Understanding turbulence-induced transport in nuclear fusion plasmas using neuro-symbolic AI: prospects and challenges

Farbod Faraji*[1]

*Department of Computing, Imperial College London, 180 Queen's Gate, South Kensington, London SW7 2RH, United Kingdom

Abstract: Plasma turbulence and turbulence-induced transport remain key unknowns in fusion science because the links between instability evolution, coherent structure formation, wave-particle interactions, and particle and energy transport are still incompletely resolved. High-fidelity kinetic simulations capture essential transport-driving physics but are costly at realistic scales, while fluid and hybrid models depend on closure relations that are often regime-specific or incomplete. Existing machine-learning approaches provide useful surrogate and reduced-order modelling capabilities, but may rely on soft physics constraints, have limited cross-regime generalization, or remain opaque in how predictions are generated. This paper presents neuro-symbolic AI as a route toward physics-constrained dynamics discovery for plasma turbulence. The proposed architecture learns compact coordinates from plasma data, searches for interpretable symbolic dynamics in those coordinates, and uses logic-driven constraints to enforce physical principles during discovery. The scientific significance of this route is to enable deriving coordinate-dynamics descriptions that can be probed and interpreted to reveal transport mechanisms.

1. Introduction

A central objective of magnetic-confinement fusion research is to understand how a hot plasma can be confined long enough, and with sufficient stability, for sustained energy generation. Realizing this objective is however limited not only by large-scale equilibrium and stability constraints, but also by smaller-scale fluctuations that can enhance the transport of particles, momentum, and heat across magnetic surfaces. Such turbulence-induced transport affects core confinement, the formation and relaxation of edge pedestal structures, transient edge events, scrape-off-layer transport, and the deposition of heat and particle loads on divertor and plasma-facing components [1-3]. Understanding how fluctuation-scale dynamics give rise to macroscopic behaviors and transport is therefore central to progress in fusion science and to improving the predictive capability of fusion plasma models.

A persistent challenge is that this transport does not arise from single isolated mechanisms. Background gradients in density, temperature, rotation, and current provide free energy for instabilities, whose growth, nonlinear interaction, and saturation can generate turbulent fluxes across a wide range of spatial and temporal scales [1,4,5]. These processes may also organize into coherent structures, including zonal flows, streamers, blobs, or filaments, depending on the plasma region and operating regime [2]. At the same time, wave-particle interactions and phase-space evolutions can influence how fluctuations grow, saturate, and exchange energy, while the resulting transport modifies the profiles that originally drive the instabilities. This feedback between profiles, fluctuations, coherent structures, and particle/energy transport poses major challenges to identify the dominant pathways under different plasma conditions. It also makes the problem especially difficult for models that aim not only to reproduce observed transport levels, but to explain how transport mechanisms act and vary across zones and regimes.

First-principles modelling has provided valuable insight into these processes, but practical predictive capability remains constrained by the competing requirements of fidelity, scale, and computational cost. Kinetic descriptions, including Vlasov-Fokker-Planck, particle-based, and gyrokinetic approaches, can resolve essential physics such as wave-particle interactions, non-Maxwellian features, anisotropy, and non-local effects that drive transport [4,6]. However, approaching realistic geometry, dimensionality, parameter coverage, and operational timescales, such simulations rapidly become computationally prohibitive. Fluid and hybrid fluid-kinetic models

[1] Eric and Wendy Schmidt AI in Science Research Fellow; email: f.faraji20@imperial.ac.uk.

are more computationally tractable and important for plasma studies at scales where kinetic descriptions may not be affordable, but they require closures to represent unresolved kinetic physics. Available closures are often regime-dependent, may miss key phenomena such as resonant wave-particle coupling, and can struggle with strong anisotropy and multiscale couplings in fusion plasmas [5,7].

Machine learning has opened new possibilities in this landscape by accelerating expensive calculations, developing reduced-order surrogate models, inferring unresolved quantities, and identifying dynamical structure from high-dimensional plasma data [8]. Several methodological directions are already relevant to plasma modelling: physics-informed neural networks and related constrained-learning approaches embed governing-equation residuals or conservation statements into training objectives [9]; operator-learning architectures learn mappings between input functions and solution fields for fast emulation of parametrized differential-equation systems [10]; sparse regression and symbolic model discovery methods seek parsimonious dynamical laws from data [11]; and neural transport surrogates have demonstrated substantial speedups in fusion plasma modelling [12]. These developments are valuable, but the requirements for turbulence-induced transport studies are more demanding than accurate interpolation or short-horizon forecasting.

For scientific discovery, a useful model should help identify which dynamical processes are being represented, how they change across plasma conditions, and whether the inferred relations remain consistent with physical principles. Many existing approaches remain limited in this respect. Physics-informed models often impose constraints through soft penalties, so constraint violation can be traded against empirical fit during training. Symbolic regression search is often shaped by expert-crafted dynamical candidate term libraries and equation complexity penalties. Deep neural models may achieve strong predictive performance but leave internal learning and prediction mechanisms opaque.

This motivates the neuro-symbolic direction considered in this paper. The central idea is to combine neural learning of compact coordinates from plasma data with symbolic discovery of interpretable dynamical models in those coordinates [13]. Logic-driven constraints are used to enforce physical principles such as symmetry, invariance, dimensional consistency, conservation or balance relations, boundedness, and stability during the discovery process. In this framing, the objective is to build a physics-aware, reasoning-enabled architecture that is supported by theory and simulation and can help expose hidden links and interaction pathways between turbulent evolution, coherent structures, wave-particle interactions, and turbulence-induced transport.

2. A physics-aware, reasoning-enabled dynamics discovery architecture

Figure 1 shows the proposed physics-aware, reasoning-enabled dynamics discovery architecture for turbulence-induced transport studies. The figure highlights how plasma information can be transformed into compact coordinate-dynamics descriptions, while enforcing physical principles during model discovery and retaining information about the bounds in which the inferred dynamics are verified.

The first component is the multi-modal plasma-state interface. The input data can be written schematically as $\mathcal{D} = \{x(t), y_m(t), \mu, u(t), \sigma_m, \mathcal{G}\}$, where $x(t)$ denotes high-dimensional simulated or camera-captured plasma states, $y_m(t)$ sparse diagnostic signals, μ operating parameters, $u(t)$ control inputs, and σ_m, $\mathcal{G}$ measurement uncertainty, metadata, and geometry/provenance information. For a magnetic-confinement plasma, μ may include quantities such as toroidal field, plasma current, density, auxiliary heating power, and pulse duration, while $u(t)$ may include time-dependent heating, coil currents, gas injection, or feedback actuation. The key requirement here is that the data are not treated as anonymous arrays. Their physical identity, units, uncertainty, source, geometry, and role as state, diagnostic, parameter, or control information should remain available to later discovery and verification steps.

The second component is a latent predictive backbone model that learns compact coordinates for the observed plasma evolution. In schematic form: $z_t = \phi_\theta\left(x_{\leq t}, y_{m,\leq t}; \mu, u_{\leq t}\right)$, where z_t is a learned coordinate representation of the plasma state over the observed history. The backbone can predict future latent states, for example $z_{t+1}, \dots, z_{t+H}$, before mapping them back to physical fields or quantities of interest. For turbulence-induced transport, these coordinates should be conditioned to represent transport-relevant patterns in the evolution of fluctuations, coherent structures, and profile changes.

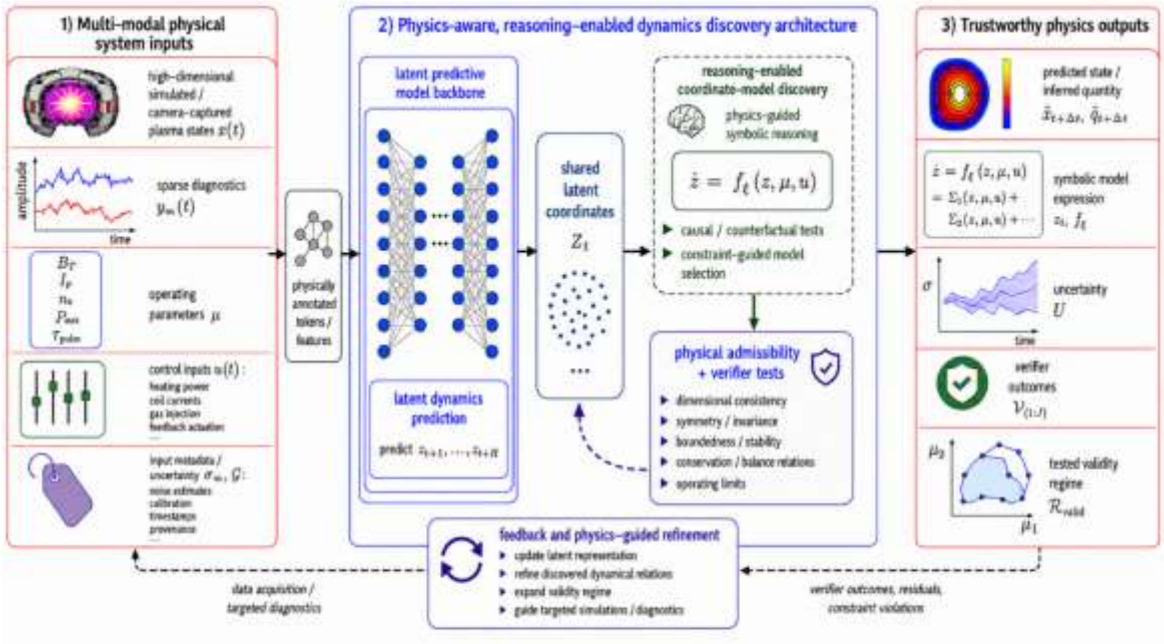

Figure 1: Schematic representation of a physics-aware, reasoning-enabled dynamics discovery architecture

The third and central component of the architecture is reasoning-enabled coordinate-model discovery. Once compact coordinates have been learned, the architecture searches for symbolic dynamical models in those coordinates, written schematically as $\dot{z} = f_\xi(z, \mu, u)$. Here, f_ξ may represent low-dimensional ODE relations for the learned dominant turbulence structures, reduced balance relations, flux laws, or closure-like expressions linking fluctuation dynamics to transport-relevant quantities. The important point is that the object being discovered is not only a symbolic equation, but a coordinate-model pair: the learned representation z_t and the dynamical relation f_ξ must be meaningful together. This differs from applying symbolic regression after fixing an arbitrary latent space. This is important because a dynamics model that fits trajectories but violates physical principles, or fails when parameters change, is not a satisfactory model for plasma physics discovery, and so should be rejected by the architecture.

The fourth component therefore imposes physical admissibility and dynamics-model verification tests during discovery. Candidate coordinate-model pairs should be constrained by principles that can be stated concretely for the plasma problem under consideration. Dimensional consistency requires, for example, that symbolic terms proposed for growth rates, diffusivities, heat fluxes, particle fluxes, or energy-transfer rates carry the appropriate units. Conservation constraints require candidate models to remain compatible, in the relevant limit, with mass, momentum, or energy balance for instance. Field-particle consistency should prevent learned coordinates or symbolic dynamics terms from implying relations incompatible with quasineutrality or field equations. Positivity and boundedness tests can reject candidates that drive density, temperature, pressure, or distribution-related quantities outside physically admissible ranges. Symmetry and invariance tests should be applied according to the geometry: in simplified settings, this may include translations, rotations, sign symmetries, or normalization changes. Logic-driven physical reasoning provides a route to enforce these requirements through admissible variable types, symbolic grammars, hard filters, and executable checks, rather than treating every physical requirement as a soft loss penalty that only promotes consistency.

The fifth component introduces causal and counterfactual testing [14]. In practice, this implies testing candidate dynamics under parameter scans, including altered heating or fueling conditions, modified forcing, or withheld simulation regimes. Schematically, one may consider $\mu \rightarrow \mu',\ u(t) \rightarrow u'(t)$, and then assess whether the inferred coordinate dynamics respond in a physically consistent way. For turbulence-induced transport, this is essential because a model may reproduce observed fluctuation fields while learning correlations that do not correspond to causally relevant mechanisms that govern, or are governed by, the fluctuation and transport dynamics. Parameter and control variations allow the discovered terms to be probed: do they continue to describe the coupling between instability evolution, coherent structures, and transport-relevant quantities, or do they fail outside the conditions from which they were inferred? This provides a structured way to test whether a learned relation describes a candidate physical mechanism.

The final component is the trustworthy physics output layer. The architecture should not only return a predicted plasma state. Its output can be written schematically as $\mathcal{O} = \{\hat{x}_{t+\Delta t}, \hat{q}_{t+\Delta t}, z_t, f_\xi, U, \mathcal{V}_{1:J}, \mathcal{R}_{\text{valid}}\}$.

$\hat{x}_{t+\Delta t}$ is a predicted physical state, $\hat{q}_{t+\Delta t}$ an inferred quantity, z_t the learned coordinate representation, f_ξ the selected symbolic dynamical model, U an uncertainty estimate, $\mathcal{V}_{1:J}$ the outcomes of physical and dynamical verifications, and $\mathcal{R}_{\text{valid}}$ the tested validity regime. In physical terms, this means that a prediction is accompanied by the coordinate description used to generate it, the symbolic dynamics inferred from the data, the uncertainty attached to the prediction, the tests the model passed or failed, and the region of parameter or operating space over which the relation has been examined and verified.

The feedback loop in Figure 1 closes the discovery process. Test outcomes, residuals, and constraint violations can be used to update the latent representation, refine the discovered dynamical relations, and guide targeted simulations or diagnostics. This loop is important for turbulence-induced transport because the most useful outcome is not a single fitted model, but an iterative route toward coordinate-dynamics descriptions that become progressively more interpretable, physically constrained, and informative about the pathways linking turbulent evolution, coherent structures, wave-particle interactions, and transport.

3. Promises and challenges

The promise of the architecture in Figure 1 is that it enables studying turbulence-induced transport through compact, interpretable coordinate-dynamics descriptions that address key limitations of current methods. If such descriptions can be learned from plasma data while satisfying physics fundamentals, they can help identify how turbulence-driven mechanisms and interactions lead to particle and energy transport.

Nevertheless, several technical challenges must be addressed to achieve this promise. First, physical constraints must be encoded in forms amenable to logic-based reasoning so that they can be enforced during discovery. Second, the learned coordinates must remain physically meaningful: they should not only compress the data, but retain traceable links to specific fluctuation dynamics, coherent structures, and transport processes. Third, candidate coordinate-dynamics models must remain valid when plasma conditions change, rather than fitting only one dataset or operating point. Demonstrating these points on controlled plasma problems, and then on progressively more fusion-relevant turbulence datasets, would help establish whether and how neuro-symbolic AI can become a useful discovery tool for revealing and explaining transport mechanisms in fusion and broader applications of plasma science and engineering.

References:

[1] Horton W, 1999, Drift waves and transport, *Rev. Mod. Phys.* **71** 735–778.

[2] Diamond P H, Itoh S-I, Itoh K and Hahm T S, 2005, Zonal flows in plasma – a review, *Plasma Phys. Control. Fusion* **47** R35–R161.

[3] Loarte A, Lipschultz B, Kukushkin A S, Matthews G F, Stangeby P C, et al., the ITPA Scrape-off Layer and Divertor Physics Topical Group 2007 Chapter 4: Power and particle control, *Nucl. Fusion* **47** S203–S263.

[4] Krommes J A, 2002, Fundamental statistical descriptions of plasma turbulence in magnetic fields, *Phys. Rep.* **360** 1–352.

[5] Dimits A M, Bateman G, Beer M A, Cohen B I, Dorland W, et al., 2000, Comparisons and physics basis of tokamak transport models and turbulence simulations, *Phys. Plasmas* **7** 969–983.

[6] Brizard A J and Hahm T S, 2007, Foundations of nonlinear gyrokinetic theory, *Rev. Mod. Phys.* **79** 421–468.

[7] Braginskii S I, 1965, Transport processes in a plasma, *Reviews of Plasma Physics* **1** 205–311.

[8] Faraji F and Reza M, 2025, Machine learning applications to computational plasma physics and reduced-order plasma modeling: a perspective, *J. Phys. D: Appl. Phys.* **58** 102002.

[9] Raissi M, Perdikaris P and Karniadakis G E, 2019, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* **378** 686–707.

[10] Li Z, Kovachki N, Azizzadenesheli K, Liu B, Bhattacharya K, Stuart A, and Anandkumar A, 2021, Fourier Neural Operator for Parametric Partial Differential Equations, *International Conference on Learning Representations*.

[11] Brunton S L, Proctor J L, and Kutz J N, 2016, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl Acad. Sci. USA* **113** 3932–3937.

[12] van de Plassche K L, Citrin J, Bourdelle C, Camenen Y, et al., 2020, Fast modeling of turbulent transport in fusion plasmas using neural networks, *Phys. Plasmas* **27** 022310.

[13] d'Avila Garcez A and Lamb L C, 2023, Neurosymbolic AI: the 3rd wave, *Artif. Intell. Rev.* **56** 12387–12406.

[14] Schölkopf B, Locatello F, Bauer S, Ke N R, Kalchbrenner N, Goyal A and Bengio Y, 2021, Toward causal representation learning, *Proc. IEEE* **109** 612–634.