Novelty Assessment Report

Paper: Textual Equilibrium Propagation for Deep Compound AI Systems
PDF URL: https://openreview.net/pdf?id=4XsV5Qoj0j
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-30

Abstract

Large language models (LLMs) are increasingly deployed as part of compound AI systems which coordinate multiple modules (e.g., retrievers, tools, verifiers) over long-horizon workflows. Although recent frameworks that propagate textual feedback globally (e.g., TextGrad make it feasible to optimize such pipelines, we identify two depth-scaling failure modes in long-horizon agentic workflows: 1) exploding textual gradient, where textual feedback grows exponentially with depth, leading to prohibitively long message and amplifies evaluation biases; and 2) vanishing textual gradient, where limited long-context ability causes models overemphasize recent or early feedback, while compression of lengthy feedback causes downstream messages to lose specificity gradually as they propagate many hops upstream. To mitigate these issues, we introduce Textual Equilibrium Propagation (TEP), a local learning principle inspired by Equilibrium Propagation in energy-based models. TEP includes two phases: 1) a free phase where a local LLM critics iteratively refine prompts until reaching equilibrium (no further improvements are suggested); and 2) a nudged phase which applies proximal prompt edits with bounded modification intensity, using task-level objectives that propagate via forward signaling rather than backward feedback chains. This design supports local prompt optimization followed by controlled adaptation toward global goals without the computational burden and signal degradation of global textual backpropagation. Across long-horizon QA benchmarks and multi-agent tool-use dataset, TEP consistently improves accuracy and efficiency over global propagation methods such as TextGrad, with gains that increase at greater depths, while preserving the practicality of black-box LLM components in deep compound AI system.

> **Disclaimer**
>
> This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.
>
> Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.
>
> If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **Optimizing Prompts in Deep Compound AI Systems**

A total of **41 papers** were analyzed and organized into a taxonomy with **16 categories**.

Taxonomy Overview

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

- **Prompt Optimization Methods for Multi-Module Systems**
- **Multi-Stage Prompt Engineering Frameworks**
- **Prompt Optimization Search and Meta-Learning**
- **Infrastructure and Orchestration for Compound Systems**
- **Domain-Specific Prompt Applications**
- **Prompt Design Principles and Formalization**
- **Security and Robustness in Prompt Systems**
- **Practical Deployment and Workflow Integration**

Complete Taxonomy Tree

- Optimizing Prompts in Deep Compound AI Systems Survey Taxonomy
- Prompt Optimization Methods for Multi-Module Systems
  - Global Gradient-Based Optimization ★ (2 papers)
  - [0] Textual Equilibrium Propagation for Deep Compound AI Systems (Anon et al., 2026) View paper
  - [2] Optimizing generative AI by backpropagating language model feedback (Mert Yuksekgonul, 2025) View paper
  - Local Optimization Strategies (2 papers)
  - [1] Optimizing instructions and demonstrations for multi-stage language model programs (Krista Opsahl-Ong, 2024) View paper
  - [6] Optimas: Optimizing Compound AI Systems with Globally Aligned Local Rewards (Wu, 2025) View paper
  - Joint Fine-Tuning and Prompt Optimization (2 papers)
  - [4] Fine-tuning and prompt optimization: Two great steps that work better together (Soylu, 2024) View paper
  - [9] Multi-module GRPO: Composing policy gradients and prompt optimization for language model programs (Ziems, 2025) View paper
- Multi-Stage Prompt Engineering Frameworks
  - Sequential Stage Decomposition (3 papers)
  - [12] MSP: Multi-stage prompting for making pre-trained language models better translators (Zhixing Tan, 2022) View paper
  - [34] A Multi-Stage Prompt Framework for High-Quality News Summarization with Large Language Models (Salma Ali, 2024) View paper
  - [37] Multi-Stage Prompting for Knowledgeable Dialogue Generation (Liu Zihan, 2022) View paper
  - Hierarchical Prompt Architectures (3 papers)
  - [13] Towards hierarchical multi-agent workflows for zero-shot prompt optimization (Liu Yuchi, 2024) View paper
  - [27] Enhancing AI-Driven Psychological Consultation: Layered Prompts with Large Language Models (Souza, 2024) View paper
  - [35] Prompt-Layered Architecture: A New Stack for AI-First Product Design (Khatri, 2024) View paper
  - Adaptive Multi-Stage Prompting (2 papers)

- ◦ [18] GraphPrompter: Multi-Stage Adaptive Prompt Optimization for Graph In-Context Learning (Rui Lv, 2025) View paper
- ◦ [38] A Multi-Stage Self-Optimization Framework for LLM Reasoning: Exploration, Structured Improvement, and Robust Inference (V Chen, n.d.) View paper
- • Prompt Optimization Search and Meta-Learning (1 papers)
  - ◦ [3] Automatic prompt optimization via heuristic search: A survey (W Cui, 2025) View paper
- • Infrastructure and Orchestration for Compound Systems
  - ◦ Network and Execution Orchestration (2 papers)
  - ◦ [5] ALTO: An Efficient Network Orchestrator for Compound AI Systems (Keshav Santhanam, 2024) View paper
  - ◦ [11] A layered architecture for developing and enhancing capabilities in large language model-based software systems (Zhang Da-wen, 2024) View paper
  - ◦ Architectural Design Patterns (2 papers)
  - ◦ [19] Orchestrating strategic control points in layered AI ecosystems: insights from startup founders and investors (Cronjäger, 2025) View paper
  - ◦ [28] Towards a Modular and Adaptive AI Architecture for Personalized Intelligent Tutoring Systems (Torró, 2025) View paper
- • Domain-Specific Prompt Applications
  - ◦ Generative Content Creation (3 papers)
  - ◦ [7] Prompt-a-video: Prompt your video diffusion model via preference-aligned llm (Ji, 2025) View paper
  - ◦ [20] From prompt to production: a design process pipeline using AI (Derksen, 2025) View paper
  - ◦ [22] From Keyframes to Narrative: A Multi-stage AI Pipeline for Scene Understanding Using Object Detection and Large Language Models (Valentín Calzada-Ledesma, 2025) View paper
  - ◦ Structured Data and Knowledge Tasks (4 papers)
  - ◦ [8] Optimization modeling and verification from problem specifications using a multi-agent multi-stage LLM framework (Mahdi Mostajabdaveh, 2024) View paper
  - ◦ [21] Liberal Entity Matching as a Compound AI Toolchain (Silvery D. Fu, 2024) View paper
  - ◦ [26] Integrating Textual Queries with AI-Based Object Detection: A Compositional Prompt-Guided Approach (Silvan Ferreira, 2025) View paper
  - ◦ [39] AI-based Information Retrieval from Structured Text Documents (Premkumar, n.d.) View paper
  - ◦ Interactive and Conversational Systems (2 papers)
  - ◦ [10] Multi-stage prompting for next best agent recommendations in adaptive workflows (Prerna Agarwal, 2024) View paper
  - ◦ [24] Large Language Models for Real-World Nutrition Assessment: Structured Prompts, Multi-Model Validation and Expert Oversight (A Ase, 2025) View paper
  - ◦ Specialized Domain Applications (6 papers)
  - ◦ [14] Multi-stage prompt tuning for political perspective detection in low-resource settings (Kang-Min Kim, 2023) View paper
  - ◦ [15] Deep Learning Architectures for Early GI Disease Detection and Stage Prediction (Fluet, 2024) View paper
  - ◦ [16] Prompt-Based Bioinformatic Pipeline Generation for a Multi-Step Metaviral Workflow (P Ma, 2025) View paper
  - ◦ [17] AI-Driven Optimization Framework for Construction Site Ecosystems (Khan, 2025) View paper
  - ◦ [23] Phase-Adaptive LLM Framework with Multi-Stage Validation for Construction Robot Task Allocation: A Systematic Benchmark Against Traditional Optimization … (Yu, 2025) View paper
  - ◦ [36] Prompts for Planning-Ai Integration: Effective Prompt Design for Large Language Models in Support of Sustainable Urban Development (Ke Liu, n.d.) View paper
- • Prompt Design Principles and Formalization (1 papers)
  - ◦ [25] XML Prompting as Grammar-Constrained Interaction: Fixed-Point Semantics, Convergence Guarantees, and Human-AI Protocols (Faruk Alpay, 2025) View paper
- • Security and Robustness in Prompt Systems (1 papers)
  - ◦ [30] Multi-Stage Prompt Inference Attacks on Enterprise LLM Systems (Andrii Balashov, 2025) View paper
- • Practical Deployment and Workflow Integration (6 papers)
  - ◦ [29] AI-Powered Roadmap Generator Prompt Engineering (A Gupta, 2025) View paper
  - ◦ [31] Leveraging AI for More Effective Teams: Evidence-Based Strategies for Diverse Idea Generation (Westover, 2025) View paper
  - ◦ [32] SocialBrain (Y Jalal, 2025) View paper
  - ◦ [33] Demystifying LLM Serving Pipeline: From Prompt to Response (Kumar, 2025) View paper
  - ◦ [40] General Purpose Technologies vs. AI: Comparative Diffusion Rates and Economic Impact (AJUZIEOGU, n.d.) View paper
  - ◦ [41] BUILDING MODULAR AND SCALABLE AI FOR CONTENT CREATION USING LLAMAINDEX (Popov, n.d.) View paper

Narrative

Core task: optimizing prompts in deep compound AI systems. Modern AI applications increasingly rely on multi-module pipelines where language models are chained together, each stage consuming the output of previous modules and producing inputs for downstream components. The taxonomy reveals several major branches addressing this complexity. Prompt Optimization Methods for Multi-Module Systems focuses on techniques that treat entire pipelines as differentiable or searchable structures, enabling end-to-end tuning across modules. Multi-Stage Prompt Engineering Frameworks and Infrastructure and Orchestration branches emphasize architectural patterns and tooling for managing these cascaded systems, while Prompt Optimization Search and Meta-Learning explores automated discovery of effective prompt configurations. Domain-Specific Prompt Applications demonstrates how these methods adapt to specialized fields, and branches on Design Principles, Security, and Deployment address formalization, robustness, and practical integration challenges.

Within the optimization methods, a particularly active line of work pursues gradient-based or gradient-inspired techniques that propagate feedback through non-differentiable language model boundaries. Textual Equilibrium Propagation[0] exemplifies this global gradient-based optimization approach, drawing on equilibrium propagation principles to update prompts across deep compound systems. It shares conceptual ground with Backpropagating Language Feedback[2], which similarly aims to flow optimization signals backward through multi-stage pipelines, and contrasts with more modular approaches like Optimizing Instructions Demonstrations[1] that tune individual components separately. These gradient-oriented methods face the fundamental challenge of bridging discrete text generation with continuous optimization, a trade-off that distinguishes them from search-based or reinforcement learning alternatives found elsewhere in the taxonomy. Textual Equilibrium Propagation[0] sits squarely in this emerging cluster, contributing a biologically-inspired mechanism for end-to-end prompt refinement in systems where traditional backpropagation is unavailable.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. Optimizing generative AI by backpropagating language model feedback

Authors: Mert Yuksekgonul, Federico Bianchi, Joseph Boen, Sheng Liu, Pan Lu, et al. (8 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

â¦ consider a prompt to be the parameter we want to optimize. TextGrad â¦ , enabling optimization across black-box AI systems. dâ g, â¦ of TextGrad to optimize compound AI systems involving â¦

Relationship Analysis

Both papers belong to the Global Gradient-Based Optimization category, focusing on propagating feedback globally across modules to optimize prompts end-to-end in compound AI systems. The candidate paper (TextGrad) introduces the foundational approach of backpropagating language model feedback through computation graphs, which the original paper (TEP) directly builds upon and critiques. The key difference is that TEP addresses TextGrad's depth-scaling limitations (exploding/vanishing textual gradients) by replacing global backpropagation with a local two-phase equilibrium-based optimization method inspired by energy-based models.

Contributions Analysis

Overall novelty summary. The paper introduces Textual Equilibrium Propagation (TEP) for optimizing prompts in deep compound AI systems, addressing failure modes in long-horizon workflows. It resides in the Global Gradient-Based Optimization leaf, which contains only two papers total. This is a notably sparse research direction within the broader taxonomy of 41 papers across the field, suggesting the work targets an emerging problem space where gradient-inspired optimization methods for multi-module LLM pipelines are still being actively developed.

The taxonomy reveals that prompt optimization for compound systems divides into global versus local strategies, with TEP's leaf focusing on end-to-end feedback propagation. Neighboring leaves include Local Optimization Strategies (module-by-module tuning) and Joint Fine-Tuning approaches (simultaneous weight and prompt updates). The scope note explicitly distinguishes global gradient flow from local methods, positioning TEP alongside one sibling paper that also propagates feedback across all modules. Related branches on Multi-Stage Frameworks and Infrastructure address architectural patterns rather than optimization mechanics, indicating TEP's focus on the optimization algorithm itself rather than system design.

Among 30 candidates examined through semantic search, none clearly refuted any of the three contributions. The identification of exploding and vanishing textual gradient failure modes examined 10 candidates with zero refutations, as did the TEP method itself and the empirical validation component. This suggests that within the limited search scope, the specific framing of depth-scaling failures and the equilibrium-based solution appear distinct from prior work. However, the analysis explicitly notes this is not an exhaustive literature review, leaving open the possibility of relevant work outside the top-30 semantic matches.

Based on the limited search scope, the work appears to occupy a sparsely populated research direction with novel problem framing. The taxonomy structure shows only one sibling paper in the same optimization category, and no examined candidates provided overlapping prior work. The analysis covers top-30 semantic matches plus citation expansion but does not claim exhaustive coverage of all gradient-based prompt optimization literature.

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

Contribution 1: Identification of exploding and vanishing textual gradient failure modes

Description: The authors identify and formalize two critical depth-dependent failure modes in global textual backpropagation for compound AI systems: exploding textual gradients (where feedback grows exponentially with depth) and vanishing textual gradients (where compression causes loss of specificity). These failure modes limit the scalability of existing optimization methods in deep workflows.

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. Physics-informed neural networks: A review of methodological evolution, theoretical foundations, and interdisciplinary frontiers toward next-generation â¦

URL: View paper

Brief Assessment

Physics-informed Networks Review[51] discusses gradient pathology in numerical physics-informed neural networks, not textual gradient failure modes in compound AI systems with language model feedback propagation.

2. Machine learning optimization techniques: a survey, classification, challenges, and future research issues

URL: View paper

Brief Assessment

Machine Learning Optimization Survey[43] mentions 'disappearing gradients and inflating gradients' in the context of traditional deep learning optimization, not textual gradient propagation in compound AI systems with LLM-based feedback chains.

3. Theoretical issues in deep networks

URL: View paper

Brief Assessment

Deep Networks Theory[46] focuses on numerical gradient dynamics in deep neural networks for classification tasks, not textual gradient propagation in compound AI systems. The candidate discusses gradient flow of weight directions and implicit regularization during training, which is fundamentally different from the original paper's analysis of textual feedback propagation failure modes in multi-agent workflows.

4. Understanding and mitigating gradient flow pathologies in physics-informed neural networks

URL: View paper

Brief Assessment

Gradient Flow Pathologies[42] focuses on numerical gradient pathologies in physics-informed neural networks (PINNs) for solving PDEs, not textual gradient propagation in compound AI systems. The failure modes are fundamentally different domains.

5. Theoretical optimization of group size in group normalization for enhanced deep neural network training

URL: View paper

Brief Assessment

Group Normalization Optimization[45] focuses on gradient propagation in deep neural networks through normalization techniques, not on textual gradient failure modes in compound AI systems with LLM-based workflows.

6. Directional convergence and alignment in deep learning

URL: View paper

Brief Assessment

Directional Convergence Alignment[49] studies gradient convergence in deep neural networks with numerical gradients, not textual gradients in compound AI systems. The failure modes analyzed are fundamentally different mathematical phenomena.

7. An enhanced deep neural network with global adaptive weighted gradient for solving hyperbolic partial differential equations

URL: View paper

Brief Assessment

Adaptive Weighted Gradient[50] focuses on numerical gradient pathology in deep neural networks for solving PDEs, not textual gradient failure modes in compound AI systems with LLM-based workflows.

8. Backward gradient normalization in deep neural networks

URL: View paper

Brief Assessment

Backward Gradient Normalization[44] addresses vanishing and exploding gradients in numerical deep neural networks during backpropagation, not textual gradient propagation in compound AI systems with LLM-based workflows.

9. Failures of gradient-based deep learning

URL: View paper

Brief Assessment

Gradient-based Learning Failures[47] focuses on numerical gradient failures in deep neural networks (vanishing/exploding numerical gradients), while the original paper addresses textual gradient failures in compound AI systems where feedback is propagated as natural language text rather than numerical values.

10. Understanding gradient descent on the edge of stability in deep learning

URL: View paper

Brief Assessment

Edge of Stability[48] analyzes gradient descent dynamics in deep neural networks with numerical gradients, focusing on sharpness stabilization and loss oscillations. The original paper addresses textual gradient propagation in compound AI systems with LLM-based feedback chains—a fundamentally different domain.

Contribution 2: Textual Equilibrium Propagation (TEP) method

Description: The authors introduce TEP, a local learning principle inspired by Equilibrium Propagation in energy-based models. TEP consists of two phases: a free phase where local LLM critics iteratively refine prompts until equilibrium, and a nudged phase that applies bounded prompt modifications guided by task objectives via forward signaling rather than backward feedback chains.

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. Developing a hybrid algorithm based on an equilibrium optimizer and an improved backpropagation neural network for fault warning

URL: View paper

Brief Assessment

Hybrid Equilibrium Optimizer[68] focuses on numerical optimization of backpropagation neural networks for fault detection in manufacturing systems, not textual learning principles for compound AI systems with LLMs.

2. Quantum equilibrium propagation: Gradient-descent training of quantum systems

URL: View paper

Brief Assessment

Quantum Gradient Training[64] focuses on quantum systems where equilibrium propagation is applied to quantum Hamiltonians and eigenstates, not textual optimization of LLM-based compound AI systems. The domains are fundamentally different: quantum mechanics versus natural language processing.

3. Scaling equilibrium propagation to deep convnets by drastically reducing its gradient estimator bias

URL: View paper

Brief Assessment

Scaling Equilibrium Propagation[67] focuses on numerical gradient computation in energy-based neural networks for vision tasks, not textual optimization in compound AI systems. The candidate addresses bias reduction in gradient estimators for deep convnets, while the original paper introduces a local learning framework using textual feedback for multi-agent workflows.

4. Equilibrium-Based Learning Dynamics in Spiking Architectures

URL: View paper

Brief Assessment

Equilibrium Spiking Architectures[69] focuses on equilibrium propagation in spiking neural networks for energy-based learning with neuromorphic hardware implementation, not on textual optimization of compound AI systems with LLM-based agents and prompt refinement.

5. Equilibrium Propagation for Periodic Dynamics

URL: View paper

Brief Assessment

Equilibrium Periodic Dynamics[62] focuses on equilibrium propagation for physical dynamical systems (mechanical structures, RLC circuits) operating under physical laws, not textual optimization of LLM-based compound AI systems. The candidate addresses gradient computation in hardware through physical dynamics, while the original addresses textual feedback propagation in multi-agent LLM workflows.

6. Equilibrium propagation for learning in Lagrangian dynamical systems

URL: View paper

Brief Assessment

Lagrangian Equilibrium Propagation[70] focuses on training dynamical systems governed by Lagrangian mechanics through action extremization, not on optimizing compound AI systems with textual feedback. The domains and mechanisms are fundamentally different.

7. Scalable Equilibrium Propagation via Intermediate Error Signals for Deep Convolutional CRNNs

URL: View paper

Brief Assessment

Scalable Equilibrium Propagation[65] focuses on numerical gradient computation in deep convolutional neural networks using energy-based models, not textual feedback optimization in compound AI systems with LLMs.

8. Quantum equilibrium propagation for efficient training of quantum systems based on Onsager reciprocity

URL: View paper

Brief Assessment

Quantum Equilibrium Propagation[63] addresses quantum systems and Hamiltonian optimization through Onsager reciprocity, fundamentally different from TEP's textual feedback optimization in compound AI systems with LLMs. The domains (quantum physics vs. textual AI workflows) and mechanisms (quantum state equilibration vs. textual gradient propagation) are entirely distinct.

9. Holomorphic equilibrium propagation computes exact gradients through finite size oscillations

URL: View paper

Brief Assessment

Holomorphic Equilibrium Propagation[66] focuses on computing exact gradients in energy-based neural networks through complex-valued oscillations, not on textual feedback optimization in compound AI systems with LLM agents.

10. Training and synchronizing oscillator networks with Equilibrium Propagation

URL: View paper

Brief Assessment

Training Oscillator Networks[71] applies equilibrium propagation to physical oscillator networks for hardware computing, not to textual optimization of LLM-based compound AI systems. The domains and problem formulations are fundamentally different.

Contribution 3: Comprehensive empirical validation across multiple benchmarks

Description: The authors provide extensive experimental validation showing that TEP consistently outperforms TextGrad and other baselines across diverse compound AI benchmarks including PubMedQA, STARK-PRIME, HotpotQA, and BigCodeBench, with performance gains that increase as workflow depth grows.

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. Deep research agents: A systematic examination and roadmap

URL: View paper

Brief Assessment

Deep Research Agents[55] focuses on surveying deep research agent systems that integrate dynamic reasoning, adaptive planning, and iterative tool use for open-ended informational research tasks. It does not present a novel optimization method like TEP, nor does it conduct empirical validation comparing optimization approaches across compound AI benchmarks.

2. Multi-Agent Actor-Critic Generative AI for Query Resolution and Analysis

URL: View paper

Brief Assessment

Multi-Agent Actor-Critic[56] focuses on natural language to visualization (NL2Vis) tasks using multi-agent systems for query resolution, not on optimizing compound AI workflows with textual gradients across multi-step QA and tool-use benchmarks like PubMedQA, STARK-PRIME, HotpotQA, and BigCodeBench.

3. Knowledge-Aware Iterative Retrieval for Multi-Agent Systems

URL: View paper

Brief Assessment

Knowledge-Aware Iterative Retrieval[57] focuses on multi-step question answering and retrieval optimization for multi-agent systems, not on optimizing compound AI workflows with textual gradient propagation. The benchmarks and evaluation focus differ fundamentally from TEP's compound AI system optimization.

4. Efficient multi-agent collaboration with tool use for online planning in complex table question answering

URL: View paper

Brief Assessment

Multi-agent Table QA[61] focuses on table question answering with multi-agent collaboration and tool use, evaluating on WTQ, TAT, CRT, and SciTab benchmarks. The original paper evaluates TEP on different benchmarks (PubMedQA, STARK-PRIME, HotpotQA, BigCodeBench) for compound AI system optimization, not table QA specifically.

5. Mac-sql: A multi-agent collaborative framework for text-to-sql

URL: View paper

Brief Assessment

Mac-sql Multi-agent[52] focuses on text-to-SQL parsing with evaluation on BIRD and Spider benchmarks, not on multi-step question answering or multi-agent tool-use benchmarks like PubMedQA, STARK-PRIME, HotpotQA, and BigCodeBench that are central to the original paper's validation.

6. Beyond single-turn: A survey on multi-turn interactions with large language models

URL: View paper

Brief Assessment

Multi-turn Interactions Survey[59] focuses on evaluating multi-turn conversational capabilities across diverse tasks (roleplay, healthcare, education), not on optimizing compound AI workflows with textual equilibrium propagation methods.

7. Chain-of-agents: End-to-end agent foundation models via multi-agent distillation and agentic rl

URL: View paper

Brief Assessment

Chain-of-agents[53] focuses on multi-agent distillation and agentic RL for web/code agents, not on optimizing compound AI workflows with textual equilibrium propagation across PubMedQA, STARK-PRIME, HotpotQA, and BigCodeBench.

8. Multi-Agent System for Comprehensive Soccer Understanding

URL: View paper

Brief Assessment

Soccer Understanding System[58] focuses on soccer-specific multimodal QA tasks with a knowledge base and benchmark construction, not on multi-step question answering and multi-agent tool-use benchmark evaluation for compound AI systems optimization.

9. MedAgentBoard: Benchmarking Multi-Agent Collaboration with Conventional Methods for Diverse Medical Tasks

URL: View paper

Brief Assessment

MedAgentBoard[60] focuses on benchmarking multi-agent collaboration versus single LLMs and conventional methods in medical tasks (medical QA, lay summary generation, EHR prediction, clinical workflow automation). The original paper evaluates TEP on multi-step QA and multi-agent tool-use benchmarks (PubMedQA, STARK-PRIME, HotpotQA, BigCodeBench) to demonstrate advantages of local textual optimization over global backpropagation methods like TextGrad. These are fundamentally different research objectives with distinct technical contributions.

10. Long context scaling: Divide and conquer via multi-agent question-driven collaboration

URL: View paper

Brief Assessment

Long Context Scaling[54] focuses on multi-agent question-driven collaboration for long-context processing (up to 1M tokens), not on optimizing deep compound AI workflows with textual gradients. The benchmarks and evaluation focus differ fundamentally from TEP's compound AI system optimization.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Textual Equilibrium Propagation for Deep Compound AI Systems View paper
- [1] Optimizing instructions and demonstrations for multi-stage language model programs View paper
- [2] Optimizing generative AI by backpropagating language model feedback View paper
- [3] Automatic prompt optimization via heuristic search: A survey View paper
- [4] Fine-tuning and prompt optimization: Two great steps that work better together View paper
- [5] ALTO: An Efficient Network Orchestrator for Compound AI Systems View paper
- [6] Optimas: Optimizing Compound AI Systems with Globally Aligned Local Rewards View paper
- [7] Prompt-a-video: Prompt your video diffusion model via preference-aligned llm View paper
- [8] Optimization modeling and verification from problem specifications using a multi-agent multi-stage LLM framework View paper
- [9] Multi-module GRPO: Composing policy gradients and prompt optimization for language model programs View paper
- [10] Multi-stage prompting for next best agent recommendations in adaptive workflows View paper
- [11] A layered architecture for developing and enhancing capabilities in large language model-based software systems View paper
- [12] MSP: Multi-stage prompting for making pre-trained language models better translators View paper
- [13] Towards hierarchical multi-agent workflows for zero-shot prompt optimization View paper
- [14] Multi-stage prompt tuning for political perspective detection in low-resource settings View paper
- [15] Deep Learning Architectures for Early GI Disease Detection and Stage Prediction View paper
- [16] Prompt-Based Bioinformatic Pipeline Generation for a Multi-Step Metaviral Workflow View paper
- [17] AI-Driven Optimization Framework for Construction Site Ecosystems View paper
- [18] GraphPrompter: Multi-Stage Adaptive Prompt Optimization for Graph In-Context Learning View paper
- [19] Orchestrating strategic control points in layered AI ecosystems: insights from startup founders and investors View paper
- [20] From prompt to production: a design process pipeline using AI View paper
- [21] Liberal Entity Matching as a Compound AI Toolchain View paper
- [22] From Keyframes to Narrative: A Multi-stage AI Pipeline for Scene Understanding Using Object Detection and Large Language Models View paper
- [23] Phase-Adaptive LLM Framework with Multi-Stage Validation for Construction Robot Task Allocation: A Systematic Benchmark Against Traditional Optimization â⌡ View paper
- [24] Large Language Models for Real-World Nutrition Assessment: Structured Prompts, Multi-Model Validation and Expert Oversight View paper
- [25] XML Prompting as Grammar-Constrained Interaction: Fixed-Point Semantics, Convergence Guarantees, and Human-AI Protocols View paper
- [26] Integrating Textual Queries with AI-Based Object Detection: A Compositional Prompt-Guided Approach View paper
- [27] Enhancing AI-Driven Psychological Consultation: Layered Prompts with Large Language Models View paper
- [28] Towards a Modular and Adaptive AI Architecture for Personalized Intelligent Tutoring Systems View paper
- [29] AI-Powered Roadmap Generator Prompt Engineering View paper
- [30] Multi-Stage Prompt Inference Attacks on Enterprise LLM Systems View paper
- [31] Leveraging AI for More Effective Teams: Evidence-Based Strategies for Diverse Idea Generation View paper
- [32] SocialBrain View paper
- [33] Demystifying LLM Serving Pipeline: From Prompt to Response View paper

- [34] A Multi-Stage Prompt Framework for High-Quality News Summarization with Large Language Models View paper
- [35] Prompt-Layered Architecture: A New Stack for AI-First Product Design View paper
- [36] Prompts for Planning-Ai Integration: Effective Prompt Design for Large Language Models in Support of Sustainable Urban Development View paper
- [37] Multi-Stage Prompting for Knowledgeable Dialogue Generation View paper
- [38] A Multi-Stage Self-Optimization Framework for LLM Reasoning: Exploration, Structured Improvement, and Robust Inference View paper
- [39] AI-based Information Retrieval from Structured Text Documents View paper
- [40] General Purpose Technologies vs. AI: Comparative Diffusion Rates and Economic Impact View paper
- [41] BUILDING MODULAR AND SCALABLE AI FOR CONTENT CREATION USING LLAMAINDEX View paper
- [42] Understanding and mitigating gradient flow pathologies in physics-informed neural networks View paper
- [43] Machine learning optimization techniques: a survey, classification, challenges, and future research issues View paper
- [44] Backward gradient normalization in deep neural networks View paper
- [45] Theoretical optimization of group size in group normalization for enhanced deep neural network training View paper
- [46] Theoretical issues in deep networks View paper
- [47] Failures of gradient-based deep learning View paper
- [48] Understanding gradient descent on the edge of stability in deep learning View paper
- [49] Directional convergence and alignment in deep learning View paper
- [50] An enhanced deep neural network with global adaptive weighted gradient for solving hyperbolic partial differential equations View paper
- [51] Physics-informed neural networks: A review of methodological evolution, theoretical foundations, and interdisciplinary frontiers toward next-generation â¦ View paper
- [52] Mac-sql: A multi-agent collaborative framework for text-to-sql View paper
- [53] Chain-of-agents: End-to-end agent foundation models via multi-agent distillation and agentic rl View paper
- [54] Long context scaling: Divide and conquer via multi-agent question-driven collaboration View paper
- [55] Deep research agents: A systematic examination and roadmap View paper
- [56] Multi-Agent Actor-Critic Generative AI for Query Resolution and Analysis View paper
- [57] Knowledge-Aware Iterative Retrieval for Multi-Agent Systems View paper
- [58] Multi-Agent System for Comprehensive Soccer Understanding View paper
- [59] Beyond single-turn: A survey on multi-turn interactions with large language models View paper
- [60] MedAgentBoard: Benchmarking Multi-Agent Collaboration with Conventional Methods for Diverse Medical Tasks View paper
- [61] Efficient multi-agent collaboration with tool use for online planning in complex table question answering View paper
- [62] Equilibrium Propagation for Periodic Dynamics View paper
- [63] Quantum equilibrium propagation for efficient training of quantum systems based on Onsager reciprocity View paper
- [64] Quantum equilibrium propagation: Gradient-descent training of quantum systems View paper
- [65] Scalable Equilibrium Propagation via Intermediate Error Signals for Deep Convolutional CRNNs View paper
- [66] Holomorphic equilibrium propagation computes exact gradients through finite size oscillations View paper
- [67] Scaling equilibrium propagation to deep convnets by drastically reducing its gradient estimator bias View paper
- [68] Developing a hybrid algorithm based on an equilibrium optimizer and an improved backpropagation neural network for fault warning View paper
- [69] Equilibrium-Based Learning Dynamics in Spiking Architectures View paper
- [70] Equilibrium propagation for learning in Lagrangian dynamical systems View paper
- [71] Training and synchronizing oscillator networks with Equilibrium Propagation View paper