretable and aligned with evolving system dynamics.

Sioux Technologies' perspective
Through multiple projects in various high-tech domains such as semiconductors and materials, Sioux Technologies has observed that the most successful surrogate modelling initiatives are those that:
- prioritise collaboration between modelling specialists, domain experts, and operational stakeholders;
- embrace hybrid modelling strategies, leveraging both simulation and real-world data;
- invest in robust validation, monitoring, and lifecycle management practices;
- remain adaptable, refining models and workflows as new data and requirements emerge.

While challenges remain – particularly in adapting surrogate models to new domains – the key takeaway is that successful deployment is highly case-specific. Each industrial application comes with its own physical constraints, data availability, and performance requirements, which means there is no one-size-fits-all solution. Engineering the right surrogate model often requires selecting from a diverse modelling toolbox: from reduced-order physics solvers and neural networks with embedded priors, to hybrid architectures that combine simulation data with real-world measurements.
Sioux Technologies has found that tailoring the modelling strategy to the specific operational context – rather than relying on generic templates – is essential for achieving robustness, interpretability, and stakeholder trust.

Integration with GenAI and agentic systems

As physics-informed surrogate modelling (PISM) continues to evolve, new possibilities are emerging for its integration with generative AI (GenAI) and agentic systems; see Listing 1. These technologies may offer novel ways to interact with, deploy, and extend the capabilities of surrogate models – potentially transforming how engineers and operators engage with complex systems. While many of these ideas are still speculative, they point towards promising directions for future development and experimentation.

Listing 1
Emerging possibilities for integrating PISM with generative AI (GenAI) and agentic systems.

Automated design iterations
AI agents could assist engineers in exploring design spaces by invoking PISMs as fast evaluators. This might enable rapid assessment of candidate designs, helping reduce the time and cost of optimisation cycles, especially in domains where full simulations are computationally intensive.

Surrogate model-oriented workflow support
Agentic systems may help manage workflows involving PISMs by selecting appropriate models, configuring inputs based on contextual data, and interpreting outputs. This could streamline the use of surrogate models in iterative design and operational decision-making.

Real-time operational support
PISMs could be deployed in real-time environments – such as manufacturing, robotics, or energy systems – where agents use surrogate models to evaluate system behaviour and suggest control actions. This might enable adaptive strategies that respond to changing conditions without the latency of full simulations.

Human-in-the-loop design exploration
Looking ahead, conversational interfaces powered by large language models (LLMs) could enable engineers to interact directly with physics-informed surrogate models in real time. In such settings, PISM would serve as a fast, physics-grounded tool that agents or assistants invoke to answer "what if" questions, explore trade-offs, and provide immediate feedback on design changes. This could support more intuitive and collaborative design workflows, where engineers remain actively involved in decision-making while benefitting from rapid model evaluations and contextual explanations.

Model-driven monitoring and alerting
Agents could use PISMs to predict expected system behaviour and compare it with live sensor data. Deviations might trigger alerts or diagnostics, supporting predictive maintenance and anomaly detection in complex systems.

Sioux Technologies' perspective
At Sioux Technologies, we have started thinking about how PISM might be enhanced through emerging GenAI and agentic technologies. While we do not yet have fully integrated workflows in place, we are tracking these developments and exploring how Sioux can leverage them to support more adaptive, efficient and accessible engineering solutions.
Our focus is on understanding where these technologies are heading and how they can be applied in practice. We are particularly interested in exploring how agent-driven optimisation can accelerate design cycles, how adaptive control strategies can be enhanced through the agentic use of surrogate models, and how natural-language interfaces can foster more intuitive collaboration between domain experts and AI systems.

As GenAI and agentic technologies continue to evolve, their potential to enhance the use of PISMs will depend on several practical aspects. In our view, ensuring model interoperability, maintaining data quality and governance, and supporting human oversight will be key to realising meaningful applications. Beyond these, opportunities such as sensor fusion for richer state estimation and physics-informed control for complex systems highlight the versatility of this approach. While much remains to be explored, the combination of PISM with GenAI and agentic systems offers a promising direction for future engineering tools, and Sioux Technologies is keen to monitor and adopt these innovations as part of its broader innovation efforts.

Conclusion and outlook

Physics-informed surrogate modelling (PISM) is rapidly becoming a cornerstone of modern engineering practice. By fusing domain knowledge, physical laws, and data-driven learning, these models offer a compelling solution to the challenges of speed, accuracy and adaptability in high-tech domains.

Applications ranging from real-time digital twins and adaptive control systems to generative design and multi-objective optimisation are already demonstrating the potential of PISM across various industrial domains such as semiconductors, materials, and precision systems.

Our lessons learned from working on industrial applications underscore the importance of using domain knowledge and leveraging physics priors for improved generalisation. It also is important to prioritise the integration of simulation and experimental data allowing for hybrid modelling strategies and robust validation. We also learned that it pays off to invest in maintainable, scalable workflows. As the ecosystem of tools and frameworks continues to expand, engineers are increasingly empowered to embed surrogate models into existing pipelines, unlocking new efficiencies and capabilities.

Looking forward, PISM will continue to impact engineering innovation. Surrogate models are expected to increasingly be integrated into autonomous design workflows, enabling faster iteration and more informed decision-making. Advances in model speed and interoperability are making real-time digital twins more practical, allowing systems to adapt dynamically to changing conditions. At the same time, hybrid physics-AI frameworks are emerging that combine interpretability and scalability, making them suitable for embedded and edge applications.

In terms of the models themselves, we expect significant progress in several areas:

- Neural operators and advanced architectures:
  Architectures such as graph-based neural networks, transformers for partial differential equations, and advanced neural operators will improve generalisation across physics problems and reduce retraining needs.
- Foundation models for physics:
  Large-scale pretrained models that capture broad physical principles and can be fine-tuned for specific domains are emerging, enabling faster deployment and cross-domain adaptability.
- Multi-fidelity and multi-scale approaches:
  Seamlessly combining coarse and fine-grained models for efficiency and accuracy.
- Uncertainty-aware surrogate models:
  Incorporating probabilistic reasoning to quantify confidence and propagate uncertainty through workflows.
- Symbolic and interpretable surrogates:
  Moving beyond black-box predictions towards human-readable physics insights.
- Integration with streaming data:
  Supporting real-time adaptation and online learning in dynamic environments.

These developments signal a shift towards more intelligent, responsive and efficient engineering systems. As these trends accelerate, the role of physics-informed surrogate modelling will only grow in importance. Sioux Technologies is committed to tracking these developments and translating them into practical, industry-ready solutions.