Novelty Assessment Report

Paper: Discovering alternative solutions beyond the simplicity bias in recurrent neural networks

PDF URL: https://openreview.net/pdf?id=8fViWZ0yZJ

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-27

Abstract

Training recurrent neural networks (RNNs) to perform neuroscience-style tasks has become a popular way to generate hypotheses for how neural circuits in the brain might perform computations. Recent work has demonstrated that task-trained RNNs possess a strong simplicity bias. In particular, this inductive bias often causes RNNs trained on the same task to collapse on effectively the same solution, typically comprised of fixed-point attractors or other low-dimensional dynamical motifs. While such solutions are readily interpretable, this collapse proves counterproductive for the sake of generating a set of genuinely unique hypotheses for how neural computations might be performed. Here we propose Iterative Neural Similarity Deflation (INSD), a simple method to break this inductive bias. By penalizing linear predictivity of neural activity produced by standard task-trained RNNs, we find an alternative class of solutions to classic neuroscience-style RNN tasks. These solutions appear distinct across a battery of analysis techniques, including representational similarity metrics, dynamical systems analysis, and the linear decodability of task-relevant variables. Moreover, these alternative solutions can sometimes achieve superior performance in difficult or out-of-distribution task regimes. Our findings underscore the importance of moving beyond the simplicity bias to uncover richer and more varied models of neural computation.

Core Task Landscape

This paper addresses: **Discovering diverse solutions in task-trained recurrent neural networks**

A total of **50 papers** were analyzed and organized into a taxonomy with **31 categories**.

Taxonomy Overview

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

- **Solution Diversity and Degeneracy in Task-Trained RNNs**
- **Mechanistic Interpretation of Task-Optimized RNNs**
- **Comparative Modeling: RNNs and Biological Neural Systems**
- **RNN Architecture Design and Optimization**
- **Methodological Tools for RNN Analysis**
- **Application Domains: Forecasting and Prediction**
- **Application Domains: Classification and Detection**
- **Application Domains: Scheduling and Resource Optimization**
- **Application Domains: Computer Vision and Image Processing**
- **General RNN Optimization and Training Methods**

Complete Taxonomy Tree

- Discovering diverse solutions in task-trained recurrent neural networks Survey Taxonomy
- Solution Diversity and Degeneracy in Task-Trained RNNs
  - Quantifying and Controlling Solution Degeneracy ★ (3 papers)
  - [0] Discovering alternative solutions beyond the simplicity bias in recurrent neural networks (Anon et al., 2026) View paper
  - [25] Measuring and Controlling Solution Degeneracy across Task-Trained Recurrent Neural Networks (Huang Ann, 2024) View paper
  - [29] Charting and Navigating the Space of Solutions for Recurrent Neural Networks (Turner, 2021) View paper
  - Universality and Individuality Across RNN Populations (1 papers)
  - [26] Universality and individuality in neural dynamics across large populations of recurrent networks (Maheswaranathan, 2019) View paper
  - Discovery of Multiple Algorithmic Strategies (3 papers)
  - [3] Automatic discovery of cognitive strategies with tiny recurrent neural networks (L Ji-An, 2023) View paper
  - [5] The Clock and the Pizza: Two Stories in Mechanistic Explanation of Neural Networks (Zhong Ziqian, 2023) View paper
  - [47] Discovering Cognitive Strategies with Tiny Recurrent Neural Networks (Jian Li, 2023) View paper
- Mechanistic Interpretation of Task-Optimized RNNs
  - Dynamical Systems and Attractor Analysis (3 papers)
  - [1] The emergence of visual simulation in task-optimized recurrent neural networks (AK Ashok, 2022) View paper
  - [22] When and why does motor preparation arise in recurrent neural network models of motor control? (Marine Schimel, 2024) View paper
  - [50] Exploring weight initialization, diversity of solutions, and degradation in recurrent neural networks trained for temporal and decision-making tasks (C. Jarne, 2019) View paper
  - Working Memory Dynamics and Coding Strategies (2 papers)
  - [6] The computational foundations of dynamic coding in working memory (Jake P. Stroud, 2024) View paper

- [18] Geometry of naturalistic object representations in recurrent neural network models of working memory (Pouya Bashivan, 2024) View paper
- Sensory-Memory Interference and Reactivation (2 papers)
- [2] Sufficient conditions for offline reactivation in recurrent neural networks (Bredenberg, 2025) View paper
- [20] Overcoming sensory-memory interference in working memory circuits (Andrii Zahorodnii, 2025) View paper
- Spatial Embedding and Resource Constraints (1 papers)
- [21] Spatially embedded recurrent neural networks reveal widespread links between structural and functional neuroscience findings (Jascha Achterberg, 2023) View paper
- Heterogeneity and Neuromodulation Effects (2 papers)
- [10] Diversity in the impact of heterogeneities on recurrent networks performing a cognitive task (Anjana Santhosh, 2025) View paper
- [30] Structured flexibility in recurrent neural networks via neuromodulation (Linderman, 2024) View paper

- Comparative Modeling: RNNs and Biological Neural Systems
  - Predicting and Aligning with Human Behavior (3 papers)
  - [15] Predicting human behavioral decisions with recurrent neural networks (YA Cheng, 2024) View paper
  - [23] Diverse perceptual biases emerge from Hebbian plasticity in a recurrent neural network model. (Francesca SchÃ¶nsberg, 2025) View paper
  - [33] RTify: Aligning deep neural networks with human behavioral decisions (Cheng Yu-ang, 2024) View paper
  - Brain Activity Encoding and Neural Dynamics (2 papers)
  - [8] Modeling the dynamics of human brain activity with recurrent neural networks (Guclu, 2017) View paper
  - [24] Dynamic tracking of objects in the macaque dorsomedial frontal cortex (Rishi Rajalingham, 2025) View paper
  - Computational Complexity and Deliberation (1 papers)
  - [27] Computational complexity drives sustained deliberation (Tao, 2023) View paper
  - Spatial Navigation and Entorhinal Cortex Modeling (1 papers)
  - [34] Explaining heterogeneity in medial entorhinal cortex with task-driven neural networks (Aran Nayebi, 2021) View paper
  - Face Perception and Recognition Signatures (1 papers)
  - [46] Behavioral signatures of face perception emerge in deep neural networks optimized for face recognition (K Dobs, 2023) View paper
  - Visual System Modeling with Recurrent Architectures (1 papers)
  - [48] Task-driven convolutional recurrent models of the visual system (Kubilius Jonas, 2018) View paper
- RNN Architecture Design and Optimization
  - Evolutionary and Neuroevolutionary Approaches (1 papers)
  - [7] From nodes to networks: Evolving recurrent neural networks (Rawal, 2018) View paper
  - Task-Parametrized and Compositional Structures (2 papers)
  - [35] Learning to compose task-specific tree structures (Jihun Choi, 2018) View paper
  - [44] Task-Parametrized Dynamics: Representation of Time and Decisions in Recurrent Neural Networks. (Cecilia Gisele Jarne, 2025) View paper
  - Quantization and Compression Techniques (1 papers)
  - [19] Approximation with neural networks and construction of quantized recurrent neural networks (Foucault, 2025) View paper
- Methodological Tools for RNN Analysis
  - Dynamical Similarity and Comparison Metrics (1 papers)
  - [38] Inputdsa: Demixing then comparing recurrent and externally driven dynamics (Huang Ann, 2025) View paper
  - Connections to Numerical Algorithms for PDEs (1 papers)
  - [17] Connections between numerical algorithms for PDEs and neural networks (Alt, 2023) View paper
- Application Domains: Forecasting and Prediction
  - Electric Load Forecasting (2 papers)
  - [39] Short-Term Electrical Load Demand Forecasting Based on LSTM and RNN Deep Neural Networks (Badar ul Islam, 2022) View paper
  - [49] Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting (Salah Bouktif, 2020) View paper
  - Weather and Environmental Forecasting (2 papers)
  - [28] Predicting concentration levels of air pollutants by transfer learning and recurrent neural network (Fong, 2025) View paper
  - [41] Comparison of RNN-LSTM, TFDF and stacking model approach for weather forecasting in Bangladesh using historical data from 1963 to 2022 (MD MAHMUDUL HASAN, 2024) View paper
  - Multi-Step Forecasting Strategies (1 papers)
  - [9] Stratify: unifying multi-step forecasting strategies: R. Green et al. (R Green, 2025) View paper
  - Degradation and Remaining Useful Life Prediction (1 papers)
  - [42] A novel WaveNet-GRU deep learning model for PEM fuel cells degradation prediction based on transfer learning (Mohammad Javad Izadi, 2024) View paper
- Application Domains: Classification and Detection
  - Sentiment Analysis and Text Classification (2 papers)
  - [13] Deep-Sentiment: An Effective Deep Sentiment Analysis Using a Decision-Based Recurrent Neural Network (D-RNN) (P. Durga, 2023) View paper
  - [43] Cyberbullying detection: advanced preprocessing techniques & deep learning architecture for Roman Urdu data (Amirita Dewani, 2021) View paper
  - Sound Event Detection and Classification (1 papers)
  - [11] Enhanced Multiple Sound Event Detection and Classification Using Physical Signal Properties in Recurrent Spiking Neural Networks (Zahra Roozbehi, 2025) View paper
  - Medical Diagnosis and Recommendation (2 papers)
  - [14] Deep learning for ECG classification: A comparative study of 1D and 2D representations and multimodal fusion approaches (Hemaxi Narotamo, 2024) View paper
  - [32] CNN-RNN based intelligent recommendation for online medical pre-diagnosis support (Xiaokang Zhou, 2020) View paper
  - Facial Expression and Emotion Recognition (1 papers)
  - [40] MMA-MRNNet: Harnessing Multiple Models of Affect and Dynamic Masked RNN for Precise Facial Expression Intensity Estimation (Dimitrios Kollias, 2023) View paper

- Application Domains: Scheduling and Resource Optimization
    - Cloud Task Scheduling and Load Balancing (3 papers)
    - [16] Dynamic Load Balancing in Cloud Computing: Optimized RL-Based Clustering with Multi-Objective Optimized Task Scheduling (Khan, 2024) View paper
    - [31] End-Edge-Cloud Heterogeneous Resources Scheduling Method Based on RNN and Particle Swarm Optimization (Haijie Wu, 2025) View paper
    - [45] HOSHMAND: Accelerated AI-Driven Scheduler Emulating Conventional Task Distribution Techniques for Cloud Workloads (Michael Bidollahkhani, 2024) View paper
    - Robotic Manipulability Optimization and Control (1 papers)
    - [36] BOMO-RNN: a novel neural network controller for industrial robots with experimental validation (Ameer Hamza Khan, 2025) View paper
- Application Domains: Computer Vision and Image Processing (2 papers)
    - [4] User-independent, accurate and pixel-wise DIC measurements with a task-optimized neural network (Pan B, 2024) View paper
    - [12] Recurrent convolutional strategies for face manipulation detection in videos (Cheng J, 2019) View paper
- General RNN Optimization and Training Methods (1 papers)
    - [37] RegGRU-Opt: A Robust GRU-Based RNN Model for High-Performance Sentiment Analysis and Binary Classification (Aashita Chhabra, 2025) View paper

Narrative

Core task: Discovering diverse solutions in task-trained recurrent neural networks. The field explores how RNNs trained on identical tasks can arrive at qualitatively different internal mechanisms—a phenomenon known as solution degeneracy. The taxonomy reflects a broad landscape organized around several major themes. One branch examines solution diversity and degeneracy directly, investigating how multiple distinct circuit-level strategies emerge from the same objective and how researchers can quantify or control this variability (e.g., Alternative Solutions Simplicity[0], Solution Degeneracy Control[25]). A second branch focuses on mechanistic interpretation, dissecting the internal dynamics and representational structure that task-optimized RNNs develop (Cognitive Strategies Discovery[3], Clock Pizza Mechanistic[5]). A third branch draws comparisons between RNN solutions and biological neural systems, asking whether artificial networks recapitulate known neural coding schemes or circuit motifs (Dynamic Coding Memory[6], Nodes to Networks[7]). Additional branches address architecture design, methodological tools for analysis, and a wide range of application domains—from forecasting and classification to computer vision and resource scheduling—demonstrating that the core questions about solution diversity arise across many practical settings.

Within the solution diversity and degeneracy branch, a particularly active line of work seeks to chart the space of possible solutions and develop principled ways to sample or steer networks toward simpler or more interpretable configurations. Alternative Solutions Simplicity[0] sits squarely in this cluster, emphasizing methods to discover and compare alternative circuit implementations that solve the same task. Nearby efforts such as Solution Degeneracy Control[25] and Charting Solution Space[29] share a focus on mapping out the landscape of degenerate solutions and understanding the factors—initialization schemes, regularization, or architectural constraints—that bias networks toward one solution over another. A key open question is whether certain solutions generalize better or align more closely with biological plausibility, and how one might systematically favor such solutions during training. By situating itself among these works, Alternative Solutions Simplicity[0] contributes tools for navigating the rich, often redundant space of learned representations, helping researchers move beyond single-solution analyses toward a more complete picture of what task-trained RNNs can learn.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Measuring and Controlling Solution Degeneracy across Task-Trained Recurrent Neural Networks

Authors: Huang Ann, Martinelli, Flavio, Rajan, Kanaka | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Task-trained recurrent neural networks (RNNs) are widely used in neuroscience and machine learning to model dynamical computations. To gain mechanistic insight into how neural systems solve tasks, prior work often reverse-engineers individual trained networks. However, different RNNs trained on the same task and achieving similar performance can exhibit strikingly different internal solutions, a phenomenon known as solution degeneracy. Here, we develop a unified framework to systematically quant...

Relationship Analysis

Both papers belong to the same taxonomy category focused on quantifying and controlling solution degeneracy in task-trained RNNs. They overlap in their core objective of understanding and manipulating the diversity of solutions RNNs find for the same task, both examining degeneracy across behavioral, dynamical, and weight-space levels. However, the original paper (INSD) proposes a training-time intervention method that actively penalizes neural similarity to discover alternative solutions beyond the simplicity bias, while the candidate paper develops a measurement framework that systematically quantifies how task complexity, learning regime, network size, and regularization affect naturally occurring degeneracy across 3,400 networks without explicit intervention during training.

---

2. Charting and Navigating the Space of Solutions for Recurrent Neural Networks

Authors: Turner, Elia, Elia Turner, Dabholkar, Kabir, et al. (9 authors total) | **Year/Venue**: 2021 • Neural Information Processing Systems | **URL**: View paper

Abstract

Recurrent Neural Networks (RNNs) were recently successfully used to model the way neural activity drives task-related behavior in animals, operating under the implicit assumption that the obtained solutions are universal. Observations in both neuroscience and machine learning challenge this assumption. Animals can approach a given task with a variety of strategies, and training machine learning algorithms introduces the phenomenon of underspecification. These observations imply that every task i...

Relationship Analysis

Both papers belong to the same taxonomy category focused on quantifying and controlling solution degeneracy in task-trained RNNs. They overlap in their investigation of diverse solutions in RNNs trained on neuroscience-style tasks and their use of dynamical systems analysis to characterize these solutions. However, the original paper proposes Iterative Neural Similarity Deflation (INSD) as an active training method to explicitly generate alternative solutions by penalizing linear predictivity, while the candidate paper focuses on characterizing and navigating the naturally existing space of solutions through post-hoc analysis, extrapolation testing, and reduced dynamics representations without introducing novel training procedures.

Contributions Analysis

Overall novelty summary. The paper introduces Iterative Neural Similarity Deflation (INSD), a method to break the simplicity bias in task-trained RNNs by penalizing linear predictivity of neural activity. It sits within the 'Quantifying and Controlling Solution Degeneracy' leaf, which contains only three papers total. This leaf focuses specifically on methods for measuring and manipulating solution degeneracy across behavioral, dynamical, and weight-space levels. The sparse population suggests this is an emerging rather than saturated research direction, with relatively few prior works directly addressing controlled generation of diverse RNN solutions.

The taxonomy reveals that solution diversity research connects to several neighboring areas. The sibling leaf 'Discovery of Multiple Algorithmic Strategies' (three papers) examines qualitatively different algorithms emerging naturally, while 'Universality and Individuality' (one paper) studies shared versus unique representations across RNN populations. The broader 'Mechanistic Interpretation' branch explores dynamical systems analysis and working memory coding strategies—analytical tools that INSD leverages to characterize discovered solutions. The paper bridges degeneracy control methods with mechanistic interpretation techniques, positioning itself at the intersection of generating diversity and analyzing what makes solutions genuinely distinct.

Among thirty candidates examined, the contribution-level analysis shows mixed novelty signals. The INSD method itself (Contribution 1) examined ten candidates with one refutable match, suggesting some methodological overlap exists within the limited search scope. Similarly, discovering alternative solutions beyond simplicity bias (Contribution 2) found one refutable candidate among ten examined. The framework for generating diverse computational hypotheses (Contribution 3) showed no refutable matches across ten candidates, indicating this framing may be more distinctive. The statistics reflect a focused but not exhaustive literature search, leaving open whether additional relevant work exists beyond the top-thirty semantic matches.

Given the sparse taxonomy leaf and limited search scope, the work appears to address a genuine gap in controlled diversity generation for task-trained RNNs. The one-to-two refutable matches per contribution suggest some methodological precedent exists, but the overall scarcity of papers in this specific research direction indicates the problem remains relatively underexplored. The analysis captures top semantic matches but cannot rule out relevant work in adjacent communities or under different terminological framings.

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

Contribution 1: Iterative Neural Similarity Deflation (INSD) method

Description: The authors introduce INSD, a training procedure that penalizes linear predictivity of neural activity from previously trained RNNs in an iterative manner. This method enables discovery of alternative task solutions that diverge from the prototypical solutions typically found due to simplicity bias in RNNs.

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. Cosine-similarity penalty to discriminate sound classes in weakly-supervised sound event detection

URL: View paper

Brief Assessment

Cosine Similarity Penalty[70] addresses weakly-supervised sound event detection using a penalty to discriminate sound classes, not iterative training to reduce simplicity bias in RNNs solving neuroscience tasks.

---

2. Word embedding based document similarity for the inferring of penalty

URL: View paper

Brief Assessment

Document Similarity Penalty[68] focuses on document similarity using word embeddings for legal case penalty inference, not on iterative training methods for neural networks to reduce simplicity bias or discover alternative task solutions.

---

3. A baseline regularization scheme for transfer learning with convolutional neural networks

URL: View paper

Brief Assessment

Baseline Regularization Transfer[67] focuses on transfer learning with CNNs using L2 penalties to promote similarity to pre-trained models, whereas INSD iteratively penalizes linear predictivity to discover alternative RNN solutions that diverge from prototypical ones.

---

4. Beyond Clean Training Data: A Versatile and Model-Agnostic Framework for Out-of-Distribution Detection with Contaminated Training Data

URL: View paper

Brief Assessment

Contaminated Training Detection[69] addresses out-of-distribution detection with contaminated training data through iterative weight adjustments, not iterative training methods to reduce simplicity bias in RNNs discovering alternative task solutions.

---

5. Evading the Simplicity Bias: Training a Diverse Set of Models Discovers Solutions with Superior OOD Generalization

URL: View paper

Prior Art Analysis

Evading Simplicity Bias[63] demonstrates that iterative training methods penalizing similarity between models to reduce simplicity bias were already proposed and implemented prior to the ORIGINAL paper. The candidate paper describes training a collection of models in parallel with a diversity regularizer that penalizes alignment of input gradients, which can be applied iteratively. This approach explicitly aims to overcome the simplicity bias by forcing networks to learn different predictive patterns, including more complex ones that would otherwise be ignored. The method shares the core concept of iteratively penalizing similarity to previously trained models to discover alternative solutions beyond those favored by simplicity bias.

Evidence

Evidence 1 - **Rationale**: Both papers describe methods that penalize similarity between models to induce learning of alternative, more complex solutions that would otherwise be missed due to simplicity bias. - **Original**: we propose iterative neural similarity deflation (insd), a simple method to break this inductive bias. by penalizing linear predictivity of neural activity produced by standard task-trained rnns, we find an alternative class of solutions to classic neuroscience-style rnn tasks. - **Candidate**: we train a collection of similar models to fit the data in different ways using a penalty on the alignment of their input gradients. we show theoretically and empirically that this induces the learning of more complex predictive patterns.

Evidence 2 - **Rationale**: Both methods iteratively penalize similarity between models during training. The candidate explicitly formulates this as a pairwise penalty across a collection of models, which can be applied iteratively. - **Original**: this method, which we call iterative neural similarity deflation (insd), is loosely analogous to the gram-schmidt procedure but in the space of rnn solutions. by iteratively penalizing the linear predictivity of neural activity produced by previously trained rnns in an online fashion, we find soluti... - **Candidate**: instead of training one model, we train a collection of models {fi} in parallel, where fi = gϕi ◦f. they share an optional feature extractor f (e.g. a resnet) for computational reasons, whereas the model-specific classifiers gϕi are small multi-layer perceptrons (mlps) in our experiments. we replace...

Evidence 3 - **Rationale**: Both papers explicitly connect the penalization of similarity to the discovery of more complex solutions that diverge from the default simple solutions favored by standard training. - **Original**: by iteratively penalizing the linear predictivity of neural activity produced by previously trained rnns in an online fashion, we find solutions that diverge from the prototypical solutions to classic neuroscience-style tasks. we show that the alternative solutions generated in this manner not only ... - **Candidate**: how diversity can induce complexity. by assumption of the simplicity bias, the default predictor learned by solving (1) with sgd is the simplest. in other words, the model learned by default lies at one end of the space of solutions. a diverse set of solutions departing from the default one will nec...

6. A Step-by-Step Gradient Penalty with Similarity Calculation for Text Summary Generation

URL: View paper

Brief Assessment

Gradient Penalty Similarity[64] focuses on text summarization with gradient penalties for similarity calculation, not on iterative training methods for RNNs to reduce simplicity bias in neural network task solutions.

7. Hardware Aware Evolutionary Neural Architecture Search using Representation Similarity Metric

URL: View paper

Brief Assessment

Hardware Aware Architecture[66] focuses on neural architecture search using representation similarity metrics to find efficient architectures for hardware constraints, not on iterative training methods to reduce simplicity bias in RNNs.

8. A generative neural network for maximizing fitness and diversity of synthetic DNA and protein sequences

URL: View paper

Brief Assessment

Maximizing Fitness Diversity[65] focuses on penalizing similarity between generated sequences in a generative model for DNA/protein design, not on iterative training to reduce simplicity bias in RNNs solving neuroscience tasks. The candidate's diversity penalty operates on generated outputs, while INSD penalizes linear predictivity of neural activity across iteratively trained RNNs.

9. Input similarity from the neural network perspective

URL: View paper

Brief Assessment

Input Similarity Perspective[62] focuses on defining similarity measures from a neural network's perspective for analyzing trained networks and understanding self-denoising, not on iterative training procedures to overcome simplicity bias in RNNs.

10. In-Depth Exploration of the Advantages of Neural Networks in English Machine Translation

URL: View paper

Brief Assessment

Neural Machine Translation[61] focuses on machine translation applications using backpropagation algorithms for iterative training, not on methods to reduce simplicity bias or discover alternative task solutions in RNNs through similarity penalties.

Contribution 2: Discovery of alternative RNN solutions beyond simplicity bias

Description: The authors demonstrate that their method uncovers a distinct class of solutions to neuroscience tasks that differ from standard solutions in representational geometry, dynamical motifs, and encoding of task variables. These alternative solutions forgo fixed-point attractors and instead maintain information in dynamically evolving subspaces.

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. EnvFormer: A Decomposition-based Transformer for Multi-step Burn-through Point Prediction in Sintering Process

URL: View paper

Brief Assessment

EnvFormer Burn-Through[60] focuses on burn-through point prediction in sintering processes using transformer architectures, not on discovering alternative RNN solutions or addressing simplicity bias in neural network training for neuroscience tasks.

2. Persistent learning signals and working memory without continuous attractors

URL: View paper

Prior Art Analysis

Persistent Learning Signals[54] demonstrates that alternative RNN solutions using dynamic subspaces (quasi-periodic attractors) versus fixed-point attractors were proposed prior to the original paper. The candidate explicitly shows that working memory and temporal learning can be achieved through oscillatory dynamics in dynamically evolving subspaces rather than fixed-point attractors. Both papers address the same fundamental problem: finding RNN solutions that maintain information without relying on stable fixed points, instead using dynamic evolution in subspaces.

Evidence

Evidence 1 - **Rationale**: Both papers propose mechanisms that avoid fixed-point attractors and instead use dynamic evolution. The candidate explicitly proposes quasi-periodic attractors as an alternative mechanism before the original paper's submission, directly challenging the novelty of discovering alternatives to fixed-point solutions. - **Original**: these solutions forgo the usage of fixed point attractors and slow manifolds for keeping track of task-relevant information, and instead tend to maintain task-relevant information in dynamically evolving subspaces of activity. - **Candidate**: we propose an alternative recurrent mechanism, (quasi-)periodic attractors, with learning signals that are both robust and asymptotically persistent, thus satisfying both requirements for unbounded temporal learning. the proposed mechanism encodes information in oscillations

Evidence 2 - **Rationale**: The candidate demonstrates that periodic and quasi-periodic attractors represent a distinct class of solutions that differ from standard continuous attractor solutions, establishing prior work on alternative RNN solution classes before the original paper. - **Original**: by penalizing linear predictivity of neural activity produced by standard task-trained rnns, we find an alternative class of solutions to classic neuroscience-style rnn tasks. these solutions appear distinct across a battery of analysis techniques, including representational similarity metrics, dyna... - **Candidate**: we show that in addition to the continuous attractors that are widely implicated, periodic and quasi-periodic attractors can also support learning arbitrarily long temporal relationships. unlike the continuous attractors that suffer from the fine-tuning problem, the less explored quasi-periodic attrac...

Evidence 3 - **Rationale**: Both papers demonstrate that their alternative solutions can achieve superior performance on challenging tasks, showing that the candidate established this finding prior to the original paper's claim of novelty. - **Original**: these alternative solutions can sometimes achieve superior performance in difficult or out-of-distribution task regimes. - **Candidate**: to demonstrate its practical benefits, we devised an initialization scheme for artificial recurrent neural networks (rnn) that enables learning challenging tasks.

Evidence 4 - **Rationale**: Both papers identify the same gap in the literature: standard RNN solutions rely on attractor states, but biological evidence suggests more dynamic representations. The candidate explicitly addresses this discrepancy before the original paper. - **Original**: for instance, rnns trained on simple memory tasks ubiquitously find solutions using persistent activity held in stable attractor states (maheswaranathan et al., 2019; turner & barak, 2023; driscoll et al., 2024; hazelden et al., 2025), yet population-level recordings have shown that the neural repre... - **Candidate**: research in working memory mechanisms provides a useful framework for bridging temporal gaps. analyzing working memory as a neural dynamical system, three types of dynamical structures have been generally considered in the literature: fading memory, multiple point attractors, and continuous attracto...

Evidence 5 - **Rationale**: Both papers describe solutions where information is maintained through oscillatory dynamics and trajectory shapes rather than fixed points, demonstrating that the candidate proposed this alternative solution mechanism prior to the original paper. - **Original**: activity in each context was readily distinguishable by the shape of trajectories, rather than the portion of state space they occupy. moreover, activity was no longer driven along slow/fixed points. instead, unstable fixed points with oscillatory eigenmodes were found - **Candidate**: stable limit cycle is composed of an attracting ring manifold, similar to the continuous attractor with ring topology. but unlike the ring attractor, the neural activation is not persistent over time, but rather forms a periodic trajectory

---

3. Recurrent neural networks with explicit representation of dynamic latent variables can mimic behavioral patterns in a physical inference task

URL: View paper

Brief Assessment

Physical Inference Latent[57] focuses on comparing RNN models with/without dynamic inference capacity for physical prediction tasks, not on discovering alternative solutions beyond simplicity bias or comparing fixed-point versus dynamic subspace solutions.

---

4. A novel BDPCA-SMLSTM algorithm for fault diagnosis of industrial process

URL: View paper

Brief Assessment

BDPCA Fault Diagnosis[52] focuses on industrial fault diagnosis using LSTM variants with principal component analysis, not on discovering alternative RNN solutions or analyzing dynamical motifs in neuroscience tasks.

---

5. DSTED: A denoising spatialâ□□temporal encoderâ□□decoder framework for multistep prediction of burn-through point in sintering process

URL: View paper

Brief Assessment

DSTED Burn-Through[58] focuses on industrial process prediction using denoising GRU for sintering operations, not on discovering alternative RNN solutions or analyzing simplicity bias in neuroscience-style tasks.

---

6. Attractor memory for long-term time series forecasting: A chaos perspective

URL: View paper

Brief Assessment

Attractor Memory Forecasting[56] focuses on time series forecasting using chaos theory and attractor dynamics, not on discovering alternative RNN solutions or addressing simplicity bias in recurrent neural networks for neuroscience tasks.

---

7. Deep concatenated features with improved heuristic-based recurrent neural network for hyperspectral image classification

URL: View paper

Brief Assessment

Hyperspectral Image Classification[59] focuses on image classification using concatenated features and heuristic-based RNNs for hyperspectral data, not on discovering alternative RNN solutions for neuroscience tasks or addressing simplicity bias in recurrent networks.

---

8. Back to the continuous attractor

URL: View paper

Brief Assessment

Continuous Attractor[53] focuses on bifurcations and approximations of continuous attractors in the context of analog memory tasks, not on discovering alternative RNN solutions that differ from standard solutions in representational geometry and dynamical motifs as the original paper does. The candidate examines how continuous attractors degrade into slow manifolds, while the original paper uses INSD to actively generate solutions that forgo fixed-point attractors.

---

9. Beyond exploding and vanishing gradients: analysing RNN training using attractors and smoothness

URL: View paper

Brief Assessment

Attractors Gradient Analysis[55] focuses on analyzing RNN training dynamics through the lens of attractors and gradient smoothness, examining how different RNN architectures (LSTM, stable LSTM, orthogonal RNN) learn attractors during training. The original paper's contribution concerns a method (INSD) to actively discover alternative solutions that avoid fixed-point attractors by penalizing linear predictivity, which is a different approach and objective.

---

10. Organizing recurrent network dynamics by task-computation to enable continual learning

URL: View paper

Brief Assessment

Task-Computation Continual[51] focuses on organizing dynamics into orthogonal subspaces for continual learning across multiple tasks, not on discovering alternative solutions to single tasks that avoid fixed-point attractors. The candidate addresses multi-task interference prevention, while the original explores alternative dynamical motifs for individual task solutions.

Contribution 3: Framework for generating diverse computational hypotheses in neuroscience

Description: The authors address the problem of generating multiple competing hypotheses for neural computation by developing a method that overcomes dynamic collapse. Their approach enables production of genuinely unique solutions that can be evaluated against experimental data, moving beyond the limitations of varying only basic hyperparameters.

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. Inferring brain-wide interactions using data-constrained recurrent neural network models

URL: View paper

Brief Assessment

Brain-Wide Interactions[79] focuses on inferring inter-region communication from existing neural data using data-constrained RNNs, not on generating diverse computational hypotheses by overcoming dynamic collapse in task-trained networks.

2. Computational models of multisensory integration with recurrent neural networks: A critical review and future directions

URL: View paper

Brief Assessment

Multisensory Integration Review[74] focuses on reviewing RNN models for multisensory integration tasks, not on methods for generating diverse computational hypotheses or overcoming dynamic collapse in neural networks.

3. Recurrent neural networks as neuro-computational models of human speech recognition

URL: View paper

Brief Assessment

Speech Recognition Models[76] focuses on using RNNs to predict human neural responses to speech stimuli and comparing architectural variations for brain-like temporal dynamics. It does not address the problem of generating multiple diverse computational hypotheses or overcoming dynamic collapse in task-trained networks.

4. Localizing syntactic predictions using recurrent neural network grammars

URL: View paper

Brief Assessment

Localizing Syntactic Predictions[72] focuses on using recurrent neural network grammars to localize syntactic predictions in brain activity, not on generating diverse computational hypotheses or overcoming dynamic collapse in RNNs.

5. Second-order forward-mode optimization of recurrent neural networks for neuroscience

URL: View paper

Brief Assessment

Second-Order Forward-Mode[71] focuses on optimization methods for training RNNs efficiently (memory and compute), not on generating diverse computational hypotheses or overcoming dynamic collapse to produce multiple competing solutions.

6. Biologically plausible models of cognitive flexibility: merging recurrent neural networks with full-brain dynamics

URL: View paper

Brief Assessment

Cognitive Flexibility Models[75] is a review paper discussing hybrid frameworks for modeling cognitive flexibility using RNNs and whole-brain models. It does not address the specific problem of generating diverse hypotheses by overcoming dynamic collapse in task-trained RNNs.

7. Modeling the dynamics of human brain activity with recurrent neural networks

URL: View paper

Brief Assessment

Brain Activity Dynamics[8] focuses on predicting fMRI responses to sensory stimuli using RNNs as response models in an encoding framework, not on generating diverse computational hypotheses by overcoming dynamic collapse in task-trained RNNs.

8. Computational neuroscience: Recent advancement

URL: View paper

Brief Assessment

Computational Neuroscience Advancement[78] only briefly mentions RNNs (LSTM networks) in passing without addressing the problem of generating diverse hypotheses or overcoming dynamic collapse in task-trained networks.

9. Recurrent neural networks as versatile tools of neuroscience research

URL: View paper

Brief Assessment

Versatile Neuroscience Tools[77] discusses RNNs as general research tools in neuroscience but does not address the specific problem of overcoming dynamic collapse or generating genuinely diverse solutions beyond simplicity bias that the original paper tackles.

10. Reconstructing computational system dynamics from neural data with recurrent neural networks

URL: View paper

Brief Assessment

Reconstructing System Dynamics[73] focuses on reconstructing computational dynamics from neural data using RNNs, not on generating diverse hypotheses by overcoming dynamic collapse. The candidate addresses a different problem in computational neuroscience.

Appendix: Text Similarity Detection

Textual similarity detection checked 32 papers and found 1 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. Persistent learning signals and working memory without continuous attractors

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] Discovering alternative solutions beyond the simplicity bias in recurrent neural networks View paper
- [1] The emergence of visual simulation in task-optimized recurrent neural networks View paper
- [2] Sufficient conditions for offline reactivation in recurrent neural networks View paper
- [3] Automatic discovery of cognitive strategies with tiny recurrent neural networks View paper
- [4] User-independent, accurate and pixel-wise DIC measurements with a task-optimized neural network View paper
- [5] The Clock and the Pizza: Two Stories in Mechanistic Explanation of Neural Networks View paper
- [6] The computational foundations of dynamic coding in working memory View paper
- [7] From nodes to networks: Evolving recurrent neural networks View paper
- [8] Modeling the dynamics of human brain activity with recurrent neural networks View paper
- [9] Stratify: unifying multi-step forecasting strategies: R. Green et al. View paper
- [10] Diversity in the impact of heterogeneities on recurrent networks performing a cognitive task View paper
- [11] Enhanced Multiple Sound Event Detection and Classification Using Physical Signal Properties in Recurrent Spiking Neural Networks View paper
- [12] Recurrent convolutional strategies for face manipulation detection in videos View paper
- [13] Deep-Sentiment: An Effective Deep Sentiment Analysis Using a Decision-Based Recurrent Neural Network (D-RNN) View paper
- [14] Deep learning for ECG classification: A comparative study of 1D and 2D representations and multimodal fusion approaches View paper
- [15] Predicting human behavioral decisions with recurrent neural networks View paper
- [16] Dynamic Load Balancing in Cloud Computing: Optimized RL-Based Clustering with Multi-Objective Optimized Task Scheduling View paper
- [17] Connections between numerical algorithms for PDEs and neural networks View paper
- [18] Geometry of naturalistic object representations in recurrent neural network models of working memory View paper
- [19] Approximation with neural networks and construction of quantized recurrent neural networks View paper
- [20] Overcoming sensory-memory interference in working memory circuits View paper
- [21] Spatially embedded recurrent neural networks reveal widespread links between structural and functional neuroscience findings View paper
- [22] When and why does motor preparation arise in recurrent neural network models of motor control? View paper
- [23] Diverse perceptual biases emerge from Hebbian plasticity in a recurrent neural network model. View paper
- [24] Dynamic tracking of objects in the macaque dorsomedial frontal cortex View paper
- [25] Measuring and Controlling Solution Degeneracy across Task-Trained Recurrent Neural Networks View paper
- [26] Universality and individuality in neural dynamics across large populations of recurrent networks View paper
- [27] Computational complexity drives sustained deliberation View paper
- [28] Predicting concentration levels of air pollutants by transfer learning and recurrent neural network View paper
- [29] Charting and Navigating the Space of Solutions for Recurrent Neural Networks View paper
- [30] Structured flexibility in recurrent neural networks via neuromodulation View paper
- [31] End-Edge-Cloud Heterogeneous Resources Scheduling Method Based on RNN and Particle Swarm Optimization View paper
- [32] CNN-RNN based intelligent recommendation for online medical pre-diagnosis support View paper
- [33] RTify: Aligning deep neural networks with human behavioral decisions View paper
- [34] Explaining heterogeneity in medial entorhinal cortex with task-driven neural networks View paper
- [35] Learning to compose task-specific tree structures View paper
- [36] BOMO-RNN: a novel neural network controller for industrial robots with experimental validation View paper
- [37] RegGRU-Opt: A Robust GRU-Based RNN Model for High-Performance Sentiment Analysis and Binary Classification View paper
- [38] Inputdsa: Demixing then comparing recurrent and externally driven dynamics View paper
- [39] Short-Term Electrical Load Demand Forecasting Based on LSTM and RNN Deep Neural Networks View paper
- [40] MMA-MRNNet: Harnessing Multiple Models of Affect and Dynamic Masked RNN for Precise Facial Expression Intensity Estimation View paper
- [41] Comparison of RNN-LSTM, TFDF and stacking model approach for weather forecasting in Bangladesh using historical data from 1963 to 2022 View paper
- [42] A novel WaveNet-GRU deep learning model for PEM fuel cells degradation prediction based on transfer learning View paper
- [43] Cyberbullying detection: advanced preprocessing techniques & deep learning architecture for Roman Urdu data View paper
- [44] Task-Parametrized Dynamics: Representation of Time and Decisions in Recurrent Neural Networks. View paper
- [45] HOSHMAND: Accelerated AI-Driven Scheduler Emulating Conventional Task Distribution Techniques for Cloud Workloads View paper
- [46] Behavioral signatures of face perception emerge in deep neural networks optimized for face recognition View paper
- [47] Discovering Cognitive Strategies with Tiny Recurrent Neural Networks View paper
- [48] Task-driven convolutional recurrent models of the visual system View paper
- [49] Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting View paper
- [50] Exploring weight initialization, diversity of solutions, and degradation in recurrent neural networks trained for temporal and decision-making tasks View paper
- [51] Organizing recurrent network dynamics by task-computation to enable continual learning View paper
- [52] A novel BDPCA-SMLSTM algorithm for fault diagnosis of industrial process View paper
- [53] Back to the continuous attractor View paper
- [54] Persistent learning signals and working memory without continuous attractors View paper
- [55] Beyond exploding and vanishing gradients: analysing RNN training using attractors and smoothness View paper
- [56] Attractor memory for long-term time series forecasting: A chaos perspective View paper
- [57] Recurrent neural networks with explicit representation of dynamic latent variables can mimic behavioral patterns in a physical inference task View paper
- [58] DSTED: A denoising spatialâtemporal encoderâdecoder framework for multistep prediction of burn-through point in sintering process View paper
- [59] Deep concatenated features with improved heuristic-based recurrent neural network for hyperspectral image classification View paper
- [60] EnvFormer: A Decomposition-based Transformer for Multi-step Burn-through Point Prediction in Sintering Process View paper

- [61] In-Depth Exploration of the Advantages of Neural Networks in English Machine Translation View paper
- [62] Input similarity from the neural network perspective View paper
- [63] Evading the Simplicity Bias: Training a Diverse Set of Models Discovers Solutions with Superior OOD Generalization View paper
- [64] A Step-by-Step Gradient Penalty with Similarity Calculation for Text Summary Generation View paper
- [65] A generative neural network for maximizing fitness and diversity of synthetic DNA and protein sequences View paper
- [66] Hardware Aware Evolutionary Neural Architecture Search using Representation Similarity Metric View paper
- [67] A baseline regularization scheme for transfer learning with convolutional neural networks View paper
- [68] Word embedding based document similarity for the inferring of penalty View paper
- [69] Beyond Clean Training Data: A Versatile and Model-Agnostic Framework for Out-of-Distribution Detection with Contaminated Training Data View paper
- [70] Cosine-similarity penalty to discriminate sound classes in weakly-supervised sound event detection View paper
- [71] Second-order forward-mode optimization of recurrent neural networks for neuroscience View paper
- [72] Localizing syntactic predictions using recurrent neural network grammars View paper
- [73] Reconstructing computational system dynamics from neural data with recurrent neural networks View paper
- [74] Computational models of multisensory integration with recurrent neural networks: A critical review and future directions View paper
- [75] Biologically plausible models of cognitive flexibility: merging recurrent neural networks with full-brain dynamics View paper
- [76] Recurrent neural networks as neuro-computational models of human speech recognition View paper
- [77] Recurrent neural networks as versatile tools of neuroscience research View paper
- [78] Computational neuroscience: Recent advancement View paper
- [79] Inferring brain-wide interactions using data-constrained recurrent neural network models View paper