Artificial Intelligence and Electrical & Electronics Engineering: AIEEE Open Access

PRIME OPEN ACCESS

Volume 2, Issue 1

Review Article

Date of Submission: 01 January, 2026
Date of Acceptance: 23 January, 2026
Date of Publication: 30 January, 2026

A Neural PDE Framework via Navier–Stokes Dynamics: Continuous-Depth Learning and Hybrid Attention-Driven Forcing

Chur Chin*

Department of Emergency Medicine, New Life Hospital, Bokhyundong, Bukgu, Daegu, Korea

Corresponding Author:
Chur Chin, Department of Emergency Medicine, New Life Hospital, Bokhyundong, Bukgu, Daegu, Korea.

Citation: Chin, C. (2026). A Neural PDE Framework via Navier–Stokes Dynamics: Continuous-Depth Learning and Hybrid Attention-Driven Forcing. *Art Intelligence and Ele & Electronics Eng: AIEEE Open Access, 2*(1), 01-06.

Abstract

We propose a reformulation of a Navier–Stokes–based dynamical learning framework as a continuous-depth neural network (Neural PDE), and develop a hybrid model in which transformer-style attention mechanisms act as adaptive forcing terms within the flow. In contrast to conventional neural networks trained by gradient descent and backpropagation, the proposed approach replaces explicit loss minimization with evolution governed by energy dissipation, spectral cascades, and resolvent-based regularity control. We show how the flow map of the Navier–Stokes equation can be interpreted as an infinite-depth residual network, establish conditional regularity criteria analogous to Beale–Kato–Majda, and position attention as a data-dependent operator that injects structured information into the dynamics. This framework provides a principled bridge between operator-theoretic PDE analysis and modern deep learning architectures, with potential applications in structured inference, multiscale representation learning, and mathematically constrained learning systems.

Keywords: Neural PDE, Continuous-Depth Neural Networks, Navier–Stokes Equations, Attention Mechanisms, Energy Cascade, Operator Theory

Introduction

Deep learning has achieved remarkable empirical success through gradient-based optimization of high-dimensional parameter spaces, particularly via backpropagation and, more recently, transformer architectures with attention mechanisms [1–3]. Despite this success, the theoretical understanding of stability, generalization, and interpretability remains limited and largely statistical [4]. In parallel, the theory of partial differential equations (PDEs), especially fluid dynamics, offers a mature analytical framework characterized by energy inequalities, spectral decompositions, and conditional regularity results [5–7].

Recent work has suggested that deep neural networks may be interpreted as discretizations of continuous dynamical systems, giving rise to the concept of neural ordinary differential equations and neural PDEs [8–10]. Motivated by this perspective, we develop a framework in which learning dynamics are governed by Navier–Stokes–type equations, emphasizing energy flow and dissipation rather than explicit loss minimization.

The main contributions of this paper are threefold. First, we recast a Navier–Stokes–based learning mechanism as a continuous-depth neural network. Second, we introduce a hybrid architecture in which attention mechanisms appear as adaptive forcing terms within the PDE. Third, we compare the resulting framework to conventional gradient-based and transformer-based methods from both computational and theoretical perspectives.

Background

Gradient-Based Neural Networks

Standard neural networks are trained by minimizing a loss function via stochastic gradient descent and backpropagation [1]. The learning dynamics are discrete and parameter-centric, with stability controlled by hyperparameters such as learning rate and regularization strength.

Transformers and Attention

Transformers employ attention mechanisms to model global dependencies in data [2,3]. Given queries, keys, and values, attention computes weighted averages that dynamically reallocate representational focus. While powerful, attention remains an algebraic operation lacking intrinsic notions of conservation or dissipation.

Neural ODEs and Neural PDEs

Neural ODEs reinterpret deep networks as continuous-time dynamical systems [8], while neural PDEs generalize this idea to spatially distributed representations [9,10]. These approaches motivate viewing depth as time and layers as flow maps.

Recasting the Navier–Stokes Framework as a Continuous-Depth Neural Network

Consider the incompressible Navier–Stokes equations on a domain $\Omega \subset R^d$:

$$\partial_t u + (u \cdot \nabla)u = -\nabla p + \nu \Delta u + F, \nabla \cdot u = 0.$$

Here $u(x,t)$ represents a feature field evolving in continuous depth t. The flow map $u(0) \mapsto u(T)$ plays the role of an infinitely deep residual network, with viscosity ν acting as an intrinsic regularizer.

In this interpretation, nonlinear advection corresponds to feature interaction, while the Laplacian enforces spectral damping. Unlike discrete networks, stability is governed by the energy inequality

$$d/dt \|u(t)\|_{L^2}^2 + 2\nu \|\nabla u(t)\|_{L^2}^2 \leq 2\langle F,u\rangle.$$

Conditional Regularity and Operator Control

The well-posedness of the proposed Neural PDE is tied to classical conditional regularity results. In analogy with the Beale–Kato–Majda criterion [6], smoothness is preserved provided

$$\int_0^T \|\omega(t)\|_{L^\infty} dt < \infty,$$

where $\omega = \nabla \times u$ denotes vorticity. Resolvent estimates for the associated Stokes operator further bound high-frequency amplification [7]. These results replace heuristic regularization techniques commonly used in deep learning.

Hybrid Model: Attention as Forcing Term

We now introduce a hybrid architecture in which attention mechanisms enter as a forcing term F_{attn}. Given an input representation X, we define

$$F_{attn}(x,t) = A(Q(X),K(X),V(X))(x,t),$$

where A denotes a standard attention operator [2]. This forcing injects data-dependent, nonlocal information into the flow while preserving the underlying PDE structure.

The resulting system combines global context selection (attention) with multiscale energy redistribution (Navier–Stokes dynamics). Attention guides the flow, while dissipation prevents uncontrolled growth.

Comparison with Conventional Architectures

Compared to gradient-based networks, the proposed framework does not rely on explicit loss gradients, instead leveraging physical regularization via viscosity and spectral decay. Relative to transformers, attention is no longer the sole organizing principle but acts in concert with a continuous dynamical system.

Discussion and Outlook

The Neural PDE and hybrid attention–Navier–Stokes framework offers a mathematically grounded alternative to conventional deep learning architectures. It is particularly suited for structured inference tasks where stability, interpretability, and multiscale behavior are paramount. Future work includes numerical discretization strategies, data-driven identification of forcing terms, and applications to scientific machine learning.

Conflict of interest

There is no conflict of interest.

Data availability statement
Data openly available in a public repository that issues datasets with DOIs.

References

1. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature, 323*(6088), 533-536.
2. Ashish, V. (2017). Attention is all you need. *Advances in neural information processing systems, 30*, I.
3. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1* (long and short papers) (pp. 4171-4186).
4. Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017). Understanding deep learning requires rethinking generalization.
5. Temam, R. (1977). : Navier-Stokes Equations. North-Holland Publishing Company.
6. Beale, J. T., Kato, T., & Majda, A. (1984). Remarks on the breakdown of smooth solutions for the 3-D Euler equations. *Communications in Mathematical Physics, 94*(1), 61-66.
7. Constantin, P., & Foiaş, C. (1988). *Navier-stokes equations*. University of Chicago press.
8. Chen, R. T., Rubanova, Y., Bettencourt, J., & Duvenaud, D. K. (2018). Neural ordinary differential equations. *Advances in neural information processing systems, 31*.
9. Ruthotto, L., & Haber, E. (2019). Deep neural networks motivated by PDEs. *Journal of Mathematical Imaging and Vision*.
10. Long, Z., Lu, Y., Ma, X., & Dong, B. (2018, July). Pde-net: Learning pdes from data. In *International conference on machine learning* (pp. 3208-3216). PMLR.
11. Weinan, E. (2017). A proposal on machine learning via dynamical systems. *Communications in Mathematics and Statistics, 5*(1), 1-11.
12. Haber, E., & Ruthotto, L. (2017). Stable architectures for deep neural networks. *Inverse problems, 34*(1), 014004.
13. Pope, S. B. (2000). *Turbulent Flows*. Cambridge University Press.
14. Lions, J. L. (1969). " Quelques Méthodes de Résolution des Problèmes aux Limites Non-Linéaires,". *Dunod*.
15. Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422-440.

Supplement

Hybrid Constraint Dynamics for Hallucination Reduction in Artificial Intelligence

Abstract
We introduce a mathematically constrained framework for large language model generation based on dissipative dynamics, spectral consistency, and energy minimization. Hallucination is formalized as instability of unconstrained evolution in representation space. A constraint operator acting on generated outputs is defined and analyzed. Under bounded precision and energy assumptions, we prove that the operator suppresses unstable modes associated with unsupported statements. The framework does not claim completeness or correctness guarantees, but establishes provable reduction of hallucination frequency in decidable subdomains.

Keywords: Large Language Models, Constraint operator, Dissipative Dynamics, Spectral Stability, Energy Minimization

Introduction
Large language models generate sequences by optimizing conditional likelihood, a criterion that does not enforce semantic or logical validity. As a result, models may produce internally consistent yet ungrounded statements, commonly referred to as hallucinations [1–3]. Rather than treating hallucination as a data deficiency, we adopt a structural viewpoint: hallucination arises from unstable modes in high-dimensional generative dynamics.

This paper proposes a constraint-based formalism inspired by mathematical physics. We introduce a deterministic operator acting on candidate outputs, designed to suppress instability through dissipative smoothing, spectral regularity, and bounded energy principles. The presentation emphasizes formal definitions and provable properties, in accordance with the style of *Communications in Mathematical Physics*.

Preliminaries
Let H be a separable Hilbert space representing internal model states, and let Y denote the space of finite symbolic outputs. A language model induces a map

g: H x C-> Y,

where C denotes contextual inputs. We assume that each output y ∈ Y admits an associated reasoning representation R(y)∈H .

Definition of the Constraint Operator
Definition 3.1 (Constraint Operator)
Let R(y) ∈ H. The constraint operator K : Y-> Y U {ø} is defined by

k (y) ={y, if y satisfies dynamical,spectral,and energy constraints,
{ø,otherwise.

The constraints are specified below.

Dynamical Stability Constraint
We define a dissipative flow on H:

∂_t h = -A h + F(h),

where A is a positive self-adjoint operator with compact inverse and v > 0. The stabilized representation h- is the long-time limit of this flow when it exists.

Lemma 4.1 (Mode Suppression)
Let {λk} *denote the eigenvalues of A*. *Then for any initial state h*_0∈ H, *high-frequency components decay exponentially*:

$||P>_N h(t)||$ e^{-vλ_{N+1} t} l|h0|l.

Proof. Standard semigroup estimates for dissipative operators [5,6].

Spectral Consistency Constraint
Let T(y) be a bounded linear operator encoding inferential transitions of y. Denote its spectrum by σ(T(y)).

Definition 5.1 (Spectral Admissibility)
An output y is spectrally admissible if

$\sup\{|\lambda| : \lambda \in \sigma(T(y))\} \leq M$,] for a fixed constant $M > 0$.

Proposition 5.2
Spectral inadmissibility implies exponential sensitivity of reasoning trajectories.
Proof. Follows from the spectral radius formula and perturbation theory [7,8].

Energy Constraint
Define the energy functional [$E(y) = |A^{1/2} R(y)|^2 + \alpha|R(y)|^2$.]

Definition 6.1 (Energy Boundedness)
An output (y) is energy admissible if $E(y) \leq E^{max}$.

Main Result
Theorem 7.1 (Quantitative Hallucination Suppression)
Assume that the operator A satisfies the assumptions of Appendix A, and fix thresholds (nu, M, E_max). Let h(t) be the stabilized flow associated with an output y. Then there exist constants C > 0 and gamma > 0, depending only on nu and the spectral gap of A, such that the unstable component satisfies

$$||P_{>N} h(t)|| \leq C e^{-\gamma t} ||h(0)||.$$

Consequently, any output y violating spectral or energy admissibility is eliminated by the constraint operator K, and within any decidable subdomain the hallucination rate is reduced by at least a factor proportional to $e^{-\gamma t}$.

Proof. By Lemma 4.1 and the spectral gap assumption on A, we obtain exponential decay with $\gamma = \nu \lambda_{N+1}$. Energy admissibility ensures boundedness of the remaining modes.

Limitations
The operator K does not guarantee correctness, nor can it resolve undecidable propositions or open problems. Its effect is purely restrictive and does not expand the computational power of the underlying model [10–12].

Conclusion
We have formulated a constraint operator grounded in dissipative dynamics and spectral theory, establishing quantitative suppression of unstable generative modes. The resulting decay estimates place hallucination reduction within a rigorous mathematical framework compatible with semigroup methods in mathematical physics.

Appendix A. Functional-Analytic Setting
We summarize the assumptions required for the operator A.

Assumption A.1
The operator A: $D(A) \subset H \to H$ is positive, self-adjoint, and has compact inverse. Its spectrum consists of eigenvalues

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots, \lambda_k \to \infty.$$

Assumption A.2
The semigroup e^{-tA} is analytic and satisfies

$$||A^\alpha e^{-tA}|| \leq C t^{-\alpha} e^{-\lambda_1 t}, \alpha \geq 0.$$

These assumptions are standard for elliptic operators on bounded domains.

Appendix B. Relation to Navier–Stokes Semigroups
Let A be identified with the Stokes operator on a bounded domain with appropriate boundary conditions. Then the dissipative flow in Section 4 coincides with the linearized Navier–Stokes semigroup:

$$\partial_t h + \nu A h = 0.$$

The exponential decay obtained in Theorem 7.1 mirrors classical energy decay estimates for viscous incompressible flows. The present framework relies only on semigroup properties of the linear operator.

Conflict of interest
There is no conflict of interest.

Data availability statement
Data openly available in a public repository that issues datasets with DOIs.

References

1. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?
   Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT), 2021.
2. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM computing surveys, 55*(12), 1-38.
3. Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. *arXiv preprint arXiv:2005.00661*.
4. Marcus, G. (2020). The next decade in AI: four steps towards robust artificial intelligence. *arXiv preprint arXiv:2002.06177*.
5. Temam, R. Navier–Stokes Equations: Theory and Numerical Analysis.
   American Mathematical Society, Providence, RI, 2001.
6. Lions, P. L. (1996). *Mathematical topics in fluid mechanics: volume 2: compressible models* (Vol. 2). oxford university press.
7. Reed, M., & Simon, B. Methods of Modern Mathematical Physics I: Functional Analysis.Academic Press, New York, 1980.
8. Tao, T. (2012). Topics in random matrix theory. American Mathematical Society. *graduate Studies in Mathematics, 132*.
9. Goldstein, H., Poole, C., & Safko, J. (1980). Classical mechanics addison-wesley. *Reading, MA*, 426, 136.
10. Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für mathematik und physik, 38*(1), 173-198.
11. Baker, T., Gill, J., & Solovay, R. (1975). Relativizations of the P=?NP question. *SIAM Journal on computing, 4*(4), 431-442.
12. Razborov, A. A., & Rudich, S. Natural Proofs. Journal of Computer and System Sciences 55 (1997), 24–35.
13. Gehrmann, S., Adewumi, T., Aggarwal, K., Ammanamanchi, P. S., Aremu, A., Bosselut, A., ... & Zhou, J. (2021, August). The gem benchmark: Natural language generation, its evaluation and metrics. In P*roceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)* (pp. 96-120).
14. Chen, M. (2021). Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*.
15. Aaronson, S., & Wigderson, A. (2009). Algebrization: A new barrier in complexity theory. *ACM Transactions on Computation Theory (TOCT), 1*(1), 1-54.