全新架构的高精度可训练 PDE 算子：AI 参数数量存在唯一性的数学证明与实验验证

杜秋实
湘潭大学
202205191101@smail.xtu.edu.cn

April 7, 2026

Abstract

AI 模型普遍存在分布外（OOD）泛化产生幻觉的问题，难以用于严格科学计算。现有物理信息神经网络（PINNs）、傅里叶神经算子（FNOs）等主流方案均为在工程层面的软正则约束提供物理先验，未能突破 AI 底层的拓扑结构缺陷。

在机器学习的流形假设（Manifold Hypothesis）中：所有的 AI 模型，本质上就是一种根据输入数据拟合特定高维流形的可训练算子。而这类高维函数逼近与全局收敛性问题的数学基础，早在 1943 年，由 Courant 奠定的古典伽辽金有限元变分理论中得到系统性建立与严格证明。

以此为基础，本文融合了有限元与线性代数理论，证明了 AI 参数数量的存在唯一性：为彻底消除非平凡零空间与 OOD 幻觉，AI 参数数量必须严格等于 AI 训练集下 Galerkin 投影的非线性基底数，即 $N_{AI} = \dim(V_h) = N_{basis}$(其中 V_h 是 Galerkin 投影的有限维子空间，N_{basis} 是该空间内的非线性基底数)。

数值实验结果表明，全新架构仅需 $O(1)$ 的可训练参数与简单的 ADAM 梯度下降方法，即可实现 FP64 双精度浮点极限精度的零幻觉 OOD 泛化：$MSELoss = O(10^{-32})$，彻底解决了 OOD 泛化问题。

Keywords:
Computational Mathematics; FEM; Linear Algebra; PDE
Deep Learning; AI4Science; Hallucination

1 引言

1.1 行业现状

现有主流深度学习模型，均建立在极大似然估计（Maximum Likelihood Estimation, MLE）驱动的自回归生成式建模与统计拟合底层范式之上。图灵奖得主 LeCun[1] 与认知科学泰斗 Marcus[2] 先后从理论与实证层面指出，基于 MLE 的过参数化神经网络本质上是统计匹配器，而非对物理规律、因果结构与微分算子的真正理解；自回归生成过程中，单步预测的微小误差会随推理深度呈指数级累积，而过参数化架构导致的信息不可逆丢失，进一步加剧了模型的虚假解生成（幻觉）与逻辑推理失效。从线性代数与代数拓扑的视角来看，现有过参数化模型上述缺陷的数学根源，在于模型参数空间到输出空间映射的雅可比矩阵中存在高维非平凡零空间（Null Space，核空间）。Sagun 等[3] 与 LeCun 合作的奠基性研究严格证明，过参数化神经网络存在无限维零空间，是神经网络幻觉、信息丢失与泛化能力退化的核心诱因；Qin 等[4]针对 Transformer 架构的研究进一步发现，自注意力矩阵存在固有低秩壁垒 (inherent low-rank barrier)，其零空间维度随输入序列长度的增长呈爆炸式扩张 (explosive expansion)，从数学上解释了 Transformer 在长程逻辑推理与全局信息整合中的固有失效问题。后续系列研究[5][6][7][8][9][10][11][12]从线性变换可逆性 (invertibility of linear transformation)、模型表达能力边界 (fun- damental limits of model expressiveness)、物理世界符号接地性 (physical symbol grounding)、过拟合量化表征 (quantitative characterization of overfitting) 等多个核心维度，系统证实了非平凡零空间是现有过参数化 AI 架构无法突破的内在理论壁垒：非平凡零空间的存在直接导致模型前向映射不可逆 (irreversibility of model forward mapping)；纯文本训练的 LLMs 零空间覆盖了几乎所有物理世界的时空与因果维度，使其无法实现语言符号与物理实体的有效接地；而单纯的模型规模扩增与调优策略，均无法从根本上消除零空间带来的信息丢失问题。

为解决上述科学计算场景中的核心痛点，人工智能赋能科学计算（Artificial Intelligence for Science, AI4Science）领域的前沿研究提出了物理信息神经网络（Physics-Informed Neural Networks, PINNs）、傅里叶神经算子（Fourier Neural Operators, FNOs）等主流建模范式，尝试将偏微分方程（Partial Differential Equation, PDE）控制方程、物理守恒定律作为软约束（soft constraints）嵌入神经网络架构，以期增强模型的物理一致性与分布外（Out-of-Distribution, OOD）泛化能力。

然而，现有研究证实，此类方法并未突破过参数化（over-parameterized）神经网络的底层拓扑架构，无法从根本上解决传统 MLE 范式的内生性固有缺陷，仍存在严重的 OOD 泛化崩溃（generalization collapse）问题：

Krishnapriyan 等[13] 的顶会研究明确指出，PINNs 在面对简单的对流方程（advection equation）求解任务时，会因过参数化导致的复杂损失地形（Loss Landscape）出现优化失效（optimization failure）问题；Wang 等[14] 进一步揭示了 PINNs 中普遍存在的梯度病态性（Gradient Pathologies）现象：多源损失项（multi-source loss terms）的相互竞争会导致梯度反向传播（back-propagation of gradients）过程中的信号湮灭（signal annihilation），使模型无法收敛至符合物理约束的全局最优解（global optimal solution）；Lu 等[15] 针对神经算子的大样本公平基准对比研究也证实，包括 FNOs 在内的主流神经算子架构，在极端 OOD 场景中均会出现显著的泛化性能衰减（generalization performance degradation）。上述研究一致表明，仅引入物理正则约束而不从根本上改变过参数化的底层拓扑结构，无法彻底解决模型的虚假解生成（spurious solution generation, 即 AI 幻觉 hallucination）与泛化崩溃问题。

1.2 核心问题

纵观现有研究，学术界已充分认识到过参数化深度学习范式的根本性缺陷，并从统计学习、线性代数与计算数学（computational mathematics）等多个维度开展了系统的缺陷归因分析。然而，现有研究仍存在三大核心理论空白，制约了 AI 方法在高可靠科学计算领域的落地应用：

其一，现有研究仅证实了非平凡零空间（non-trivial null space）与模型缺陷的统计关联性，尚未基于线性代数基本定理严格证明非平凡零空间是 AI 模型 OOD 幻觉产生的充要条件（necessary and sufficient condition），无法从数学上严格锁定 OOD 幻觉产生的内生性本质根源；

其二，现有研究尚未从理论上推导并证明 AI 模型可训练参数量选取的唯一性准则（uniqueness criterion），无法回答“为彻底根除非平凡零空间与 OOD 幻觉，AI 模型可训练参数量的最优取值应满足何种数学准则”这一核心问题，导致 AI 架构的参数设计仍停留在“试错式调参”（heuristic parameter tuning，业内俗称“炼丹”）的工程化阶段，缺乏确定性的数学判据；

其三，现有研究大多停留在理论批判与局部工程化修补层面，尚未提出能够从拓扑结构上彻底消除非平凡零空间、实现训练损失与 OOD 损失同阶收敛（same-order convergence）的可实现 AI 架构。与此同时，现有 AI 架构设计普遍忽视了古典计算数学中已被严格证明的变分逼近理论（variational approximation theory）与有限元 (FEM) 思想——早在 1943 年，Courant[16] 首次提出通过分段连续函数（piecewise continuous functions）逼近变分问题（variational problem）的核心思想，奠定了有限元方法的理论基石；Strang 与 Fix[17] 建立的有限元收敛性理论与 Céa 引理（Céa’s Lemma），进一步严格证明了当离散基底与系统内禀泛函（intrinsic

functional）满足同构条件（isomorphism condition）时，局部单元的适定性（well-posedness）可严格保证全局解在拓扑流形上的收敛性，为解决上述核心理论空白提供了坚实的数学理论支撑。

1.3 本文创新点

针对上述研究空白，本文彻底摒弃了基于 MLE 的过参数化深度学习范式，直击人工智能概念的本质：所有人工智能本质上都可视为将输入数据映射到特定高维流形的可训练算子，将古典伽辽金（Galerkin）有限元理论、等参变换以及线性代数中的秩-零化度定理与 AI 训练进行交叉结合，提出了参数规模与非线性物理基底数量耦合的白盒人工智能架构——Pure Science AI，并通过严格的数学证明与系统的数值实验，验证了该架构在非线性 PDE 算子训练中的零幻觉 OOD 泛化能力。

本文的主要创新与核心贡献如下：

（1） 基于秩 - 零化度定理与 Galerkin 投影子空间，证明了 AI 参数量选取的唯一性准则：为彻底消除非平凡零空间与 OOD 幻觉，AI 参数量必须严格等于训练集下 Galerkin 投影的非线性基底数，即 $N_{AI} = \dim(V_h) = N_{basis}$(其中 V_h 是 Galerkin 投影的有限维子空间，N_{basis} 是该空间内的非线性基底数)。为了方便行文论述，我们将此等式命名为 AI 参数与基底数等式, 为 AI 架构的参数设计提供了数学准则。

（2） 基于线性代数齐次线性方程组解空间理论，证明了 AI 模型 OOD 幻觉产生的充要条件是 AI 参数空间中大于零的非平凡零空间（$\dim(NullSpace) > 0$），明确了 AI 幻觉是拓扑结构缺陷而非优化问题，无法通过工程手段彻底消除。

（3） 基于 Céa 引理与等参变换的全局微分同胚特性，严格证明了遵循 AI 参数与基底数等式的架构满足 $O(TrainLoss) = O(OODLoss)$，实现了训练精度与 OOD 泛化精度的全局一致性，从拓扑结构上彻底根除了新 AI 的 OOD 幻觉，使之可用于严格的科学计算。

（4） 通过 Q4 双线性形函数、2D 泰勒 - 格林涡、高斯钟形曲线三类经典非线性物理流形的对照数值实验，验证了本文提出的 Pure Science AI 架构的性能：仅需个位数级别的可训练参数和训练集，搭配最简单的 Adam 梯度下降训练方法，即可实现 FP64 双精度浮点格式极限精度（MSE Loss 达 $O(10^{-}32)$）的 OOD 泛化。而共享 PDE 系统精确解析解的 $O(10^4)$ 参数量的传统多层感知机（MLP）对照组出现严重的 OOD 泛化误差（$O(10^{+1})$ 到 $O(10^{-3})$），为本文的理论证明提供了坚实的实验验证。

2 非线性系统与线性化的 AI 参数

Theorem 2.1. 状态非线性与 AI 参数线性的数学解耦：通过 Galerkin 投影，实现有限参数量计算和使用秩-零化度定理 (Rank-Nullity Theorem) 在下文参与证明的数学依据：

Proof. Galerkin 投影：在适当的定解条件下，PDE 系统的真解 $u(x, y)$ 存在于一个特定的 Sobolev 空间 V 中。该空间是无限维的，且存在一个可数的、完备的基函数 $\{\Phi_i\}_{i=1}^{\infty}$ 在实际计算中，通过 Galerkin 投影，将真解 u 投影到一个有限维的逼近子空间 $\mathsf{V}_h \subset \mathsf{V}$，得到数值近似解。

假设 PDE 方程为：

$$\mathsf{L}(u) = f \quad \text{in } \Omega$$

其中 L 是包含非线性项（如对流项 $u \cdot \nabla u$）的微分算子。

引入测试函数 $v \in V$，获得弱形式：

$$\int_{\Omega} \mathsf{L}(u) v \, dx = \int_{\Omega} f v \, dx$$

使用 Galerkin 投影，将无限维空间 V 投影到有限维子空间 V_h。设 V_h 的非线性基底为 $\{\Phi_1, \ldots, \Phi_N\}$。

将试探解 u_h 和测试函数 v_h 展开为这组基底的线性组合：

$$u_h(x) = \sum_{j=1}^{N} a_j \Phi_j(x), \qquad v_h(x) = \Phi_i(x)$$

代入弱形式，得到：

$$\int_{\Omega} \mathsf{L}\left(\sum_{j=1}^{N} a_j \Phi_j(x) \right) \Phi_i(x) \, dx = \int_{\Omega} f \, \Phi_i(x) \, dx \quad \text{for } i = 1, \ldots, N$$

其离散近似解 u_h 可表示为一组有限数量的非线性物理基函数 $\Phi_j(X)$ 的线性组合：

$$u_h(x, y) = \sum_{j=1}^{N} a_j \Phi_j(x, y)$$

于是，PDE 系统的状态非线性完全由非线性基底 $\Phi_j(x, y)$ 提供，AI 仅需训练拟合非线性基底的系数 $\sum_{j=1}^{N} a_j$，即 AI 的参数空间 N 满足 $\sum_{j=1}^{N} a_j \in \mathrm{N}$

这种解耦确保了 AI 与 PDE 系统为线性映射关系，提供了秩-零化度定理 (Rank-Nullity Theorem) 的数学基础。 □

下面，我们将 Theorem2.1 的局部性质推广到全局性质

Theorem 2.2. 等参变换和拉回：将局部性质推广到全局，Pure Science AI 在测试集中训练出的映射具有全局不变性

Proof. 引入等参变换和拉回:

设定真实的全局计算域为 Ω，其由任意形状的畸变局部单元 Ω_e 构成，物理坐标记为 $\mathbf{x} = (x, y)$。

引入一个不变的标准参考单元（Reference Element）$\hat{\Omega}$（例如 $[-1,1] \times [-1,1]$），参考坐标记为 $\boldsymbol{\xi} = (\xi, \eta)$。

存在一个微分同胚的几何映射 $F_e : \hat{\Omega} \to \Omega_e$，使得：

$$\mathbf{x} = F_e(\boldsymbol{\xi}) = \sum_{k=1}^{N_{\text{geo}}} \mathbf{x}_k^{(e)} \hat{\Psi}_k(\boldsymbol{\xi})$$

其中 $\mathbf{x}_k^{(e)}$ 是真实物理网格的节点坐标，$\hat{\Psi}_k$ 是参考空间 $\hat{\Omega}$ 映射的雅可比矩阵 上的几何形状函数。（Jacobian）定义为：

$$J_e = \frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}}$$

根据等参变换原理，物理空间上的基底函数 $\Phi_j^{(e)}(\mathbf{x})$ 与参考空间上的标准基底 $\hat{\Phi}_j(\boldsymbol{\xi})$ 是同一个函数在不同坐标下的表示：

$$\Phi_j^{(e)}(\mathbf{x}) \equiv \hat{\Phi}_j(\boldsymbol{\xi}(\mathbf{x}))$$

因此，任意物理单元上的离散解 $u_h^{(e)}$，均为同一组参考空间上标准基底 $\hat{\Phi}_j(\boldsymbol{\xi})$ 的线性组合，代入可得：

$$u_h^{(e)}(\mathbf{x}) = \sum_{j=1}^{N} a_j^{(e)} \Phi_j^{(e)}(\mathbf{x}) = \sum_{j=1}^{N} a_j^{(e)} \hat{\Phi}_j(\boldsymbol{\xi})$$

在 PDE 系统中，我们要计算基底的空间梯度。根据链式法则引入雅可比矩阵（Jacobian），物理梯度的矩阵变换为：

$$\nabla_{\mathbf{x}} \Phi_j^{(e)}(\mathbf{x}) = J_e^{-T} \nabla_{\boldsymbol{\xi}} \hat{\Phi}_j(\boldsymbol{\xi})$$

将其代入上文中的弱形式：$\int_\Omega \left(\sum_{j=1}^{N} a_j \Phi_j(x) \right) \Phi_i(x)\, dx = \int_\Omega f\, \Phi_i(x)\, dx \quad \text{for } i = 1, \ldots, N$

组装单元刚度矩阵：

$$A_{ij}^{(e)} = \int_{\Omega_e} \nabla_{\mathbf{x}}\Phi_i \cdot \nabla_{\mathbf{x}}\Phi_j \, d\mathbf{x} = \int_{\hat{\Omega}} \left(J_e^{-T} \nabla_{\xi}\hat{\Phi}_i \right) \cdot \left(J_e^{-T} \nabla_{\xi}\hat{\Phi}_j \right) \det(J_e) \, d\boldsymbol{\xi}$$

其中：

J_e^{-T} 和 $\det(J_e)$ 包含了所有的变换性质，是任意坐标空间提供的确定性张量，不需要 AI 去学习。

$\nabla_{\xi}\hat{\Phi}_i$ 和 $\nabla_{\xi}\hat{\Phi}_j$ 分别是参考单元中的测试函数和基底。

对于参考单元中的系数 $a_j^{(e)}$，在组装单元刚度矩阵 $A_{ij}^{(e)}$ 的过程中，系数 $a_j^{(e)}$ 并不在表达式 $A_{ij}^{(e)}$ 中。这意味着 $A_{ij}^{(e)}$ 的计算仅依赖于基函数的选取与单元的几何（通过雅可比矩阵体现），而与系数 $a_j^{(e)}$ 无关。因此，系数 $a_j^{(e)}$ 并不随物理坐标 $\mathbf{x}$ 变化，而是基底本身线性组合的系数。而在上文等参变换的公式 $\Phi_j^{(e)}(\mathbf{x}) \equiv \hat{\Phi}_j(\boldsymbol{\xi}(\mathbf{x}))$ 中，系数 $a_j^{(e)}$ 作为组合权重，它在映射前后始终是同一个数，大小并不改变。

□

我们根据上述证明，得出以下四个重要结论：

Corollary 2.3. 1. 传统大模型在 $\mathbf{x}$ 空间 (真实的全局计算域) 尝试拟合，因此无法拟合雅可比矩阵 J_e 产生的无限数量的变化。Pure Science AI 在 $\boldsymbol{\xi}$ 空间（参考单元内）直接对线性系数 a_j 建立算子的线性映射，以有限元作为根基，在数学底层规避全局坐标畸变与介质非均匀性：在参考单元上训练出的系数系数 $a_j^{(e)}$，在映射后仍不改变大小，即 AI 参数在等参变换前后不发生改变。

2.Pure Science AI 仅需要一组有限个非线性基底的系数作为自身参数，组装成离散近似解 $u_h(x,y) = \sum_{j=1}^{N} a_j \Phi_j(x,y)$，即可完成对全局非线性函数的拟合（第二章实验的根基之一）

3.Pure Science AI 仅需要少量参数作为训练集，即可完成对局部的训练，从而完成对全局非线性函数的拟合（第二章实验的根基之一）

4. 有时我们无法获得 PDE 系统的精确解析解，进而导致伽辽金投影的非线性基底选取的很差，例如 train loss 只有 $O(10^{-3})$，但是等参变换下的拓扑双射仍然保证了 $O(\text{trian Loss}) \equiv O(\text{oodLoss})$(下面给出证明)，在低精度要求的计算任务中，$O(10^{-3})$ 全局拟合精度依然可用，即 Pure Science AI 具有误差上限，可直接用于科学计算。

Lemma 2.4. 对重要结论 4 的补充证明，即在 Pure Science AI 的架构下，其O(Train Loss)=O(OOD Loss) 恒成立，这是由等参变换和拓扑同构的性质决定的，与伽辽金投影的基底选取、测试集的训练精度等实现手段无关。有限维空间的数值误差，被该空间内最佳逼近误差所绝对控制

在经典的伽辽金有限元或谱方法中，Céa 引理（或其在非椭圆问题中的广义形式 Babuška 理论）为数值解提供了严格的误差界。

设真实解 $u \in V$ 满足变分问题 $\mathsf{B}(u, v) = \mathsf{L}(v), \forall v \in V$（假设双线性型 B 满足连续性与强制性/inf-sup 条件）。架构的解空间 $V_h \subset V$ 定义为由 N 个正交基底张成的有限维空间，其参数化解形式为：

$$u_h(x, y) = \sum_{j=1}^{N} a_j \Phi_j(x, y)$$

根据 Céa 引理，全局连续误差被最佳逼近误差所控制：

$$\|u - u_h\|_{H^1(\Omega)} \leq C \inf_{v \in V_h} \|u - v_h\|_{H^1(\Omega)}$$

在参数化逼近（如机器学习优化）中，模型并非通过解析投影求解，而是通过在有限训练点集 $\Omega_{\text{train}} \subset \Omega$ 上极小化经验残差（Empirical Risk）来获取系数向量 $\mathbf{A} = [a_1, a_2, \ldots, a_N]^T$。

如果 Ω_{train} 中的数据点分布满足对于基底空间 V_h 是一个适定集（Unisolvent Set），即评估矩阵（或雅可比矩阵）$\mathbf{J}_{ij} = \Phi_j(x_i, y_i)$ 满足列满秩条件（$\text{rank}(\mathbf{J}) = N$），则方程系统无非平凡零空间：

$$\mathbf{J}\Delta\mathbf{A} = \mathbf{0} \implies \Delta\mathbf{A} = \mathbf{0}$$

这意味着，只要局部离散损失 $\text{Loss}_{\text{train}}$ 收敛至极小值，参数组 $\{a_j\}_{j=1}^{N}$ 在整个物理域 Ω 上即被唯一确定。

定义未知的预测区域（Out-of-Distribution 域）为 $\Omega_{\text{OOD}} = \Omega \setminus \Omega_{\text{train}}$。在过参数化的传统深度学习模型（$N \gg N_{\text{basis}}$）中，由于雅可比矩阵严重秩亏，存在庞大的零空间，导致 Ω_{OOD} 上的函数值完全不受约束而发散。（本文在下文中会更加详细的证明）

但在本架构下，由于参数已被 Ω_{train} 唯一映射，全局解析解形态已确定。根据范数的单调性，局部区域的误差范数必然受制于全局误差范数：

$$\|u - u_h\|_{H^1(\Omega_{\text{OOD}})} \leq \|u - u_h\|_{H^1(\Omega)}$$

结合 Céa 引理，得到 OOD 区域的误差上限约束：

$$\|u - u_h\|_{H^1(\Omega_{\text{OOD}})} \leq C \inf_{v \in V_h} \|u - v_h\|_{H^1(\Omega)}$$

这表明，预测区域的误差严格受到基底空间 V_h 本身逼近能力的控制。

在复杂的实际 PDE 求解任务中，我们往往只能获取一个非完美的基底空间 V_h^{poor}。令 $\inf_{v_h \in V_h^{\text{poor}}} \|u - v_h\|_{H^1(\Omega)} = E_{\text{error}} > 0$。

在满足离散-连续范数等价性（Marcinkiewicz-Zygmund 不等式）的条件下，训练集上的离散最小化损失 $\min \text{Loss}_{\text{train}}$ 会趋近于最佳逼近误差 E_{error}。代入上述不等式可得：

$$\text{Loss}_{\text{OOD}} \approx \|u - u_h\|_{H^1(\Omega_{\text{OOD}})} \leq C \cdot E_{\text{error}} \approx C \cdot \min \text{Loss}_{\text{train}}$$

即证明了在非完美的基底空间 V_h^{poor} 下，$O(\text{Loss}_{\text{OOD}}) = O(\text{Loss}_{\text{train}})$ 仍然成立

3 非平凡零空间和幻觉

上一章的证明将非线性系统的 AI 参数进行线性化，我们引入线性代数中的秩-零化度定理

Theorem 3.1. $\dim(N) = \dim(\text{Image}) + \dim(\text{Null Space})$

Proof. AI 仅需对非线性基底的系数做线性映射。对于任何一个线性映射 $T: \mathbf{R}^n \to \mathbf{R}^m$，有秩-零化度定理（Rank-Nullity Theorem）：定义域 V 的维数等于核空间 $Ker(T)$ 的维数与值域 $Ran(T)$ 的维数之和，

即 $dim(V) = \dim(Ker(T)) + dim(Ran(T))$

AI 训练可视为训练从输入数据到输出数据的映射 $f: X \to Y$。在物理场景中，给定输入观测数据 X_{data}，前向传播实际上构成了一个由参数空间映射到物理观测空间 $Image$ 的映射 T：

$$\mathsf{T}_{X_{\text{data}}}: \mathbf{R}^{\dim(N)} \to \mathbf{R}^{\dim(\text{Image})}$$

其中 $\mathsf{T}_{X\text{data}}(W) = \Phi W$，$Y = \Phi(X_{\text{data}})W$，$WAI$，$\Phi$ 是训练集下 Galerkin 投影的非线性基底。

因为输入观测 X_{data} 与基底 Φ 在优化瞬间是固定的矩阵，那么该映射 $\mathsf{T}_{X\text{data}}(W) = \Phi W$ 关于我们要优化的参数空间 W 是严格的线性映射（Linear Mapping）

因此, 我们就可以使用秩-零化度定理：

$$\dim(N) = \dim(\text{Image}) + \dim(\text{Null Space})$$

物理意义：AI 的总参数量 $\dim(N)$，等于它能拟合出的有效物理特征维度 $\dim(\text{Image})$，加上零空间维度 $\dim(\text{Null Space})$。

□

Theorem 3.2. $\dim(N) = \text{Rank}(\text{Image})$,$\dim(\text{Null Space}) \equiv 0$

Proof. 求解线性方程组 $AX_1 = b_1$，所要求解的 X_1 就是 AI 训练出来的权重。每次求解方程组就是相当于 AI 对训练集数据的一次观察和对应输出，矩阵 A 就相当于观测矩阵（也就是训练集），b_1 就是 AI 输出的拟合数据，

PDE 系统的真实参数 X_{true}：这是大自然生成这段流体所用的“真实系数”。PDE 系统的真实输出是 $AX_{\text{true}} = b_{\text{true}}$

那么 AI 训练的本质就是根据 AI 输出的数据 b_1 和真实 PDE 系统的数据 b_{true} 进行对比，即求解 loss 值

$$Loss = ||b_2 - b_1||^{1/2}$$

降低 Loss 值, 就是让 AI 输出逼近 PDE 系统的真实输出

Loss 的是残差向量（Residual Vector）的范数：

$$R = b_{\text{true}} - b_{\text{pred}} = AX_{\text{true}} - AX = A(X_{\text{true}} - X)$$

令 $\Delta X = X_{\text{true}} - X$。这个 ΔX 代表 AI 的权重与 PDE 真解权重的距离

$$A\Delta X = R$$

当 AI 通过梯度下降训练，$Loss \to 0$ 时，即残差 $R \to 0$。

于是：

$$A\Delta X = 0$$

传统深度学习是过参数化的，它们的参数量 N 远远大于约束 M。在高等代数里，这体现为 A 的行数远远小于列数，即 $row(A) << col(A)$

这意味着矩阵 A 的列向量线性相关。

根据秩-零化度定理，齐次方程组 $A\Delta X = 0$ 存在基础解系，这意味着 ΔX 可以有无穷多个非零解。

代表 AI 的参数权重 X 的范数 $||\Delta X|| >> 0$，依有残差 $R = 0$（训练 Loss 降到了零）

若引入训练集分布外的测试集矩阵 A_{new} 进行 OOD（分布外泛化）训练，其中 $A_{\text{new}} \neq A$，则有

$$A_{\text{new}}\Delta X \neq 0$$

这在泛化测试中表示为，OOD 泛化幻觉

这就是大模型产生幻觉的数学本质，在观测矩阵 A（即训练集）下，Loss 训练无法影响到非平凡零空间（本文后续给出详细证明）

因此，非平凡零空间的存在，从数学上充要地决定了 OOD 测试集上幻觉的必然爆发。

我们必须且只能施加强制性的数学约束，避免幻觉产生的可能性：

即

$$\dim(\text{Null Space}) \equiv 0$$

代入

$$\dim(N) = \text{Rank(Image)} + \dim(\text{Null Space})$$

得到

$$\dim(N) = \text{Rank(Image)}$$

即，AI 参数量必须和观测矩阵 Image 的有效维度保持一致 □

下面我们证明，观测矩阵 Image 与 AI 参数所在空间的维数，即 Galerkin 投影的有限维子空间的维数保持一致。

即 Galerkin 投影的有限维子空间的基底数 $\dim(N_{basis})=\dim(\text{Image})$

Theorem 3.3. $\text{Rank(Image)} \geq \dim(N_{\text{basis}})$

Proof. 反证法（假设发生欠拟合）：

假设 $\text{Rank(Image)} < \dim(N_{\text{basis}})$，即 $r < m$。

根据线性代数的基础子空间理论，映射矩阵 A 的列空间 $C(A)$ 仅仅是 m 维空间 $\mathbf{R}^m$ 中的一个 r 维子空间（可视为一个低维的超平面）。

由于 $r < m$，必然存在属于大自然真实物理流形、但正交于 AI 列空间的维度。当大自然给出一个包含了这些维度的复杂物理状态 b 时，目标向量 b 不在矩阵 A 的列空间内，即：

$$b \notin C(A)$$

此时，方程 $Ax = b$ 无解

因此，$\text{Rank(Image)} \geq \dim(N_{\text{basis}})$ □

Theorem 3.4. $\text{Rank(Image)} \leq \dim(N_{\text{basis}})$

Proof. AI 的最终输出是被表达为非线性基底的线性组合的：

$$U_h(\mathbf{x}) = \sum_{j=1}^{N_{\text{basis}}} a_j \Phi_j(\mathbf{x})$$

1. 根据 Galerkin 投影原理，所有的基底函数 $\{\Phi_1, \Phi_2, \ldots, \Phi_{N_{\text{basis}}}\}$ 张成了一个有限维的试探子空间（Trial Space），我们记为 V_h。

根据线性代数基底的定义，这个空间的维度是确定的：$\dim(V_h) \equiv \dim(N_{\text{basis}})$。

2. AI 无论怎么生成预测结果，都绝对且只能是这 N_{basis} 个基底的某种线性组合。这意味着，AI 映射的像空间（Image），是试探空间 V_h 的一个子集（Subset）或子空间（Subspace）

即 $\text{Image(AI)} \subseteq V_h$

3. 子空间维度定理：任何有限维向量空间的子空间，其维度绝对不可能超过母空间。

因此，我们得出了这个不等式：

$$\text{Rank(Image)} = \dim(\text{Image}) \leq \dim(V_h) \equiv \dim(N_{\text{basis}})$$

□

Theorem 3.5. $\dim(N) = \dim(N_{\text{basis}})$

Proof. 收束证明，根据上文 theorem 3.1、3.2、3.3、3.4：

$$\dim(N) = \dim(\text{Image}) + \dim(\text{Null Space})$$

$$\dim(N) = \text{Rank(Image)}$$

$$\text{Rank(Image)} \geq \dim(N_{\text{basis}})$$

$$\text{Rank(Image)} \leq \dim(N_{\text{basis}})$$

得到

$$\dim(N) = \dim(N_{\text{basis}})$$

命题得证，$\dim(N) = \dim(N_{\text{basis}})$，即 AI 参数量具有存在唯一性 □

Lemma 3.6. 推论：AI 幻觉是拓扑结构缺陷，不是优化问题

Proof. 任意 AI 训练方法，其每次参数矩阵 N 权重更新，依赖于训练数据的观测矩阵 X_{data} 的行空间 (Row Space)。而零空间 (Null Space) 与行空间是正交补的。

非平凡零空间的存在是 AI 幻觉的充要条件

假设零空间中存在任意一个非零向量 $V_{\text{null}} \in N(A_{\text{train}})$。

即 $A_{\text{train}} V_{\text{null}} = \mathbf{0}$。

我们在参数空间中，沿着这个零空间方向移动一段距离 ϵ。：

$$L(W + \epsilon V_{\text{null}}) = f(A_{\text{train}}(W + \epsilon V_{\text{null}}))$$

$$= f(A_{\text{train}} W + \epsilon A_{\text{train}} V_{\text{null}})$$

$$= f(A_{\text{train}} W + \mathbf{0})$$

$$= L(W)$$

因此，AI 幻觉无法通过工程手段解决

□

4 实验验证

4.1 实验设计与公平性原则

4.1.1 控制变量

传统 AI 模型难以在极简的参数量和小规模训练集下，对高度非线性的 PDE 系统进行精准拟合

在对照组（Black Box AI）的设计中，我们保留了 AI 使用非线性解析基底进行参数拟合的架构。黑盒与白盒模型共享相同的输入、输出以及伽辽金投影组成的非线性基底。也就是说，黑盒 AI 使用了白盒 AI 相同的架构，只是将白盒 AI 的四个参数改成了 4 个独立的 MLP 头。

得到黑盒模型产生 OOD 泛化崩溃的原因就是过参数化:$O(10^4)$，以此验证 AI 参数量与伽辽金投影基底数等式：$N = N_{\text{basis}}$

4.1.2 PDE 系统的选取

本章选取了三种具有代表性的经典 PDE 用于验证

Q4 双线性形函数 这是有限元领域广泛使用的经典测试函数，而本文证明的大部分理论基础来自于有限元。

2D 泰勒-格林涡 这是 Navier-Stokes 方程的 2D 经典特殊解析解，直接且综合地体现了测试架构对高度非线性、多尺度 PDE 系统的适应与求解能力。

高斯钟形曲线 这是自然界中最普遍的格林函数，是热传导方程的基础解析解，在全局积分意义下体现出能量守恒。

4.1.3 训练集和测试集的选取

根据前文推导，AI 学习的参数本身具有空间不变性，在训练集学习到的线性组合系数仍是全局的线性组合系数。

因此，我们构建了极其苛刻的训练集和测试集环境：

在整个 PDE 系统全局 Ω 中，仅随机选取一个点 (x_0, y_0)，在其周围提取包含中心点在内的 9 个离散点作为唯一的训练集 Ω_{train}。

则 Ω 中除该极其狭小的 9 点 Ω_{train} 外的任何其他点，均为 OOD 测试集 Ω_{OOD}，即 $\Omega_{\text{OOD}} = \Omega - \Omega_{\text{train}}$

这种小尺度训练集和大尺度测试集的实验，构成了对 AI 模型泛化能力而非记忆能力的严格测试。

4.2 实验结果

4.2.1 Q4 双线性形实验

实验结果如下：

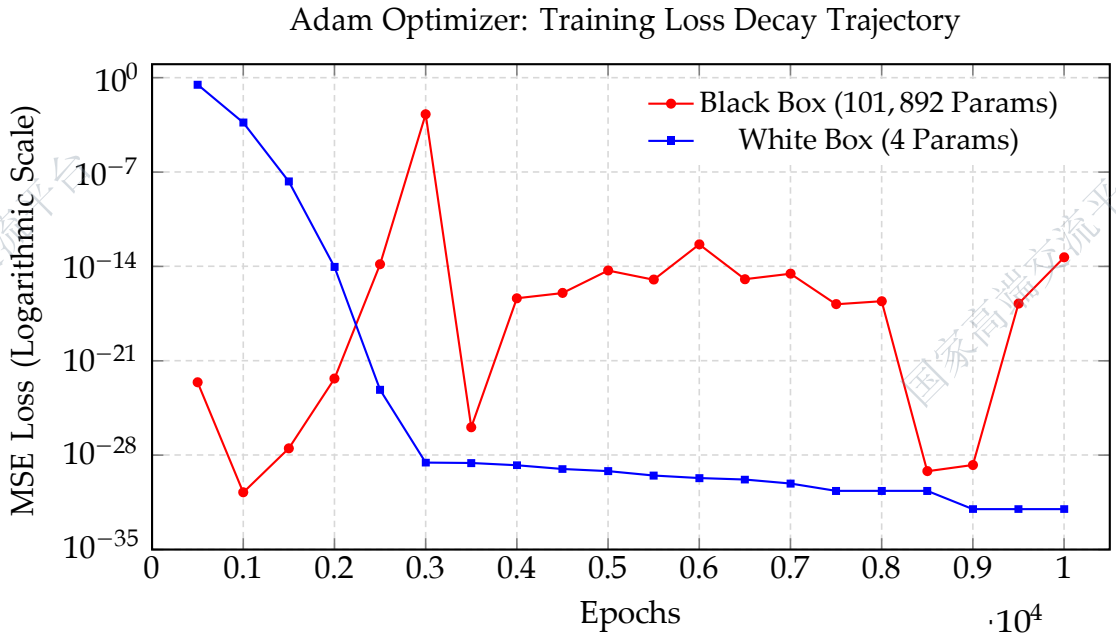

Figure 1: 白盒与黑盒的 train Loss 对比

Table 1: 最终结果

模型架构	可训练参数量	最终 Train Loss	最终 OOD Loss	泛化状态
Black Box MLP	101,892	4.65×10^{-14}	$\mathbf{3.57 \times 10^{+1}}$	OOD **幻觉**
White Box AI	4	**0.00**	**0.00**	**精准拟合**

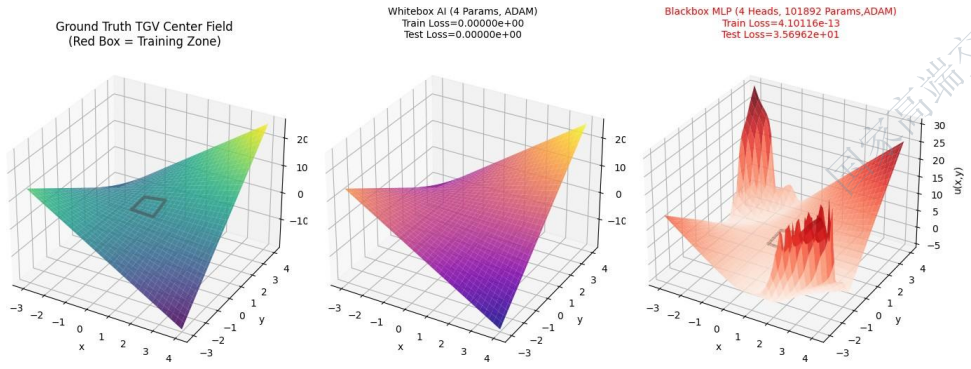

Figure 2: Q4 双线性形函数拟合对比结果的可视化图

4.2.2 2D 泰勒格林涡流

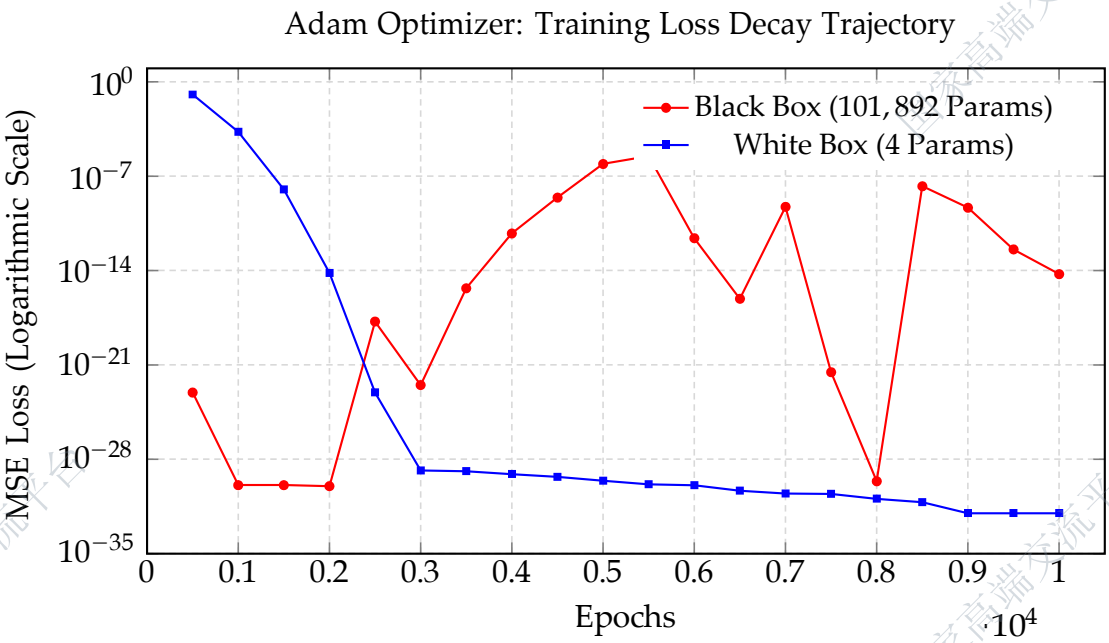

Figure 3: 白盒与黑盒的 train Loss 对比

Table 2: 最终结果

模型架构	可训练参数量	最终 Train Loss	最终 OOD Loss	泛化状态
Black Box MLP	101,892	7.23×10^{-15}	$\mathbf{2.05 \times 10^{+0}}$	OOD **幻觉**
White Box AI	4	**0.00**	**0.00**	**精准拟合**

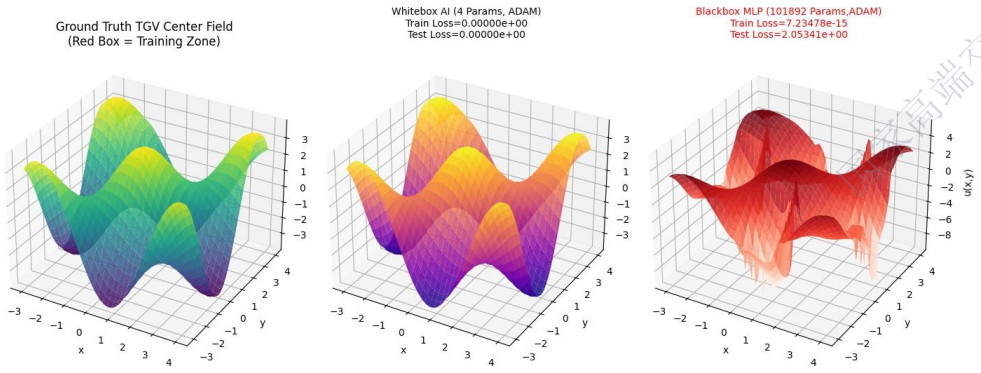

Figure 4: 在 2D 泰勒格林涡流中，Black Box AI ($O(10^4)$ 参数，4MLP 头) 产生严重 OOD 幻觉，而白盒仅用 4 参数完成精准拟合。

4.2.3 高斯钟形曲线

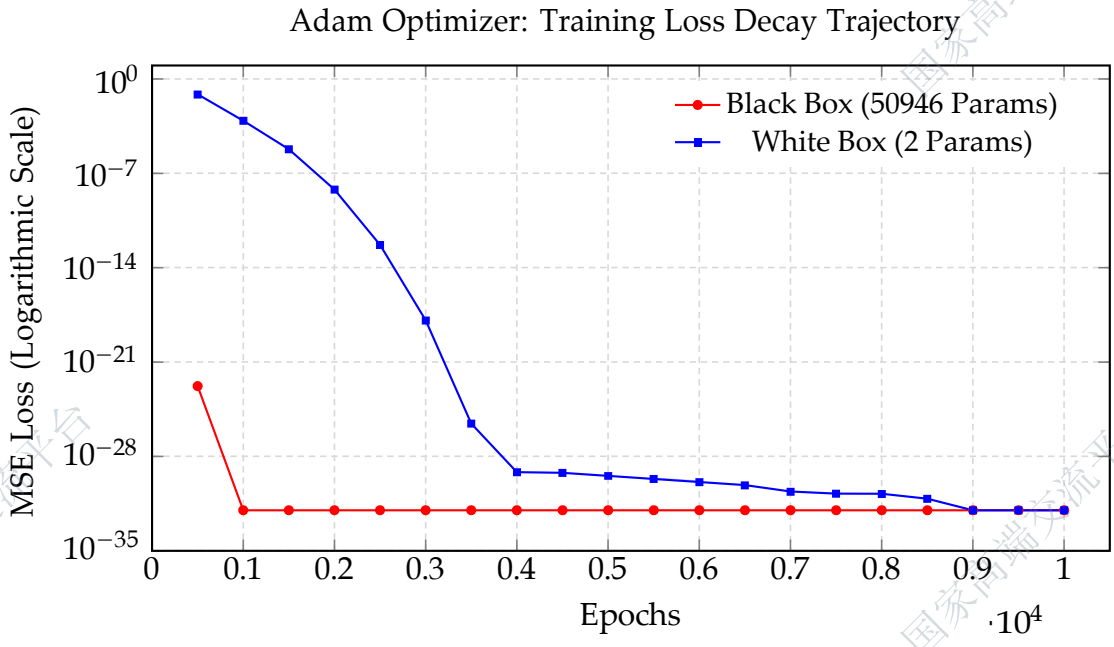

Figure 5: 白盒与黑盒的 train Loss 对比

Table 3: 最终结果

模型架构	可训练参数量	最终 Train Loss	最终 OOD Loss	泛化状态
Black Box MLP	101,892	1×10^{-32}	$\mathbf{2.07 \times 10^{+1}}$	OOD **幻觉**
White Box AI	2	**0.00**	**0.00**	**精准拟合**

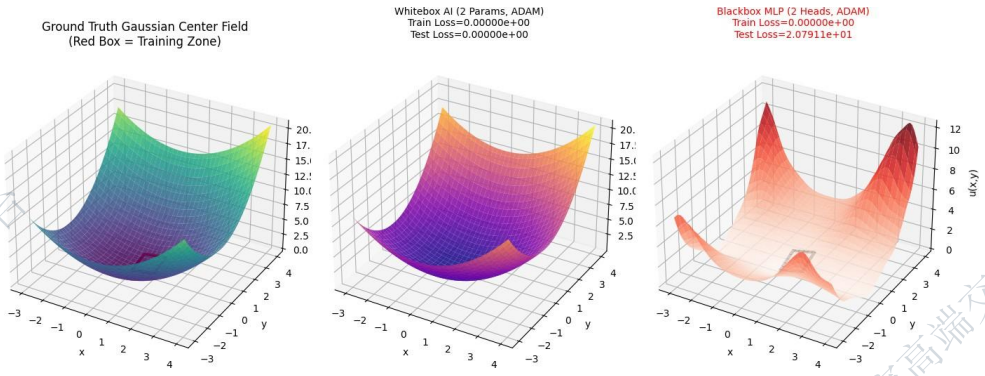

Figure 6: 在高斯钟形曲线中，Black Box AI ($O(10^3)$ 参数，2MLP 头) 产生严重 OOD 幻觉，而白盒仅用 2 参数完成精准拟合。

值得一提的是，在高斯函数实验中，黑盒架构抢先训练至 1×10^{-32} 的训练精度，但是仍然出现了非常严重的 OOD 幻觉 ($2.07 \times 10^{+1}$)，这更加印证了结论。

以上是本文的全部实验。

5 总结

本文通过将古典伽辽金变分理论与线性代数满秩条件深度融入 深度学习架构，严格证明了 $\dim(N) = \dim(N_{basis})$ 是彻底消除非平凡零空间的充 要条件。数值实验结果充分表明：在处理 2D 泰勒格林涡这种高度非线性的 NS 方 程时，遵循该准则的 Pure Science AI 架构仅需 O(1) 级别的参数量，即可在 OOD 区域实现 $O(10^{32})$ 双精度极限精度的零幻觉泛化。

与此同时，基于 Céa 引理与等参变换构建的理论框架，为本架构提供了严格的误差上界保证：即使人为选择的解析基底并不是完美精确的，$O(OODLoss) = O(TrainLoss)$ 的同阶收敛性依然成立。

PDE 求解仅是本架构的基础应用场景。从秩-零化度定理与信息几何等视角审视，无论是物理世界的演化规律，还是自然语言的复杂逻辑，其本质均可视为高维环境空间中嵌套的低维光滑流形，这一核心规律已被机器学习领域的流形假设（Manifold Hypothesis）广泛证实，LeCun、Bengio 等学者通过理论推导与可视化实验，系统性验证了 AI 模型的核心任务是学习高维数据背后的低维流形结构。

相较于现有依赖海量数据的统计学 MLE 范式，基于本文提出的算子同构思想，通过人工构建向量且保证满秩的“人工语义向量空间”，将该架构的核心设计逻辑从 PDE 求解拓展至 NLP 与大语言模型领域，有望从全领域中根除 AI 幻觉，为下一代无幻觉通用人工智能的研究提供全新的理论路径。

References

[1] LECUN Y. A path towards autonomous machine intelligence[EB/OL]. 2022 [2026-04-07]. https://openreview.net/forum?id=BZ5LxuD6OS.

[2] MARCUS G. Deep learning: A critical appraisal[EB/OL]. 2018 [2026-04-07]. https://arxiv.org/abs/1801.00631. arXiv: 1801.00631 [cs.LG].

[3] SAGUN L, GURBUZBALABAN M, MA Y A, et al. Ghosts in neural networks: Existence, structure and role of infinite-dimensional null space[EB/OL]. 2021 [2026-04-07]. https://arxiv.org/abs/2106.04770. arXiv: 2106.04770 [cs.LG].

[4] QIN Z, LI X, ZHANG Y, et al. On the limitation and redundancy of transformers: A rank perspective[EB/OL]. 2025 [2026-04-07]. https://openreview.net/forum?id=GoK1PfVgXU.

[5] LI X, SHORT K M. Null Space Properties of Neural Networks with Applications to Image Steganography[J/OL]. Mathematics, 2025, 13(21): 3394 [2026-04-07].

[6] KAMMAR O, LINDLEY S, ATKEY R. On limitations of the transformer architecture[EB/OL]. 2024 [2026-04-07]. https://arxiv.org/abs/2402.08164. arXiv: 2402.08164 [cs.LG].

[7] STORK D G, NEUMAN J, HILLIS W D. Lost in transmission: Language models do not realize when they fail to understand[EB/OL]. 2024 [2026-04-07]. https://arxiv.org/abs/2402.03114. arXiv: 2402.03114 [cs.CL].

[8] KUDITIPUDI S, WANG X, LEE H, et al. Dense neural networks are not universal approximators[EB/OL]. 2026 [2026-04-07]. https://arxiv.org/abs/2602.07618. arXiv: 2602.07618 [cs.LG].

[9] HU J Y C, CHEN H, LI Y. Fundamental limits of prompt tuning transformers: Universality, capacity and efficiency[EB/OL]. 2024 [2026-04-07]. https://arxiv.org/abs/2411.16525. arXiv: 2411.16525 [cs.LG].

[10] BURNS J, YE H, KLEIN A, et al. The inverse curse: LLMs fail to learn inverse relations[EB/OL]. 2023 [2026-04-07]. https://arxiv.org/abs/2309.12288. arXiv: 2309.12288 [cs.CL].

[11] MITROVIC A, KLIMOV O, REYNOLDS M, et al. Grounding language models to physical worlds: A zero-space perspective[C]//Proceedings of the

NeurIPS 2024 Workshop on Foundation Models and Physics. Vancouver: Curran Associates, Inc., 2024.

[12] WANG Y, ZHANG L, LIU J. Quantifying overfitting: Evaluating neural network performance through analysis of null space[EB/OL]. 2023 [2026-04-07]. https://arxiv.org/abs/2305.19424. arXiv: 2305.19424 [cs.LG].

[13] KRISHNAPRIYAN A, GHOLAMI A, ZHE S, et al. Characterizing possible failure modes in physics-informed neural networks[C]//Advances in Neural Information Processing Systems 34 (NeurIPS 2021). Virtual Event: Curran Associates, Inc., 2021.

[14] WANG S, TENG Y, PERDIKARIS P. Understanding and mitigating gradient pathologies in physics-informed neural networks[J/OL]. SIAM Journal on Scientific Computing, 2021, 43(5): A3055-A3081.

[15] LU L, MENG X, CAI S, et al. A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data[J/OL]. Computer Methods in Applied Mechanics and Engineering, 2022, 393: 114778.

[16] COURANT R. Variational methods for the solution of problems of equilibrium and vibrations[J/OL]. Bulletin of the American Mathematical Society, 1943, 49(1): 1-23.

[17] STRANG G, FIX G J. An Analysis of the Finite Element Method[M]. Englewood Cliffs: Prentice-Hall, 1973.