Novelty Assessment Report

Paper: Can Language Models Discover Scaling Laws?

PDF URL: https://openreview.net/pdf?id=TPTtWC0pGk

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-29

Abstract

Discovering scaling laws for predicting model performance at scale is a fundamental and open-ended challenge, mostly reliant on slow, case specific human experimentation. To investigate the potential for LLMs to automate this process, we collect over 5,000 experiments from existing literature and curate seven diverse scaling law discovery tasks. While existing agents struggle to produce accurate law formulas, this paper introduces SLDAgent, an evolution-based agent that co-optimize the scaling law model and the parameters, enabling it to autonomously explore complex relationships between variables. For the first time, we demonstrates that SLDAgent can automatically discover laws that exhibit consistently more accurate extrapolation than their established, human-derived counterparts across all tasks. Through comprehensive analysis, we elucidate why these discovered laws are superior and verify their practical utility in both pretraining and finetuning applications. This work establishes a new paradigm for agentic scientific discovery, showing that AI systems can understand their own scaling behavior, and can contribute novel and practical knowledge back to the research community.

Core Task Landscape

This paper addresses: **Automated Discovery of Scaling Laws for Language Model Performance**

A total of **50 papers** were analyzed and organized into a taxonomy with **33 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Empirical Scaling Law Characterization**
- **Architecture-Specific Scaling Laws**
- **Data-Centric Scaling Laws**
- **Training Method and Hyperparameter Scaling**
- **Capability-Specific Scaling Laws**
- **Predictive Modeling and Observational Methods**
- **Model Composition and Merging Scaling**
- **System-Level Scaling Analysis**
- **Theoretical and Information-Theoretic Scaling**
- **Large-Scale Model Development and Empirical Studies**
- ... and 3 more categories

Complete Taxonomy Tree

- Automated Discovery of Scaling Laws for Language Model Performance Survey Taxonomy
- Empirical Scaling Law Characterization
  - Fundamental Compute-Loss Scaling (2 papers)
  - [7] DeepSeek LLM: Scaling Open-Source Language Models with Longtermism (DeepSeek-AI, 2024) View paper
  - [14] Scaling Laws for Neural Language Models (Kaplan, 2020) View paper
  - Temporal and Training Dynamics Scaling (3 papers)
  - [2] Temporal scaling law for large language models (Xiong Yizhe, 2025) View paper
  - [32] Scaling with collapse: Efficient and predictable training of llm families (Bergsma, 2025) View paper
  - [42] A Multi-Power Law for Loss Curve Prediction Across Learning Rate Schedules (Luo Kairong, 2025) View paper
  - Inference-Time Compute Scaling (2 papers)
  - [4] Inference scaling laws: An empirical analysis of compute-optimal inference for problem-solving with language models (Yangzhen Wu, 2024) View paper
  - [39] Provable Scaling Laws for the Test-Time Compute of Large Language Models (Chen Yan-xi, 2024) View paper
  - Algorithmic Progress and Efficiency Scaling (1 papers)
  - [1] Algorithmic progress in language models (Ho, 2024) View paper
- Architecture-Specific Scaling Laws
  - Sparse and Mixture-of-Experts Scaling (4 papers)
  - [6] Unified Scaling Laws for Routed Language Models (Clark, 2022) View paper
  - [15] Towards Greater Leverage: Scaling Laws for Efficient Mixture-of-Experts Language Models (Tian, 2025) View paper
  - [21] Parameters vs FLOPs: Scaling Laws for Optimal Sparsity for Mixture-of-Experts Language Models (Abnar, 2025) View paper
  - [34] Scaling Laws Across Model Architectures: A Comparative Analysis of Dense and MoE Models in Large Language Models (Wang Siqi, 2024) View paper
  - Linear Complexity Architecture Scaling (1 papers)
  - [30] Scaling Laws for Linear Complexity Language Models (Shen, 2024) View paper
  - Quantized and Low-Bit Model Scaling (2 papers)

    - [11] BitNet: Scaling 1-bit Transformers for Large Language Models (Wang Hongyu, 2023) View paper
    - [16] The Era of 1-bit LLMs: All Large Language Models are in 1.58 Bits (Ma, 2024) View paper
    - Parallel Computation Scaling (1 papers)
    - [3] Parallel scaling law for language models (Hui, 2025) View paper
- Data-Centric Scaling Laws
    - Data Quality and Composition Scaling (2 papers)
    - [20] Scaling laws revisited: modeling the role of data quality in language model pretraining (Subramanyam, 2025) View paper
    - [43] AutoScale: Automatic Prediction of Compute-optimal Data Compositions for Training LLMs (F Kang, 2024) View paper
    - Synthetic Data Scaling (1 papers)
    - [9] Scaling Laws of Synthetic Data for Language Models (Qin Ze-yu, 2025) View paper
    - Multilingual and Cross-Lingual Scaling (1 papers)
    - [24] Scaling Laws for Multilingual Language Models (He Yifei, 2024) View paper
    - Mixed-Modal Data Scaling (1 papers)
    - [18] Scaling Laws for Generative Mixed-Modal Language Models (Aghajanyan, 2023) View paper
- Training Method and Hyperparameter Scaling
    - Fine-Tuning and Adaptation Scaling (2 papers)
    - [27] When Scaling Meets LLM Finetuning: The Effect of Data, Model and Finetuning Method (Zhang Biao, 2024) View paper
    - [28] Scaling Laws for Forgetting When Fine-Tuning Large Language Models (Kalajdzievski, 2024) View paper
    - Self-Training and Iterative Improvement Scaling (1 papers)
    - [13] Beyond Human Data: Scaling Self-Training for Problem-Solving with Language Models (Singh Avi, 2023) View paper
    - Hyperparameter and Training Configuration Scaling (2 papers)
    - [26] Scaling laws with vocabulary: Larger models deserve larger vocabularies (Tao, 2024) View paper
    - [46] Predictable Scale: Part I - Optimal Hyperparameter Scaling Law in Large Language Model Pretraining (Li Houyi, 2025) View paper
    - Privacy-Preserving Training Scaling (1 papers)
    - [19] Scaling Laws for Differentially Private Language Models (McKenna Ryan, 2025) View paper
    - Continued Training and Transfer Scaling (1 papers)
    - [33] Transcending scaling laws with 0.1% extra compute (Tay, 2023) View paper
- Capability-Specific Scaling Laws
    - Reasoning and Problem-Solving Scaling (1 papers)
    - [8] Scaling Relationship on Learning Mathematical Reasoning with Large Language Models (YUAN Zheng, 2023) View paper
    - Factual Knowledge Memorization Scaling (2 papers)
    - [22] Scaling Laws for Fact Memorization of Large Language Models (Lu Xingyu, 2024) View paper
    - [48] Physics of Language Models: Part 3.3, Knowledge Capacity Scaling Laws (Allen-Zhu, 2024) View paper
    - Geographical and Domain-Specific Knowledge Scaling (1 papers)
    - [12] On the Scaling Laws of Geographical Representation in Language Models (Godey, 2024) View paper
- Predictive Modeling and Observational Methods
    - Observational Scaling Law Inference (2 papers)
    - [5] Observational Scaling Laws and the Predictability of Language Model Performance (Ruan, 2024) View paper
    - [47] Observational scaling laws and the predictability of langauge model performance (Tatsunori Hashimoto, 2024) View paper
    - Downstream Performance Prediction (2 papers)
    - [17] Unveiling downstream performance scaling of llms: A clustering-based perspective (Xu Cheng-Yin, 2025) View paper
    - [45] The Fine Line: Navigating Large Language Model Pretraining with Down-streaming Capability Analysis (Yang Chen, 2024) View paper
    - Relative and Distributional Scaling Laws (1 papers)
    - [36] Relative Scaling Laws for LLMs (Held, 2025) View paper
- Model Composition and Merging Scaling (1 papers)
    - [40] Model merging scaling laws in large language models (Wang Yuanyi, 2025) View paper
- System-Level Scaling Analysis
    - Distributed Training System Scaling (2 papers)
    - [25] MegaScale: Scaling Large Language Model Training to More Than 10, 000 GPUs (Jiang Ziheng, 2024) View paper
    - [37] Performance modeling and workload analysis of distributed large language model training and inference (Joyjit Kundu, 2024) View paper
    - Inference and Deployment Efficiency Scaling (1 papers)
    - [41] PowerInfer: Fast Large Language Model Serving with a Consumer-grade GPU (Yi-xin Song, 2023) View paper
- Theoretical and Information-Theoretic Scaling (2 papers)
    - [38] Densing law of llms (Chaojun Xiao, 2025) View paper
    - [49] The Information of Large Language Model Geometry (Tan, 2024) View paper
- Large-Scale Model Development and Empirical Studies
    - Billion-Scale Model Training Reports (1 papers)
    - [23] PaLM: Scaling Language Modeling with Pathways (Chowdhery, 2022) View paper
    - Controlled Scaling Study Suites (1 papers)
    - [29] Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling (Biderman, 2023) View paper
    - Multimodal Large Model Scaling (1 papers)
    - [10] SPHINX-X: Scaling Data and Parameters for a Family of Multi-modal Large Language Models (Liu Dong-yang, 2024) View paper
- Robustness and Safety Scaling (2 papers)
    - [31] ANAH-v2: Scaling Analytical Hallucination Annotation of Large Language Models (Gu, 2024) View paper
    - [50] Mass-scale analysis of in-the-wild conversations reveals complexity bounds on llm jailbreaking (Fernandez, 2025) View paper
- Automated Scaling Law Discovery ★ (1 papers)
    - [0] Can Language Models Discover Scaling Laws? (Anon et al., 2026) View paper

- Scaling Law Applications and Extensions (2 papers)
    - [35] d1: Scaling Reasoning in Diffusion Large Language Models via Reinforcement Learning (Zhao Siyan, 2025) View paper
    - [44] Do larger language models imply better reasoning? a pretraining scaling law for reasoning (X Wang, 2025) View paper

Narrative

Core task: automated discovery of scaling laws for language model performance. The field has matured into a rich taxonomy spanning empirical characterization of how loss and performance scale with compute, data, and model size; architecture-specific investigations into transformers, mixture-of-experts, and quantized models; data-centric studies examining quality, diversity, and synthetic data effects; and capability-specific analyses for reasoning, memorization, and multilingual performance. Branches also address training methods and hyperparameter tuning, predictive modeling techniques that enable observational inference without exhaustive training, model composition and merging dynamics, system-level considerations for distributed training, theoretical foundations rooted in information theory, large-scale empirical studies from industry labs, robustness and safety implications, and applications extending scaling insights to new domains. Representative works illustrate this breadth: Neural Scaling Laws[14] and Observational Scaling Laws[5] anchor empirical and predictive methods, while Pythia[29] and DeepSeek LLM[7] exemplify large-scale empirical studies, and Inference Scaling Laws[4] and Test-Time Compute Scaling[39] explore compute allocation beyond pretraining.

Recent activity highlights tensions between observational efficiency and experimental rigor, with Observational Scaling Laws[5] enabling low-cost prediction while works like Algorithmic Progress LMs[1] and Temporal Scaling Law[2] track how algorithmic improvements shift scaling curves over time. The original paper, LMs Discover Scaling[0], sits squarely within the Automated Scaling Law Discovery branch, proposing that language models themselves can identify and formulate scaling relationships—a meta-level approach contrasting with manual empirical fitting or observational extrapolation. This automation theme connects to AutoScale[43] and Optimal Hyperparameter Scaling[46], which similarly seek to reduce human effort in characterizing scaling behavior. By leveraging models' own reasoning capabilities, LMs Discover Scaling[0] offers a novel complement to traditional methods, potentially accelerating the discovery process as models grow more capable and the space of architectural and training choices expands.

Related Works in Same Category

No sibling papers were found in the same taxonomy leaf. A taxonomy-subtopic-level comparison will be produced instead.

Taxonomy-Level Summary

The original leaf focuses specifically on automated methods and AI agents that discover novel scaling law formulas from data, representing a meta-level approach to scaling law research. The sibling subtopics cover different substantive domains of scaling law research: model composition, robustness/safety properties, specialized applications, and theoretical foundations. While siblings study what scaling laws exist in various contexts, the original leaf studies how to automate the discovery process itself.

Similarities: - All subtopics relate to understanding scaling behavior in language models - All involve analyzing relationships between model scale and various properties or outcomes - All contribute to the broader goal of predicting model behavior at different scales

Differences: - Original leaf is methodological (how to discover laws) while siblings are domain-specific (what laws exist in particular contexts) - Original leaf emphasizes automation and AI-driven discovery; siblings focus on manual empirical studies or theoretical derivations - Siblings examine specific phenomena (merging, robustness, specialized contexts, theory) while original leaf is agnostic to the phenomenon being studied - Original leaf excludes manual empirical work which is the primary approach in sibling categories

Suggested Search Directions: - Automated discovery methods applied to robustness or safety scaling patterns - Meta-learning approaches for identifying scaling laws across multiple domains - AI agents that propose and test theoretical scaling law hypotheses

Sibling Subtopics

- **Model Composition and Merging Scaling** (leaves: 1, papers: 1)
- Scope: Studies scaling laws for combining multiple trained models through merging or ensemble methods.
- Exclude: Excludes single-model training and mixture-of-experts architectures; those belong in training or architecture categories.
- **Robustness and Safety Scaling** (leaves: 1, papers: 2)
- Scope: Studies how model vulnerabilities, hallucinations, or adversarial robustness scale with model size.
- Exclude: Excludes general capability scaling and performance metrics; those belong in empirical or capability categories.
- **Scaling Law Applications and Extensions** (leaves: 1, papers: 2)
- Scope: Applies or extends scaling law concepts to specialized contexts not covered by core empirical studies.
- Exclude: Excludes foundational scaling law characterization; those belong in empirical categories.
- **Theoretical and Information-Theoretic Scaling** (leaves: 1, papers: 2)
- Scope: Develops theoretical frameworks using information theory or statistical principles to explain scaling phenomena.
- Exclude: Excludes purely empirical studies and system-level analysis; those belong in empirical or system categories.

Contributions Analysis

Overall novelty summary. The paper introduces SLDAgent, an evolution-based system that autonomously discovers scaling law formulas from experimental data, and SLDBench, a benchmark comprising over 5,000 experiments across seven tasks. This work occupies the 'Automated Scaling Law Discovery' leaf in the taxonomy, which currently contains no sibling papers—making it the sole representative of this research direction. While the broader taxonomy encompasses 50 papers across 33 leaf nodes, this particular branch remains sparse, suggesting that automated discovery of scaling laws is an emerging rather than crowded area.

The taxonomy reveals substantial activity in adjacent branches: empirical characterization methods (Fundamental Compute-Loss Scaling, Temporal Dynamics), predictive modeling approaches (Observational Scaling Law Inference, Downstream Performance Prediction), and hyperparameter optimization (Hyperparameter and Training Configuration Scaling). The original paper diverges from these by proposing meta-level automation—using language models to discover laws rather than manually fitting empirical data or observationally inferring relationships. This positions the work at the intersection of predictive modeling and training method optimization, but with a fundamentally different mechanism: agentic exploration rather than human-guided experimentation or statistical extrapolation.

Among 26 candidates examined, the contribution-level analysis reveals mixed novelty signals. The benchmark contribution (SLDBench) examined 10 candidates with no clear refutations, suggesting this curation effort addresses a gap in standardized evaluation. The agent contribution (SLDAgent) examined 6 candidates and found 1 refutable match, indicating some overlap with prior automated discovery or optimization methods within this limited search scope. The superhuman performance claim examined 10 candidates without refutation, though this reflects the search scale rather than exhaustive validation. The statistics suggest moderate prior work density for the agent mechanism, but sparser coverage for benchmark construction and performance claims.

Given the limited search scope of 26 semantically similar papers, this analysis captures nearby work but cannot claim exhaustive coverage of all relevant optimization, meta-learning, or symbolic regression methods. The taxonomy structure and contribution-level statistics together suggest the work occupies a genuinely sparse research direction (automated scaling law discovery), though individual technical components (evolution-based search, formula optimization) may connect to broader literatures in automated machine learning and symbolic discovery not fully represented in this domain-specific search.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: SLDBench: A comprehensive scaling law discovery benchmark

Description: The authors introduce SLDBench, a benchmark containing seven diverse scaling law discovery tasks derived from over 5,000 experiments in existing literature. Each task requires identifying a symbolic expression that accurately extrapolates to unseen test data, providing a rigorous testbed for evaluating agentic scientific discovery systems.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling

URL: View paper

Brief Assessment

Pythia[29] focuses on providing a suite of language models with consistent training data ordering and checkpoints for analyzing model development, not on creating benchmarks for discovering scaling laws through automated systems.

---

2. Rewardbench: Evaluating reward models for language modeling

URL: View paper

Brief Assessment

RewardBench[67] focuses on evaluating reward models for RLHF in language modeling, not on discovering scaling laws in training experiments. The two benchmarks address fundamentally different problems.

---

3. Revisiting Neural Scaling Laws in Language and Vision

URL: View paper

Brief Assessment

Neural Scaling Vision[61] focuses on validating existing scaling law formulas through extrapolation methodology in vision and NLP domains, not on creating a benchmark for discovering new scaling laws from experimental data.

---

4. Scaling Laws for Generative Mixed-Modal Language Models

URL: View paper

Brief Assessment

Mixed-Modal Scaling Laws[18] focuses on empirical scaling laws for mixed-modal generative models across different modalities (text, speech, images, code), not on creating a benchmark for discovering scaling laws or evaluating agentic systems for scaling law discovery.

---

5. PaLM: Scaling Language Modeling with Pathways

URL: View paper

Brief Assessment

PaLM[23] focuses on demonstrating few-shot learning performance of a 540B parameter language model, not on creating benchmarks for discovering scaling laws from experimental data.

---

6. Scaling Laws of Synthetic Data for Language Models

URL: View paper

Brief Assessment

Synthetic Data Scaling[9] focuses on scaling laws of synthetic data for language model pre-training, not on benchmarking scaling law discovery systems or evaluating agentic scientific discovery capabilities.

---

7. Scaling data-constrained language models

URL: View paper

Brief Assessment

Data-Constrained Scaling[68] focuses on empirical scaling laws for data-constrained language model training, not on creating a benchmark for discovering scaling laws. The candidate proposes specific scaling law formulas and validates them experimentally, rather than providing a testbed for evaluating agentic discovery systems.

---

8. Exploring scaling laws for local SGD in large language model training

URL: View paper

Brief Assessment

Local SGD Scaling[70] focuses on discovering scaling laws for a specific distributed optimization algorithm (Local SGD) in LLM training, not on creating a general benchmark for scaling law discovery tasks across diverse scenarios.

---

9. DeepSeek LLM: Scaling Open-Source Language Models with Longtermism

URL: View paper

Brief Assessment

DeepSeek LLM[7] focuses on applying existing scaling laws to guide the development of open-source language models (7B and 67B configurations), rather than proposing a benchmark for discovering or evaluating scaling laws themselves. The paper does not present a testbed or evaluation framework for scaling law discovery systems.

---

10. Reproducible scaling laws for contrastive language-image learning

URL: View paper

Brief Assessment

Contrastive CLIP Scaling[69] focuses on empirical scaling laws for contrastive language-image pre-training (CLIP models) with specific vision-language datasets, not on creating a benchmark for discovering scaling laws across diverse experimental scenarios as SLDBench does.

Contribution 2: SLDAgent: An evolution-based agent for scaling law discovery

Description: The authors propose SLDAgent, a novel evolution-based agent that co-optimizes both the scaling law expression and its parameter fitting routine. This evolutionary approach enables autonomous exploration of complex variable relationships and achieves state-of-the-art performance on scaling law discovery tasks.

This contribution was assessed against **6 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Spatiotemporal co-optimization of agricultural management practices towards climate-smart crop production

URL: View paper

Brief Assessment

Climate-Smart Crop Production[51] focuses on agricultural management optimization and soil carbon modeling, not on scaling law discovery or evolution-based agents for AI systems. The domains are entirely different.

2. Joint scaling laws in functional and evolutionary categories in prokaryotic genomes

URL: View paper

Brief Assessment

Prokaryotic Scaling Laws[55] focuses on biological genome evolution and functional category scaling in prokaryotes, not on automated discovery of scaling laws through evolutionary optimization of mathematical expressions and parameter fitting routines for AI systems.

3. Multi-criteria selection and scaling of ground motion records using Evolutionary Algorithms

URL: View paper

Brief Assessment

Ground Motion Scaling[54] applies differential evolution to ground motion record selection for structural engineering, not to discovering scaling laws or co-optimizing symbolic expressions with parameter fitting routines in machine learning contexts.

4. : Democratized LLM Scaling for A Large Model Zoo in the Wild

URL: View paper

Brief Assessment

Democratized LLM Scaling[52] focuses on model merging, mixture-of-experts, and stacking techniques for combining pre-trained LLMs, not on evolutionary methods for discovering scaling laws through expression and parameter co-optimization.

5. Evosld: Automated neural scaling law discovery with large language models

URL: View paper

Prior Art Analysis

EvoSLD[53] demonstrates that prior work exists using evolutionary algorithms guided by LLMs to co-optimize scaling law expressions and parameter fitting routines. The candidate paper presents a nearly identical approach: both use evolutionary search to co-evolve symbolic expressions and optimization subroutines, both leverage LLMs for intelligent mutations, and both employ multi-island evolutionary strategies. The core technical approach of co-optimizing expression and fitting procedure through LLM-guided evolution was already established in EvoSLD[53], challenging the novelty claim that SLDAgent is the first to propose this co-optimization framework.

Evidence

Evidence 1 - **Rationale**: Both papers explicitly describe decomposing the problem into two co-evolving subroutines (expression and optimization), demonstrating the same architectural design was already present in EvoSLD[53]. - **Original**: the core design of sldagent is the evolution ofa pair of subroutines,(expression,optimization), which together define the model and the fitting function. we adopt this co-evolution design due to the nature of the sld problem and that a general, problemagnostic evolution (novikov et al., 2025) faces ... - **Candidate**: the core idea is to decompose the problem from sec. 3.1 into two co-evolving components, which are represented as code subroutines: 1. the expression subroutine: this defines the symbolic structure of the scaling law,f(x; θ), with undefined coefficients θ. this structure represents a family of poten...

Evidence 2 - **Rationale**: Both papers employ multi-island evolutionary strategies to prevent premature convergence, showing that this specific technical implementation approach was already established in EvoSLD[53]. - **Original**: following alphaevolve (novikov et al., 2025), we increase sample diversity and prevent premature convergence by using a multi-island strategy (five islands) and map-elites, which structures the population by fitting score, complexity, and novelty. - **Candidate**: the search runs for 50 generations using a multi-population setup of 3 separate islands, between which a 10% migration occurs every 20 generations.

6. Wavelet denoising with evolutionary algorithms

URL: View paper

Brief Assessment

Wavelet Evolutionary Denoising[56] applies evolutionary algorithms to wavelet denoising by co-optimizing basis representation and thresholding parameters. This is a signal processing application, not a framework for discovering scaling laws in machine learning systems through expression and parameter co-optimization.

Contribution 3: Demonstration of superhuman scaling law discovery

Description: The authors demonstrate for the first time that an AI agent can autonomously discover scaling laws that consistently outperform human-derived counterparts in extrapolation accuracy across all benchmark tasks. They validate the practical utility of these discovered laws in pretraining and fine-tuning applications.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Considerations on Stellaratorâs Optimization from the Perspective of the Energy Confinement Time Scaling Laws

URL: View paper

Brief Assessment

Stellarator Energy Confinement[65] focuses on energy confinement time scaling laws for stellarator fusion reactors using symbolic regression, not on AI agents autonomously discovering scaling laws for language models with superior extrapolation accuracy.

2. Revisiting Neural Scaling Laws in Language and Vision

URL: View paper

Brief Assessment

Neural Scaling Vision[61] proposes the M4 estimator for better extrapolation of known scaling law parameters, but does not demonstrate autonomous AI agent discovery of scaling laws that outperform human-derived counterparts.

3. Human mobility is well described by closed-form gravity-like models learned automatically from data

URL: View paper

Brief Assessment

Gravity Mobility Models[64] focuses on discovering closed-form models for human mobility flows between urban areas, not on automated discovery of scaling laws for AI model performance with superior extrapolation accuracy.

---

4. Model Steering: Learning with a Reference Model Improves Generalization Bounds and Scaling Laws

URL: View paper

Brief Assessment

Reference Model Steering[59] focuses on using trained models as references to guide target model training through data selection/weighting, not on automated discovery of scaling laws. The paper demonstrates superior scaling laws for their CLIP method but does not claim to autonomously discover scaling laws that outperform human-derived counterparts across diverse benchmark tasks.

---

5. Machine Learningâ□□Informed Predictive Design and Analysis of Electrohydrodynamic Printing Systems

URL: View paper

Brief Assessment

Electrohydrodynamic Printing ML[63] focuses on applying existing ML models (ridge regression, random forest, etc.) to predict electrohydrodynamic printing outcomes, comparing them to previously reported scaling laws in that specific domain. This is fundamentally different from autonomously discovering novel scaling laws for AI systems that outperform human-derived counterparts across diverse benchmarks.

---

6. Bayesian Neural Scaling Law Extrapolation with Prior-Data Fitted Networks

URL: View paper

Brief Assessment

Bayesian Prior-Fitted Networks[66] focuses on Bayesian uncertainty quantification for scaling law extrapolation, not on autonomous AI agent discovery of scaling laws that outperform human-derived counterparts across benchmark tasks.

---

7. Scaling laws for predicting downstream performance in llms

URL: View paper

Brief Assessment

Predicting Downstream Performance[58] focuses on predicting downstream task performance in LLMs using pre-training loss as an intermediate variable, rather than on automated discovery of scaling law formulas that outperform human-derived counterparts across diverse tasks.

---

8. Broken neural scaling laws

URL: View paper

Brief Assessment

Broken Neural Scaling[62] focuses on discovering a specific functional form (smoothly broken power laws) for modeling neural scaling behavior, not on automated discovery systems that outperform human-derived laws. The candidate demonstrates a mathematical framework rather than an AI agent autonomously discovering laws.

---

9. Beyond neural scaling laws: beating power law scaling via data pruning

URL: View paper

Brief Assessment

Beyond Power Law[57] focuses on data pruning strategies to beat power law scaling in neural networks, not on automated discovery of scaling laws themselves. The candidate demonstrates improved scaling through careful data selection rather than autonomous scaling law discovery.

---

10. Bayesian Neural Scaling Laws Extrapolation with Prior-Fitted Networks

URL: View paper

Brief Assessment

Bayesian Scaling Extrapolation[60] focuses on uncertainty quantification in scaling law extrapolation using Bayesian methods and prior-data fitted networks, not on autonomous discovery of scaling laws that outperform human-derived counterparts. The candidate addresses prediction accuracy and uncertainty estimation, while the original paper demonstrates an AI agent autonomously discovering novel scaling law formulas.

Appendix: Text Similarity Detection

Textual similarity detection checked 25 papers and found 3 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Evosld: Automated neural scaling law discovery with large language models

Detected in: Contribution: contribution_2

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] Can Language Models Discover Scaling Laws? View paper
- [1] Algorithmic progress in language models View paper
- [2] Temporal scaling law for large language models View paper
- [3] Parallel scaling law for language models View paper
- [4] Inference scaling laws: An empirical analysis of compute-optimal inference for problem-solving with language models View paper
- [5] Observational Scaling Laws and the Predictability of Language Model Performance View paper
- [6] Unified Scaling Laws for Routed Language Models View paper
- [7] DeepSeek LLM: Scaling Open-Source Language Models with Longtermism View paper

- [8] Scaling Relationship on Learning Mathematical Reasoning with Large Language Models View paper
- [9] Scaling Laws of Synthetic Data for Language Models View paper
- [10] SPHINX-X: Scaling Data and Parameters for a Family of Multi-modal Large Language Models View paper
- [11] BitNet: Scaling 1-bit Transformers for Large Language Models View paper
- [12] On the Scaling Laws of Geographical Representation in Language Models View paper
- [13] Beyond Human Data: Scaling Self-Training for Problem-Solving with Language Models View paper
- [14] Scaling Laws for Neural Language Models View paper
- [15] Towards Greater Leverage: Scaling Laws for Efficient Mixture-of-Experts Language Models View paper
- [16] The Era of 1-bit LLMs: All Large Language Models are in 1.58 Bits View paper
- [17] Unveiling downstream performance scaling of llms: A clustering-based perspective View paper
- [18] Scaling Laws for Generative Mixed-Modal Language Models View paper
- [19] Scaling Laws for Differentially Private Language Models View paper
- [20] Scaling laws revisited: modeling the role of data quality in language model pretraining View paper
- [21] Parameters vs FLOPs: Scaling Laws for Optimal Sparsity for Mixture-of-Experts Language Models View paper
- [22] Scaling Laws for Fact Memorization of Large Language Models View paper
- [23] PaLM: Scaling Language Modeling with Pathways View paper
- [24] Scaling Laws for Multilingual Language Models View paper
- [25] MegaScale: Scaling Large Language Model Training to More Than 10, 000 GPUs View paper
- [26] Scaling laws with vocabulary: Larger models deserve larger vocabularies View paper
- [27] When Scaling Meets LLM Finetuning: The Effect of Data, Model and Finetuning Method View paper
- [28] Scaling Laws for Forgetting When Fine-Tuning Large Language Models View paper
- [29] Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling View paper
- [30] Scaling Laws for Linear Complexity Language Models View paper
- [31] ANAH-v2: Scaling Analytical Hallucination Annotation of Large Language Models View paper
- [32] Scaling with collapse: Efficient and predictable training of llm families View paper
- [33] Transcending scaling laws with 0.1% extra compute View paper
- [34] Scaling Laws Across Model Architectures: A Comparative Analysis of Dense and MoE Models in Large Language Models View paper
- [35] d1: Scaling Reasoning in Diffusion Large Language Models via Reinforcement Learning View paper
- [36] Relative Scaling Laws for LLMs View paper
- [37] Performance modeling and workload analysis of distributed large language model training and inference View paper
- [38] Densing law of llms View paper
- [39] Provable Scaling Laws for the Test-Time Compute of Large Language Models View paper
- [40] Model merging scaling laws in large language models View paper
- [41] PowerInfer: Fast Large Language Model Serving with a Consumer-grade GPU View paper
- [42] A Multi-Power Law for Loss Curve Prediction Across Learning Rate Schedules View paper
- [43] AutoScale: Automatic Prediction of Compute-optimal Data Compositions for Training LLMs View paper
- [44] Do larger language models imply better reasoning? a pretraining scaling law for reasoning View paper
- [45] The Fine Line: Navigating Large Language Model Pretraining with Down-streaming Capability Analysis View paper
- [46] Predictable Scale: Part I - Optimal Hyperparameter Scaling Law in Large Language Model Pretraining View paper
- [47] Observational scaling laws and the predictability of langauge model performance View paper
- [48] Physics of Language Models: Part 3.3, Knowledge Capacity Scaling Laws View paper
- [49] The Information of Large Language Model Geometry View paper
- [50] Mass-scale analysis of in-the-wild conversations reveals complexity bounds on llm jailbreaking View paper
- [51] Spatiotemporal co-optimization of agricultural management practices towards climate-smart crop production View paper
- [52] : Democratized LLM Scaling for A Large Model Zoo in the Wild View paper
- [53] Evosld: Automated neural scaling law discovery with large language models View paper
- [54] Multi-criteria selection and scaling of ground motion records using Evolutionary Algorithms View paper
- [55] Joint scaling laws in functional and evolutionary categories in prokaryotic genomes View paper
- [56] Wavelet denoising with evolutionary algorithms View paper
- [57] Beyond neural scaling laws: beating power law scaling via data pruning View paper
- [58] Scaling laws for predicting downstream performance in llms View paper
- [59] Model Steering: Learning with a Reference Model Improves Generalization Bounds and Scaling Laws View paper
- [60] Bayesian Neural Scaling Laws Extrapolation with Prior-Fitted Networks View paper
- [61] Revisiting Neural Scaling Laws in Language and Vision View paper
- [62] Broken neural scaling laws View paper
- [63] Machine Learningâ□□Informed Predictive Design and Analysis of Electrohydrodynamic Printing Systems View paper
- [64] Human mobility is well described by closed-form gravity-like models learned automatically from data View paper
- [65] Considerations on Stellaratorâ□□s Optimization from the Perspective of the Energy Confinement Time Scaling Laws View paper
- [66] Bayesian Neural Scaling Law Extrapolation with Prior-Data Fitted Networks View paper
- [67] Rewardbench: Evaluating reward models for language modeling View paper
- [68] Scaling data-constrained language models View paper
- [69] Reproducible scaling laws for contrastive language-image learning View paper
- [70] Exploring scaling laws for local SGD in large language model training View paper