

Physics-Informed Autonomous LLM Agents for Explainable Power Electronics Modulation Design

Junhua Liu^{1*}, Fanfan Lin^{2*}, Xinze Li^{3†}, Shuai Zhao⁴, Kwan Hui Lim^{5,1}

¹Forth AI, Singapore

²Zhejiang University - University of Illinois Urbana-Champaign Institute, China

³University of Arkansas, USA

⁴Aalborg University, Denmark

⁵Singapore University of Technology and Design, Singapore

j@forth.ai, fanfanlin@intl.zju.edu.cn, xinzel@uark.edu, szh@energy.aau.dk, kwanhui@acm.org

Abstract

LLM-based autonomous agents have recently shown strong capabilities in solving complex industrial design tasks. However, in domains aiming for carbon neutrality and high-performance renewable energy systems, current AI-assisted design automation methods face critical challenges in explainability, scalability, and practical usability. To address these limitations, we introduce Physics-Informed Autonomous Agent (PHIA), an LLM-driven system that automates modulation design for power converters in Power Electronics Systems with minimal human intervention. In contrast to traditional pipeline-based methods, PHIA incorporates an LLM-based planning module that interactively acquires and verifies design requirements via a user-friendly chat interface. This planner collaborates with physics-informed simulation and optimization components to autonomously generate and iteratively refine modulation designs. The interactive interface also supports interpretability by providing textual explanations and visual outputs throughout the design process. Experimental results show that PHIA reduces standard mean absolute error by 63.2% compared to the second-best benchmark and accelerates the overall design process by over 33 times. A user study involving 20 domain experts further confirms PHIA's superior design efficiency and usability, highlighting its potential to transform industrial design workflows in power electronics.

Introduction

As global temperatures rise due to climate change, transitioning power generation systems toward carbon neutrality is increasingly urgent. A key component of this transition is the integration of renewable energy sources (RES) into the power grid. Power Electronics Systems (PES) are critical to this effort for three primary reasons (Lin 2022). Firstly, power converters within PES transform the direct current (DC) output from solar panels or wind turbines to alternating current (AC), facilitating the seamless integration of RES into the AC grid. Secondly, PES manages power during peak and off-peak periods, guiding the charge and discharge

processes of energy storage units. Lastly, it upholds grid stability through functions like voltage regulation and reactive power compensation. Achieving the desired outcomes necessitates meticulous modulation of power converters in PES, which involves controlling their switching to manage output voltage or current.

The escalating adoption of RES and the burgeoning scale of multi-resource power systems complicate the modulation design for power converters within PES. The diversity of energy sources linked to the power grid intensifies system complexity. As more converter subsystems interlink, the modulation design problem's dimensionality surges. Consequently, identifying an optimized design solution through manual computation becomes infeasible given this complexity. As cities worldwide embrace multi-resource power systems, modulation designs must be customized, catering to particular geographies and application objectives. Oftentimes, these objectives can conflict, further complicating matters.

The recent advancements in artificial intelligence (AI) and Large Language Models (LLMs) offer a promising solution to these modulation design challenges through automation. Emerging research has introduced AI-centric modulation design techniques. For instance, recent works explore the use of an LLM with in-context learning for modulation design (Lin et al. 2024). Extreme gradient boosting algorithm (XGBoost) was utilised to construct the surrogate model for the triple phase shift modulation strategy for the dual active bridge (DAB) converter using training data sourced from simulation tools or hardware prototypes (Lin et al. 2023a). Subsequently, a differential evolution algorithm collaborates iteratively with XGBoost until optimal modulation parameters are identified. Likewise, the Q-learning algorithm, a conventional reinforcement learning technique, can be trained offline to derive optimized modulation parameters (Tang et al. 2020). While innovative, these methods suffer from several drawbacks:

1. The training of these AI models is often data-intensive, with significant demands for an extended and tedious data collection through extensive simulations or hardware experiments.
2. For complex tasks, training or finetuning large models

*These authors contributed equally.

†Corresponding author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

are also computationally intensive and consume a large amount of energy to train and serve.

3. Deploying AI models as unexplainable black boxes severely restricts their industrial adoption.
4. The current techniques focus on specific modulation strategies or pre-established design goals, which limits the scalability of such automation.
5. Existing methods require extensive human involvement in the whole design process which makes the design process inefficient.

In recent years, LLMs, such as GPT-4 (OpenAI 2023), Palm (Anil et al. 2023) and LLaMa (Touvron et al. 2023) demonstrates superior performance in natural language understanding and generation. Furthermore, recent works introduce LLM-based autonomous agents which extend LLM's capacity from simply performing reasoning and generating content to actions and control (Wang et al. 2024).

In our work, we investigate the potential of using LLM-based autonomous agents for downstream task automation in Power Electronics Systems. Specifically, we introduce **PHIA**, a Physics-Informed Autonomous Agent for power converter modulation design, enabling users to produce high quality modulation designs with minimal human supervision.

PHIA first uses a LLM-based planner to process user requirements and generate a set of design specifications. Subsequently, the agent invokes a set of system design tools that consist of physics-informed surrogate models and optimization algorithms. In the design process, the tools iteratively derives optimal modulation parameters tailored to users' design specifications. Finally, the optimal design parameters and performance metrics will be returned to and visualised for the users to achieve an explainable design process.

The contributions of this paper are as follows:

- We propose PHIA, a physics-informed autonomous agent that streamlines the power converter modulation design process. Comprehensive experiments show the strong performance of PHIA, statistically outperforming the best baseline model with 63.2% lower error for low-data scenario, and 23.7% lower error for high-data scenario.
- We propose a hierarchical physics-informed surrogate model as a design automation tool for power converter modulation, which outperforms benchmarks in accurately predicting power converter performance even in extreme scenarios where the data is sparse.
- Our user study with 20 experts demonstrates that the design time with PHIA is over 33 times faster than conventional methods, showing superior efficiency in empirical practice with minimal human supervision.

Problem Statements

Power Converter Phase Shift Modulation

Within the renewable energy systems sector, power converters need to work in coordination with other components to

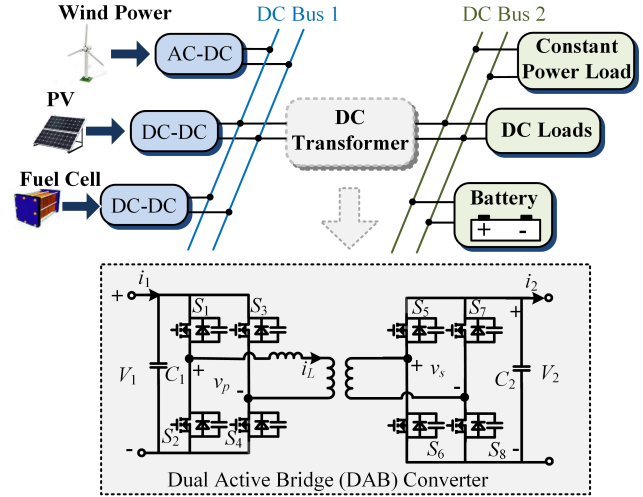


Figure 1: An example of power converter application: the DAB (Dual Active Bridge) converter serves as a DC transformer of the DC-DC power grid, offering galvanic isolation and regulating power and voltage between DC buses. Modulating its switches directly affects the system operating performance, including power transfer efficiency, voltage regulation, and stability of the interconnected buses.

ensure optimal functionality. Figure 1 depicts a classic application of a particular power converter, the dual active bridge (DAB) converter. This converter functions as a DC transformer, connecting various DC buses to facilitate the utilization or storage of power produced by RES (Lin, Zhang, and Li 2021). The phase shift modulation technique is employed to regulate the current flow and guarantee efficient power transmission. This is done by adjusting the timing of different power switch sets, represented as S_1 to S_8 in Figure 1. Table 1 details the distinct phase shift modulation strategies associated with the DAB converter, highlighting their respective degrees of freedom (DoF) and modifiable parameters (Hou and Li 2019). It is worth noting that the adjustable parameter for $\langle S_1, S_5 \rangle$ indicates the phase difference between S_1 and S_5 .

For designing the parameters of the modulation strategy, users are required to define three essential pieces of specification consistent with the application prerequisites, namely:

1. Expected operating conditions for the converter: This involves specifying the rated power, rated input voltage (current) and rated output voltage (current), as well as their actual operating values.
2. Selected modulation strategy: The modulation strategy selection should factor in both the acceptable level of control complexity and the feasibility of achieving the desired performance for the specific strategy.
3. Performance objectives: Various application scenarios have different priorities for performance objectives, which directly affect the modulation design. Common objectives in practical applications encompass power efficiency, power density, zero voltage switching, zero current switching, current stress, dynamic response, electro-

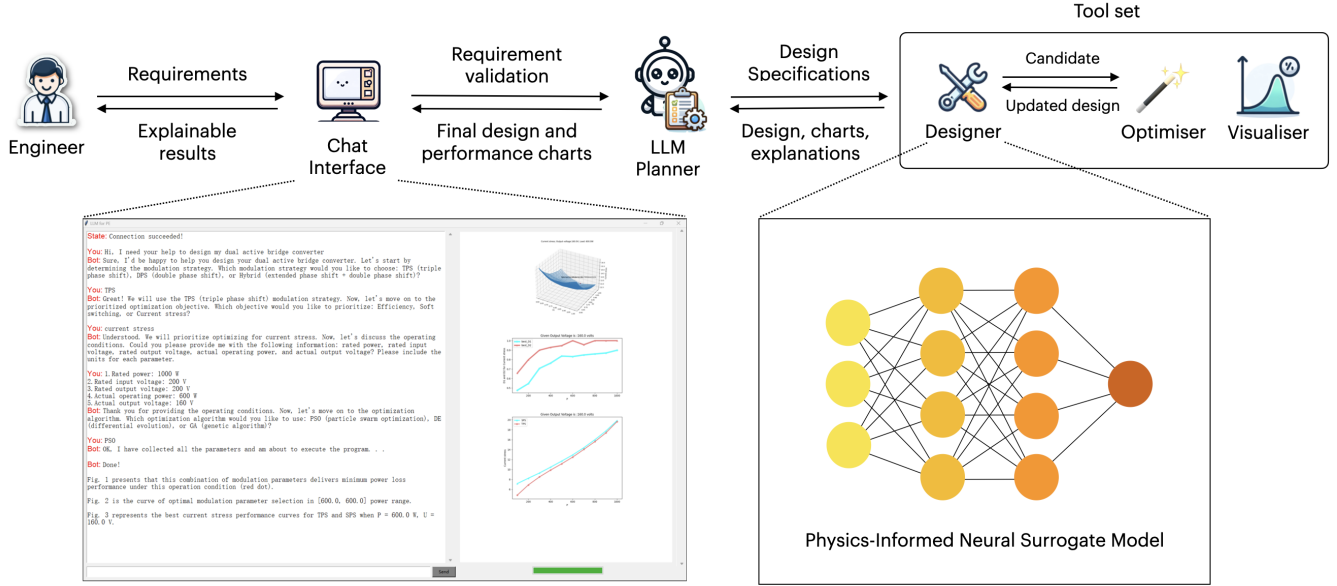


Figure 2: System architecture of PHIA: an engineer provides design requirements to PHIA via a chat interface connecting to its planner. Once the full requirements are determined, the planner coordinates and invokes tools from the tool set to iteratively generate the modulation design without human supervision. After the design is done, the planner displays the final results and explainable process on the chat interface.

Phase shift Strategies	No. of DoF	$\langle S_1, S_3 \rangle$	$\langle S_1, S_5 \rangle$	$\langle S_5, S_7 \rangle$
Single phase shift	1	0	D_o	0
Double phase shift	2	D_i	D_o	D_i
Extended phase shift 1	2	D_i	D_o	0
Extended phase shift 2	2	0	D_o	D_i
Triple phase shift	3	D_1	D_o	D_2
Hybrid phase shift	2	D_i	D_o	0
	2	0	D_o	D_i
	2	D_i	D_o	D_i

Table 1: Phase shift modulation strategies with the number of degrees of freedom (DoF) from 1 to 3. $\langle S_x, S_y \rangle$ defines the phase shift between S_x and S_y .

magnetic interference, stability, and compatibility with other subsystems, etc.

Problem Statements

Given a set of operating conditions of the power converter (rated power P_r , actual power P_a , rated input voltage V_{1r} , rated output voltage V_{2r} , actual operating output voltage V_{2a}), chosen modulation strategy S and prioritized performance objectives O , the aim is to design modulation parameters for the chosen modulation strategy S to achieve optimal performance for O . This design system is expected to have the following features:

- Automatic design outcome generation, eliminating the need for users to perform complex analysis or computations.
- Precise design outcomes that translate into exceptional operational performance in real-world scenarios.
- Good usability and explainability to accelerate engineering design and scientific discovery.
- Simple setup for the design system, minimizing the demand for extensive data resources.

Related Work

LLM-based Autonomous Agents

The evolution of LLMs has unlocked immense potential for automating various processes in engineering and natural sciences through simple, plain-language queries. Numerous studies have explored LLM capability in reasoning and formal system applications, such as programming copilots (Ross et al. 2023; McNutt et al. 2023), multi-lingual tasks (Xu et al. 2024; Liu et al. 2024b), and mathematical problem-solving (Shakarian et al. 2023; Shi et al. 2024).

To address more complex industrial tasks, recent work proposed the use of LLM-based autonomous agents (LLM-agents), which extends the capability of LLMs (Wang et al. 2024). LLMs are deployed as the planning backbone to understand user request and invoke suitable tools to perform actions (Yao et al. 2024). Besides the planning module, LLM-agents also include other components, such as memory and tools. With advanced techniques such as Retrieval Augmented Generation (RAG) (Liu et al. 2024b; Xu et al. 2024), LLM-agents become powerful tools for solving complex tasks with little to no human instruction.

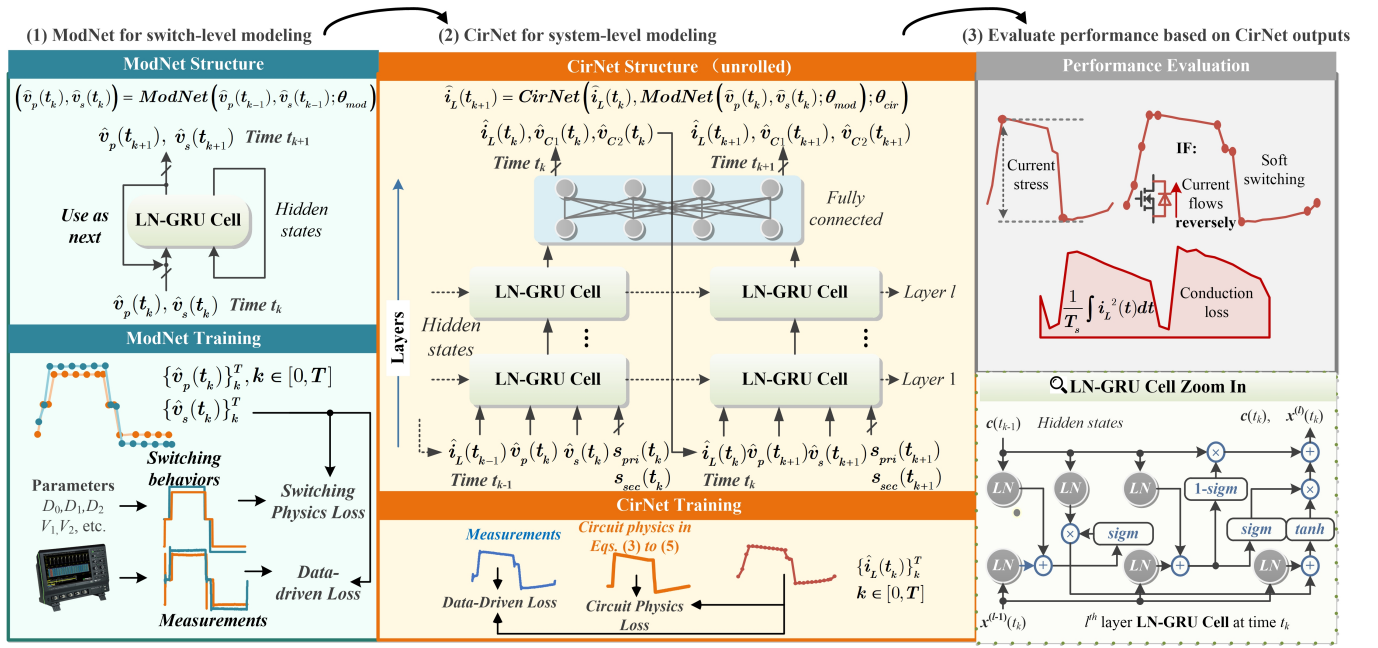


Figure 3: The proposed surrogate model consists of two physics-informed neural networks, namely, **ModNet** for switch-level modeling to learn the switching behaviors, and **CirNet** for system-level modeling to learn the circuit physics. The hierarchical structure enhances the overall accuracy of the power converter’s modeling of the complex behaviors of the switches.

Recent works reported outstanding performance of LLM-agents in industrial applications. For instance, Cao and Lee harnessed LLMs and Phase-Step prompts for cross-domain robot task generation (Cao and Lee 2023). Li et al. introduced an agentic framework to for strategic and interactive decision-making (Li et al. 2024a). LLM-based agents are found effective in other domain-specific tasks, such as spatial-temporal strategic planning (Li et al. 2024b; Liu et al. 2024a; Xiang et al. 2024), e-commerce customer supports (Liu et al. 2024b, 2026, 2025) and financial trading (Zhang et al. 2024b; Koa et al. 2024; Zhang et al. 2024a; Gao, Wang, and Yang 2023).

Physics-Informed Neural Networks

Conventional neural networks are commonly data-hungry and not explainable. By integrating physics principles into neural networks (Raissi 2018), studies on Physics-Informed Neural Networks (PINNs) show promising results in overcoming such challenges.

PINN encompasses various approaches to incorporate physics-informed (PI) components into the deep learning pipeline, such as loss functions, parameter initializations and the neural network architectures (Huang and Wang 2022). As a result, the neural networks are able to learn latent features that are governed by physical laws.

Recent studies also enhance PINN’s convergence, stability, and accuracy, yielding improvements (Kang et al. 2023; Yang, Qiu, and Fu 2023; Wandel et al. 2022). PINN has found success in a wide range of applications. Some examples include predicting traffic conditions and modeling traffic flow (Shi, Mo, and Di 2021; Ji et al. 2022), determining circuit parameters for power converter health mon-

itoring (Zhao et al. 2022), computing power flow in electrical systems (Hu et al. 2020), modeling temperature dynamics in lakes (Jia et al. 2021), managing real-time reservoir gas production (Mudunuru et al. 2020), designing quantum circuits (Liu et al. 2021; Guo et al. 2019), and simulating turbulent flows in urban environments (Xiao et al. 2019). These applications highlight PINNs’ promising performance across many domains by infusing physics knowledge into deep learning.

Methodology

Agentic AI System

Figure 2 presents the architecture of the proposed PHIA agentic AI system. The front-end consists of a chat interface and API interaction with GPT-4 as reasoning engine, whereas the back-end consists of the agent’s tooling, i.e., an optimization algorithm and the PINN-based surrogate model. In terms of workflow, PHIA first collects design specifications from the user and passes to an LLM-based reasoning engine (GPT-4) for planning. Subsequently, the agent invokes the back-end tooling, where the surrogate model produces the performance metrics and passes to the optimization algorithm searches for the optimal modulation parameters. The back-end will respond with a set of optimal design parameters to the front-end to display and visualise as charts for explainability and readability.

Physics-Informed Power Electronics Modeling

Power electronics systems are highly nonlinear induced by the complex switching behaviors of semiconductor devices and the coupling of energy storage components. The generic

representation of a power electronics system formulated by the nonlinear time-variant state-space equations is given in Eq. 1, where θ are the circuit parameters, $u(t)$ and $x(t)$ are the state variables and the input variables, $g(\cdot)$ are general nonlinear functions governing state transition behaviors, and $h(\cdot)$ denotes the input physical laws.

Overall, the proposed surrogate model consists of two hierarchical PINNs, one ModNet for switch-level modeling to learn the switching behaviors and another one CirNet for the system-level modeling to learn the circuit physics. ModNet is trained to learn the intermediate waveforms v_p and v_s and CirNet infers key characteristic waveforms in the circuit including i_L , v_{c1} , v_{c2} , with which the operating performance such as the current stress, soft switching range, and efficiency can be evaluated. As shown in Figure 1, v_p and v_s denote the ac terminal voltages of the primary and secondary full bridges, respectively, and i_L , v_{c1} , and v_{c2} represent the inductor current, input capacitor voltage, and output capacitor voltage, respectively.

$$\frac{\partial u(t)}{\partial t} = g(x(t); \theta)u(t) + h(x(t); \theta) \quad (1)$$

ModNet for Switch-level Modeling of Switching Behaviors: In the modeling of semiconductor switches, the non-ideal equivalent circuit considers the parasitic inductances, capacitances, and resistances. The complicated interactions within circuit components result in nontrivial oscillations and overshoots, which will affect the overall operating performance of the power converter. To address this, the switching behaviors is proposed to be modeled by a physics-informed network, ModNet. ModNet is composed of several layers of gated recurrent units with layer normalization (LN-GRU). The temporal feature of LN-GRU accords well with the task of modulation waveform prediction. As shown in Figure 3 Part (1) and Eq. 2, ModNet infers $v_p(t_k)$ and $v_s(t_k)$ based on the information of the previous timestamp t_{k-1} and the hidden states, and all the intermediate predictions within a switching cycle are stored for training. The training of ModNet utilizes two kinds of losses: (1) the loss based on physics information which helps with switching synchronization; (2) the loss based on experimental data captured by oscilloscope, which helps the model to capture oscillations and overshoots in practical operations.

$$(\hat{v}_p(t_k), \hat{v}_s(t_k)) = \text{ModNet}(\hat{v}_p(t_{k-1}), \hat{v}_s(t_{k-1}); \theta_{mod}) \quad (2)$$

where $\hat{v}_p(t_k)$ and $\hat{v}_s(t_k)$ denote the ModNet predictions for v_p and v_s at time t_k given the previous voltages $\hat{v}_p(t_{k-1})$ and $\hat{v}_s(t_{k-1})$.

$$L \frac{\partial i_L(t)}{\partial t} = -R_L i_L(t) + v_p(t) - n v_s(t) \quad (3)$$

where L , R_L , and n denote features of magnetic components: leakage inductance, equivalent inductor resistance, and turn ratio of transformer, respectively. $i_L(t)$ is the key waveform characterizing main circuit performances such as power transfer, efficiency, thermal behavior, soft switching, electromagnetic interference, etc.

$$C_1 \frac{\partial^2 v_{C1}(t)}{\partial^2 t} = - \left(\frac{\partial s_{pri}(t)}{\partial t} - \frac{s_{pri}(t)R_L}{L} \right) i_L(t) + \frac{\partial i_1(t)}{\partial t} - \frac{s_{pri}(t)}{L} (v_p(t) - n v_s(t)) \quad (4)$$

$$C_2 \frac{\partial^2 v_{C2}(t)}{\partial^2 t} = n \left(\frac{\partial s_{sec}(t)}{\partial t} - \frac{s_{sec}(t)R_L}{L} \right) i_L(t) - \frac{\partial i_2(t)}{\partial t} + \frac{n s_{sec}(t)}{L} (v_p(t) - n v_s(t)) \quad (5)$$

where C_1 and C_2 represent input and output capacitors, and $s_{pri}(t)$ and $s_{sec}(t)$ are functions describing circuit switching behaviors. v_{C1} and v_{C2} govern the stability, robustness, and power quality when interfacing other power conversion circuits.

CirNet for System-level Modeling of Circuit Physics: Subsequently, the system-level modeling is attained with the physics-informed circuit net, CirNet. CirNet retrofits a recurrent LN-GRU net to encode circuit physical laws in its inherent feature space. Taking non-resonant DAB converter as an example, main state waveforms are i_L , v_{C1} and v_{C2} . By leveraging the Kirchhoff's, Faraday's, and Gauss's laws, the system-level dynamic equations are derived in Eqs. 3 to 5, where the electrical notations are given in Figure 1. Eq. 3 describes the dynamics of high-frequency ac current $i_L(t)$, and Eq. 4 and Eq. 5 present the electrical behaviors of input and output capacitors, which follow second-order differential equations.

To predict the key state waveforms, CirNet takes the outputs of ModNet as its inputs and iteratively infers the next state based on the predictions of previous states, as the structure in Figure 3 Part (3) and Eq. 7 shown.

In terms of the training for CirNet, similar to ModNet, the loss based on physics information l_p and the loss based on experimental data l_d are both deployed, as shown in Eq. 7, in which λ_d and λ_p are their loss factors. In l_p , the circuit physical dynamics expressed in differential equations 3 to 5 are embedded in the loss functions for physics learning.

In l_d , several data points of the inductor currents i_L in the hardware experiments are taken as the ground truth, where the datapoints are denoted as $\{i_{L,j}^*(t_1), \dots, i_{L,j}^*(t_T)\}_j^N$. The waveform predictions are saved for a switching period for the performance evaluation.

$$\hat{i}_L(t_{k+1}) = \text{CirNet}(\hat{i}_L(t_k), \text{ModNet}(\hat{v}_p(t_k), \hat{v}_s(t_k); \theta_{mod}); \theta_{cir}) \quad (6)$$

$$l_{cir} = \lambda_d l_d + \lambda_p l_p = \frac{\lambda_d}{NT} \sum_j \sum_k \left[\hat{i}_{L,j}(t_{k+1}) - i_{L,j}^*(t_{k+1}) \right]^2 + \frac{\lambda_p}{NT} \sum_j \sum_k \left[\begin{aligned} &L \left(\hat{i}_{L,j}(t_{k+1}) - i_{L,j}^*(t_k) \right) - \\ &(\hat{v}_{p,j}(t_{k+1}) - n \hat{v}_{s,j}(t_{k+1}) - \\ &R_L \hat{i}_{L,j}^*(t_k)) \Delta t_k \end{aligned} \right]^2 \quad (7)$$

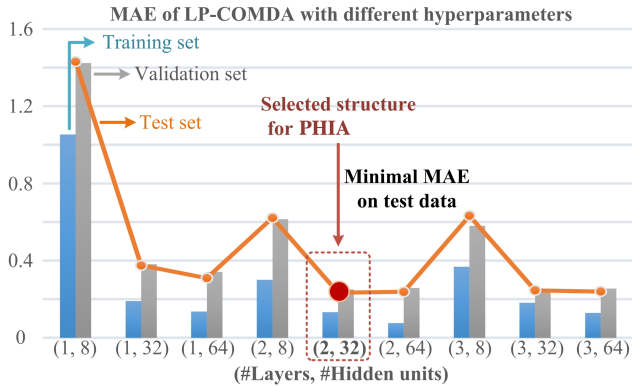


Figure 4: PHIA with different structures of CirNet. The minimal MAE of 0.235 on validation set is reached when there are 2 hidden layers with 32 hidden neurons.

Circuit Performance Evaluation Based on CirNet:

With the output of the two hierarchical PINNs, ModNet and CirNet, the power converter performance can be evaluated. With the key waveform inference results from the previous CirNet, a variety of electrical performance metrics can be measured, aiming to provide critical and comprehensive assessments for power electronics systems. As shown in Figure 3 Part (3), the difference of the maximum and minimum values of $i_L(t)$ gives the peak-to-peak current stress. The numerical integration of $[s_1(t)i_L(t)]^2$ provides the conduction loss evaluation for semiconductor switches. Soft switching performance, which is especially important for high-frequency scenarios like the electric vehicle charging, is analyzed by imposing constraints on $i_L(t)$ at the device commutation moment, where $s_{pri}(t)$ and $s_{sec}(t)$ are needed for synchronization. Other electrical performances can be gauged with the domain expertise in power electronics.

As described in Figure 2, in the core engine, optimization algorithms are deployed to cooperate with ModNet and CirNet to find the optimal modulation parameters. The optimization procedure is an iterative cycle, in which the optimization algorithm passes the operating conditions and modulation parameters to the physics-informed surrogate models ModNet and CirNet, and these models provide evaluated performance metrics to steer the optimization process.

By incorporating physics information within the neural networks, the model’s interpretability is heightened, concurrently leading to a reduction in the necessary training data points. Furthermore, the entire design automation system exhibits seamless scalability through the integration of surrogate models for additional functions into the core engine. Importantly, users can access these functionalities without any alterations to their experience.

Experiments

Experiments Setup

We primarily aim to ascertain the accuracy of the ModNet and CirNet surrogate models, where heightened model accuracy in the core engine translates to superior system design performance. Therefore, we carried out several experiments

Model	#Parameters	Selected Hyperparameters
LSTM	21,671	3 layers, 32 hidden size
GRU	13,319	1 layer, 64 hidden size
LN-GRU	13,319	1 layer, 64 hidden size
TCN	88,199	2 layers, 7 kernel size, 64 hidden size
GRU-VAE	88,583	2 layers, 64 hidden size
TST	6,913	1 layer, 2 attention heads, 32 hidden size
TSiTPlus	7,042	1 layer, 2 attention heads, 32 hidden size
MiniRocket	-	700 features, 6 dilatation size
CirNet	9,930	2 layers, 32 hidden size
PHIA	19,857	2 layers, 32 hidden size

Table 2: Hyperparameter selection of PHIA and benchmarks. The chosen hyperparameters yield the lowest loss on validation set.

to validate PHIA’s model performance in a low-resource setting, i.e., using a tiny dataset with only 200 samples.

Data Preparation: In this work, we focus on a modeling task of time-series forecasting for the inductor current $i_L(t)$ of DAB converters under triple phase-shift (TPS) modulation as a representative example. Using a hardware experimental prototype, we acquired waveform data spanning 200 sequences. The configuration of the hardware prototype and the methodology employed for data acquisition are elucidated in reference (Lin et al. 2023b).

One of the advantages of PINN models is that it requires very few samples to learn as compared to other machine learning models. To demonstrate this advantage, we conduct our experiments with two different data splitting ratios. For the first set, we split the data into 5% training (10 samples), 10% validation (20 samples) and 85% test (170 samples). The second set is split into 50%, 10% and 40% for training, validation and test, respectively.

Baseline: We evaluate PHIA’s design outcomes using: (1) the Bayesian Network (BN) (Pearl 1985), (2) Support Vector Regression (SVR) (Smola and Schölkopf 2004), (3) XG-Boost (Chen et al. 2015), (4) Random Forest (RF) (Breiman 2001), (5) Long-Short-Term-Memory (LSTM) (Hochreiter and Schmidhuber 1997), (6) GRU net (Chung et al. 2014), (7) LN-GRU net (Ba, Kiros, and Hinton 2016), (8) Temporal Convolutional Net (TCN) (Lea et al. 2017), (9) GRU-Based Variational Autoencoder (GRU-VAE) (An and Cho 2015), (10) Time-Series Transformer (TST) (Zerveas et al. 2021), (11) Time-Series Net Adapted from Vision Transformer (TSiT-Plus) (Oguiza 2022), and (12) MiniRocket (Tan et al. 2022).

Hyperparameter Search: For fair comparison, we conduct hyperparameter search on the validation set for PHIA and all baseline algorithms. We perform grid search on im-

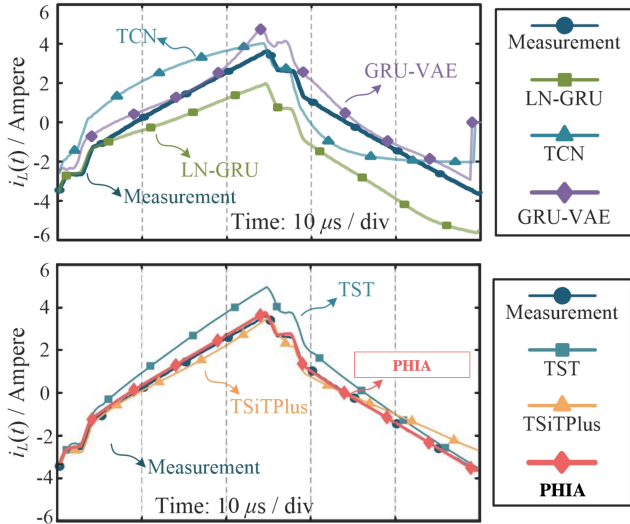


Figure 5: Modeling performance of the top 5 benchmarks and the proposed PHIA. The modelling results of PHIA is closest to the measurement, demonstrating its outstanding modeling accuracy.

portant hyperparameters for each model. Figure 4 illustrates the MAEs of PHIA with different hyperparameters, shedding light on the selected best structure. The final selected sets of hyperparameters are summarized in Table 2.

Experimental Results

Table 3 and 4 summarise the experimental results for PHIA and baseline for two different splits. Each algorithm with the respective optimal hyperparameters was ran for ten times. The average MAE for training, validation and test sets are reported. The reported P-values are the results of T-tests at a significance level of 0.02.

For both splits, experimental results show that PHIA outperforms conventional machine learning approaches, such as BN, SVR, XGBoost and RF, and exhibits superior performance over deep learning predictive models, such as LSTM, GRU, LN-GRU and TCN, and predictive models such as GRU-VAE. When compared to state-of-the-art models like TST, TSiTPlus, and MiniRocket, PHIA again demonstrates notably enhanced forecasting accuracy. To further affirm the hierarchical PINNs structure, we contrasted CirNet with PHIA. Noticeably, PHIA with 10 training samples outperforms most benchmarks with 100 training samples, and align with CirNet-Only model.

To visualize the modeling performance, Figure 5 presents the modeled i_L waveform for both PHIA and benchmark models. Notably, the PHIA modeling outcomes align more congruently with the measurements than any of the other models, demonstrating its outstanding modeling precision.

Overall, we conclude that the proposed PHIA statistically achieves better modeling results than other baseline models for power electronics modeling, even with extremely low-data scenario.

Model	Training	Validation	Test	p-value
BN	2.152±0.0	2.436±0.0	2.515±0.0	6.81E-11*
SVR	1.984±0.0	2.313±0.0	2.286±0.0	1.16E-10*
XGBoost	2.299±0.0	2.108±0.0	2.792±0.0	3.83E-11*
RF	4.983±0.0	4.181±0.0	4.434±0.0	3.18E-12*
LSTM	1.164±0.151	1.380±0.042	1.560±0.095	7.23E-8*
GRU	1.287±0.174	1.336±0.101	1.604±0.064	6.35E-10*
LN-GRU	0.834±0.272	1.289±0.165	1.257±0.117	1.71E-6*
TCN	1.571±0.210	1.996±0.257	1.538±0.501	1.436E-3*
TST	0.668±0.062	0.750±0.028	0.666±0.063	1.47E-6*
TSiTPlus	0.707±0.158	1.106±0.12	1.302±0.078	3.46E-8*
MiniRocket	0.113±0.122	0.742±0.365	0.763±0.338	0.0131*
GRU-VAE	1.928±0.201	2.190±0.180	2.219±0.172	6.47E-7*
CirNet Only	0.121±0.014	0.282±0.014	0.310±0.013	1.442E-3*
PHIA	0.140±0.054	0.235±0.025	0.245±0.029	N/A

Table 3: Experimental results with training / validation / test set split of 5% (10 samples), 10% (20 samples) and 85% (170 samples), respectively. PHIA statistically outperforms baselines (marked with *), with a 63.2% lower error as compared to the second-best benchmark (TST), measured in Mean Absolute Errors (MAE).

Interpretation

Accuracy and Reliability The improved accuracy observed in the experiments, particularly in low-data scenarios, underscores the robustness of the proposed framework. This enhancement ensures reliable performance, which is crucial for reducing the risk of failures in real-world applications. The system's ability to achieve high precision despite limited data sets demonstrates its adaptability and suitability for resource-constrained environments.

Efficiency Gains The significant reduction in error rates translates directly to efficiency gains. By optimizing workflows, the framework enables engineers to achieve desired outcomes with fewer iterations and reduced computational resources. This efficiency not only accelerates the design process but also minimizes the operational costs associated with traditional methods.

Scalability Performance improvements across metrics indicate better scalability of the framework. This scalability allows the system to handle more complex systems or broader application domains without requiring proportional increases in computational or human resources. Such scalability is essential for extending the applicability of the framework to diverse use cases.

Explainability and Transparency The integration of hierarchical modeling enhances the explainability of the results. By aligning the model predictions with physical principles, the framework produces outputs that are both interpretable and trustworthy. This transparency is crucial for fostering confidence in the system's decisions, especially in critical domains like power electronics.

Model	Training	Validation	Test	p-value
BN	2.455±0.0	2.206±0.0	2.348±0.0	5.03E-14*
SVR	2.049±0.0	1.780±0.0	1.991±0.0	1.25E-13*
XGBoost	1.321±0.0	2.194±0.0	1.986±0.0	1.27E-13*
RF	3.167±0.0	3.937±0.0	3.622±0.0	4.89E-15*
LSTM	1.246±0.165	1.129±0.078	1.193±0.098	1.89E-6*
GRU	1.196±0.127	1.356±0.134	1.270±0.114	2.80E-6*
LN-GRU	0.534±0.097	0.480±0.075	0.527±0.086	2.33E-4*
TCN	0.685±0.018	0.744±0.062	0.726±0.061	3.79E-6*
TST	0.264±0.033	0.240±0.033	0.264±0.031	3.93E-3*
TSiTPlus	0.585±0.099	0.567±0.082	0.569±0.099	2.61E-4*
MiniRocket	0.522±0.305	0.616±0.346	0.646±0.343	0.0246*
GRU-VAE	0.686±0.074	0.663±0.036	0.721±0.056	2.42E-6*
CirNet Only	0.162±0.032	0.239±0.014	0.234±0.007	3.98E-3*
PHIA	0.170±0.013	0.191±0.005	0.201±0.006	N/A

Table 4: Experimental results with training / validation / test set split of 50% (100 samples), 10% (20 samples) and 40% (80 samples), respectively. PHIA statistically outperforms baselines (marked with *), with a 23.7% lower error as compared to the second-best benchmark (TST), measured in Mean Absolute Errors (MAE).

Implications for Decision-Making The precision and reliability of the framework support informed decision-making in key applications. For example, in renewable energy systems and power grid operations, the metrics directly influence operational success and system stability. The enhanced performance ensures that the framework can be effectively deployed in high-stakes scenarios.

This interpretation highlights the broad implications of the observed performance improvements, showcasing the framework’s potential to revolutionize practical applications and advance theoretical understanding in the domain.

User Study

To assess PHIA’s empirical performance in improving engineering efficiency, we design a practical use case for the modulation strategy of the DAB converter and conduct an empirical experiment with 20 industrial practitioners.

Specifically, the user requirements for this design process include: (1) Guidance in defining design specifications; (2) Optimal design outcomes for the tailored scenario; and (3) Analysis on the design outcomes.

Based on the use case, we conduct a user study with 20 industrial practitioners recruited via research labs and collaborating external companies on a voluntary basis. The participants consist of 10 junior engineers with less than three years of experience, and 10 senior engineers with over five years of relevant experience. All participants declared to have no prior experience in using AI systems to assist engineering design works.

All participants are given two independent variants of the same task, namely to design the TPS modulation strategy

for a Dual Active Bridge (DAB) converter to serve as a DC transformer of the DC-DC power grid, as shown in Fig. 1. We ask the participant to first use PHIA’s chat interface to address task variant one. After completion, they then use standard practice (i.e. analytical and manual design) based on the given design requirements and data to address task variant two. To assess the effectiveness, the effective working hours are measured in 30-minute time blocks while tackling the tasks.

The results of the user study are highly promising, in terms of both functionality and efficiency. In terms of functionality, all 20 participants are able to complete both assigned tasks successfully. In terms of efficiency, the 10 junior engineers spent an average of 1.2 time blocks and 115.5 time blocks in completing the tasks using PHIA and analytical approach, respectively; where the 10 senior engineers recorded an average of 0.9 time blocks and 30.5 time blocks with PHIA and analytical approach, respectively. Overall, we observe that Junior engineers and senior engineers experienced 96.3 times and 33.9 times efficiency, respectively.

Conclusion

Modulation strategy design of power converters is pivotal for the optimal functioning of renewable energy power systems. However, numerous AI-centric modulation design automation systems today grapple with challenges tied to explainability, usability, scalability, and data intensity. We introduced PHIA, a LLM-based Physics-informed autonomous agent to perform effective power converter modulation design automation. Our physics-informed approach combines the reasoning capabilities of large language models with rigorous physical constraints, enabling robust and interpretable design solutions. Experiments demonstrate PHIA’s proficiency in accurate power converter modeling, even with a data-constrained environment. The hierarchical physics-informed neural network architecture ensures both accuracy and interpretability, while the LLM-based planning module provides an intuitive interface for engineers. Our user study also shows that using PHIA helps engineers improve efficiency by over 33 times in practice, highlighting its potential for widespread adoption in industrial applications. Future work could explore extending PHIA to other power electronics applications and incorporating additional physics-based constraints to further enhance its capabilities.

Acknowledgements

This work is led and supported in part by Forth AI for the use of computational resources, software licenses, in-kind R&D contributions and domain expertise.

This research is supported in part by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (Award No. MOE-T2EP20123-0015). Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not reflect the views of the Ministry of Education, Singapore.

This work is also supported in part by the National Natural Science Foundation of China under Grant 52407225, and in part by the ZJU-UIUC Institute Startup Funding.

References

- An, J.; and Cho, S. 2015. Variational autoencoder based anomaly detection using reconstruction probability. *Special lecture on IE*, 2(1): 1–18.
- Anil, R.; Dai, A. M.; Firat, O.; Johnson, M.; Lepikhin, D.; Passos, A.; Shakeri, S.; Taropa, E.; Bailey, P.; Chen, Z.; et al. 2023. Palm 2 technical report. *arXiv preprint arXiv:2305.10403*.
- Ba, J. L.; Kiros, J. R.; and Hinton, G. E. 2016. Layer normalization. *arXiv preprint arXiv:1607.06450*.
- Breiman, L. 2001. Random forests. *Machine learning*, 45: 5–32.
- Cao, Y.; and Lee, C. 2023. Robot Behavior-Tree-Based Task Generation with Large Language Models.
- Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T.; et al. 2015. Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4): 1–4.
- Chung, J.; Gulcehre, C.; Cho, K.; and Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- Gao, S.; Wang, Y.; and Yang, X. 2023. StockFormer: Learning Hybrid Trading Machines with Predictive Coding. In *IJCAI*, 4766–4774.
- Guo, C.; Liu, Y.; Xiong, M.; Xue, S.; Fu, X.; Huang, A.; Qiang, X.; Xu, P.; Liu, J.; Zheng, S.; et al. 2019. General-Purpose Quantum Circuit Simulator with Projected Entangled-Pair States and the Quantum Supremacy Frontier. *PRL*, 123(19): 190501.
- Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, 9(8): 1735–1780.
- Hou, N.; and Li, Y. W. 2019. Overview and comparison of modulation and control strategies for a nonresonant single-phase dual-active-bridge DC–DC converter. *IEEE Transactions on Power Electronics*, 35(3): 3148–3172.
- Hu, X.; Hu, H.; Verma, S.; and Zhang, Z.-L. 2020. Physics-guided deep neural networks for power flow analysis. *IEEE Transactions on Power Systems*, 36(3): 2082–2092.
- Huang, B.; and Wang, J. 2022. Applications of physics-informed neural networks in power systems—a review. *IEEE Transactions on Power Systems*, 38(1): 572–588.
- Ji, J.; Wang, J.; Jiang, Z.; Jiang, J.; and Zhang, H. 2022. STDEN: Towards physics-guided neural networks for traffic flow prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 4048–4056.
- Jia, X.; Willard, J.; Karpatne, A.; Read, J. S.; Zwart, J. A.; Steinbach, M.; and Kumar, V. 2021. Physics-guided machine learning for scientific discovery: An application in simulating lake temperature profiles. *ACM/IMS Transactions on Data Science*, 2(3): 1–26.
- Kang, N.; Lee, B.; Hong, Y.; Yun, S.-B.; and Park, E. 2023. PIXEL: Physics-Informed Cell Representations for Fast and Accurate PDE Solvers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 8186–8194.
- Koa, K. J.; Ma, Y.; Ng, R.; and Chua, T.-S. 2024. Learning to generate explainable stock predictions using self-reflective large language models. In *Proceedings of the ACM on Web Conference 2024*, 4304–4315.
- Lea, C.; Flynn, M. D.; Vidal, R.; Reiter, A.; and Hager, G. D. 2017. Temporal convolutional networks for action segmentation and detection. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 156–165.
- Li, C.; Yang, R.; Li, T.; Bafarassat, M.; Sharifi, K.; Bergemann, D.; and Yang, Z. 2024a. STRIDE: A Tool-Assisted LLM Agent Framework for Strategic and Interactive Decision-Making. *arXiv preprint arXiv:2405.16376*.
- Li, Z.; Xia, L.; Tang, J.; Xu, Y.; Shi, L.; Xia, L.; Yin, D.; and Huang, C. 2024b. Urbangpt: Spatio-temporal large language models. *arXiv preprint arXiv:2403.00813*.
- Lin, F. 2022. The dual active bridge converter design with artificial intelligence.
- Lin, F.; Liu, J.; Li, X.; Zhao, S.; Zhao, B.; Ma, H.; Zhang, X.; and Liao, X. 2024. PE-GPT: A Physics-Informed Interactive Large Language Model for Power Converter Modulation Design. In *Proceedings of the IEEE Energy Conversion Congress and Exposition (ECCE'24)*.
- Lin, F.; Zhang, X.; and Li, X. 2021. Design Methodology for Symmetric CLLC Resonant DC Transformer Considering Voltage Conversion Ratio, System Stability, and Efficiency. *IEEE Transactions on Power Electronics*, 36(9): 10157–10170.
- Lin, F.; Zhang, X.; Li, X.; Sun, C.; Zsurzsan, G.; Cai, W.; and Wang, C. 2023a. AI-Based Design with Data Trimming for Hybrid Phase Shift Modulation for Minimum-Current-Stress Dual Active Bridge Converter. *IEEE Journal of Emerging and Selected Topics in Power Electronics*.
- Lin, F.; Zhang, X.; Li, X.; Sun, C.; Zsurzsan, G.; Cai, W.; and Wang, C. 2023b. AI-Based Design with Data Trimming for Hybrid Phase Shift Modulation for Minimum-Current-Stress Dual Active Bridge Converter. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 1–1.
- Liu, J.; Albrethsen, J.; Goh, L.; Yau, D.; and Lim, K. H. 2024a. Spatial-Temporal Graph Representation Learning for Tactical Networks Future State Prediction. In *Proceedings of The International Joint Conference on Neural Networks (IJCNN'24)*.
- Liu, J.; Keat, T. Y.; Fu, B.; and Lim, K. H. 2024b. LARA: Linguistic-Adaptive Retrieval-Augmentation for Multi-Turn Intent Classification. In Dernoncourt, F.; Preoțiuc-Pietro, D.; and Shimorina, A., eds., *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, 1096–1106. Miami, Florida, US: Association for Computational Linguistics.
- Liu, J.; Lim, K. H.; Wood, K. L.; Huang, W.; Guo, C.; and Huang, H.-L. 2021. Hybrid quantum-classical convolutional neural networks. *Science China Physics, Mechanics & Astronomy*, 64(9): 290311.
- Liu, J.; Tan, Y. K.; Fu, B.; and Lim, K. H. 2025. From Intent to Conversations: Generating Intent-Driven Dialogues with Contrastive Learning for Multi-Turn Classification. In *CIKM*.

- Liu, J.; Tan, Y. K.; Fu, B.; and Lim, K. H. 2026. Balancing Accuracy and Efficiency in Multi-Turn Intent Classification for LLM-Powered Dialog Systems in Production. *AAAI'26*.
- McNutt, A. M.; Wang, C.; Deline, R. A.; and Drucker, S. M. 2023. On the design of ai-powered code assistants for notebooks. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–16.
- Mudunuru, M. K.; O'Malley, D.; Srinivasan, S.; Hyman, J. D.; Sweeney, M. R.; Frash, L.; Carey, B.; Gross, M. R.; Welch, N. J.; Karra, S.; et al. 2020. Physics-informed machine learning for real-time unconventional reservoir management. 1–10.
- Oguiza, I. 2022. tsai - A state-of-the-art deep learning library for time series and sequential data. Github.
- OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774.
- Pearl, J. 1985. Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of the 7th conference of the Cognitive Science Society, University of California, Irvine, CA, USA*, 15–17.
- Raissi, M. 2018. Deep hidden physics models: Deep learning of nonlinear partial differential equations. *The Journal of Machine Learning Research*, 19(1): 932–955.
- Ross, S. I.; Martinez, F.; Houde, S.; Muller, M.; and Weisz, J. D. 2023. The programmer's assistant: Conversational interaction with a large language model for software development. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 491–514.
- Shakarian, P.; Koyyalamudi, A.; Ngu, N.; and Mareedu, L. 2023. An Independent Evaluation of ChatGPT on Mathematical Word Problems (MWP).
- Shi, R.; Mo, Z.; and Di, X. 2021. Physics-informed deep learning for traffic state estimation: A hybrid paradigm informed by second-order traffic models. 35(1): 540–547.
- Shi, W.; Hu, Z.; Bin, Y.; Liu, J.; Yang, Y.; Ng, S.-K.; Bing, L.; and Lee, R. K.-W. 2024. Math-LLaVA: Bootstrapping Mathematical Reasoning for Multimodal Large Language Models. arXiv preprint arXiv:2406.17294.
- Smola, A. J.; and Schölkopf, B. 2004. A tutorial on support vector regression. *Statistics and computing*, 14: 199–222.
- Tan, C. W.; Dempster, A.; Bergmeir, C.; and Webb, G. I. 2022. MultiRocket: multiple pooling operators and transformations for fast and effective time series classification. *Data Mining and Knowledge Discovery*, 36(5): 1623–1646.
- Tang, Y.; Hu, W.; Xiao, J.; Chen, Z.; Huang, Q.; Chen, Z.; and Blaabjerg, F. 2020. Reinforcement learning based efficiency optimization scheme for the DAB DC–DC converter with triple-phase-shift modulation. *IEEE Transactions on Industrial Electronics*, 68(8): 7350–7361.
- Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. 2023. Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.
- Wandel, N.; Weinmann, M.; Neidlin, M.; and Klein, R. 2022. Spline-pinn: Approaching pdes without data using fast, physics-informed hermite-spline cnns. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 8529–8538.
- Wang, L.; Ma, C.; Feng, X.; Zhang, Z.; Yang, H.; Zhang, J.; Chen, Z.; Tang, J.; Chen, X.; Lin, Y.; et al. 2024. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6): 186345.
- Xiang, Z.; Zhao, H.; Zhao, C.; He, M.; and Fan, J. 2024. Performative Debias with Fair-exposure Optimization Driven by Strategic Agents in Recommender Systems. arXiv preprint arXiv:2406.17475.
- Xiao, D.; Heaney, C.; Mottet, L.; Fang, F.; Lin, W.; Navon, I.; Guo, Y.; Matar, O.; Robins, A.; and Pain, C. 2019. A reduced order model for turbulent flows in the urban environment using machine learning. *Building and Environment*, 148: 323–337.
- Xu, Y.; Cai, T.; Jiang, J.; and Song, X. 2024. Face4RAG: Factual Consistency Evaluation for Retrieval Augmented Generation in Chinese. arXiv preprint arXiv:2407.01080.
- Yang, Z.; Qiu, Z.; and Fu, D. 2023. Dmis: Dynamic mesh-based importance sampling for training physics-informed neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 5375–5383.
- Yao, X.; Shinn, A.; Liu, K.; and Xu, L. 2024. Large Language Model-based Human-Agent Collaboration for Complex Task Solving. arXiv preprint arXiv:2402.12914.
- Zerveas, G.; Jayaraman, S.; Patel, D.; Bhamidipaty, A.; and Eickhoff, C. 2021. A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2114–2124.
- Zhang, C.; Liu, X.; Jin, M.; Zhang, Z.; Li, L.; Wang, Z.; Hua, W.; Shu, D.; Zhu, S.; Jin, X.; et al. 2024a. When AI Meets Finance (StockAgent): Large Language Model-based Stock Trading in Simulated Real-world Environments. arXiv preprint arXiv:2407.18957.
- Zhang, W.; Zhao, L.; Xia, H.; Sun, S.; Sun, J.; Qin, M.; Li, X.; Zhao, Y.; Zhao, Y.; Cai, X.; et al. 2024b. A Multimodal Foundation Agent for Financial Trading: Tool-Augmented, Diversified, and Generalist. Technical report.
- Zhao, S.; Peng, Y.; Zhang, Y.; and Wang, H. 2022. Parameter estimation of power electronic converters with physics-informed machine learning. *IEEE Transactions on Power Electronics*, 37(10): 11567–11578.