Academic Simulacra: Forecasting Research Ideas through Multi-Agent LLM Simulations

JIWOONG CHOI, University of Chicago, USA
DONGHYUN KANG, University of Chicago, USA
YINGRONG MAO, University of Chicago, USA
JAMES EVANS, University of Chicago, USA

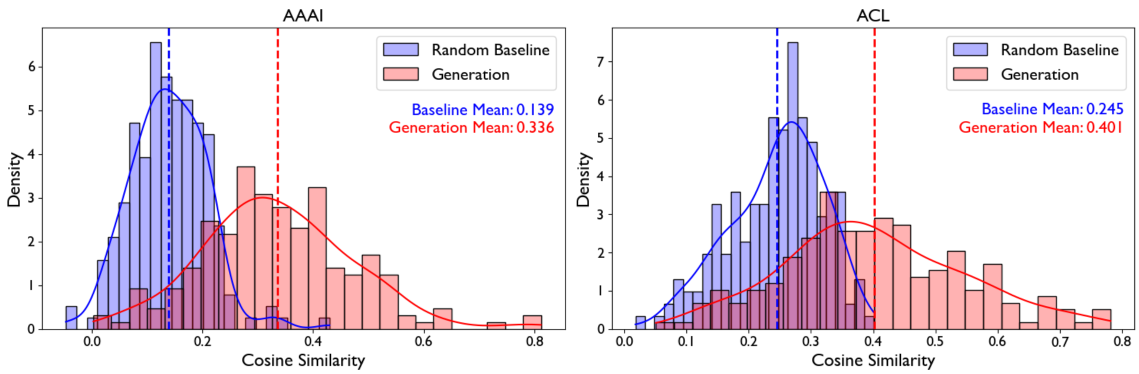

Fig. 1. Cosine Similarity Distributions for AAAI and ACL: Actual vs. Random (Blue) and Actual vs. Generated (Red)

We introduce a multi-agent simulation framework for forecasting research ideas using "scholar agents" powered by large language models (LLMs). We instantiate approximately 2,686 scholar agents based on their publication histories prior to 2024 and simulate discussions to collectively generate key research ideas for 1,400 papers targeting seven major computer science conferences in 2024. We then evaluate the proximity of these generated ideas by comparing their semantic embeddings with those of the actual target papers written by the corresponding researchers. Our results suggest that LLM-based multi-agent simulations yield substantially higher similarity scores with real publications than two baselines: (1) the average pairwise similarity among papers within the same 2024 conference, and (2) a random set of past papers from the same conference. This demonstrates the predictive capacity of our scholar agent framework. We then further analyze how diversity in ethnic composition and institutional affiliations may correlate with the predictability of research, or inversely, the degree of surprise relative to the past. Our preliminary analysis suggests that the least predictable and thus most surprising research ideas emerge from teams affiliated with Chinese institutions but not composed of ethnically Chinese authors. These findings offer promising initial evidence that simulating knowledge-driven scholar agents can anticipate directions of scientific discovery and help explain the influence of social and institutional factors on innovation.

Authors' addresses: Jiwoong Choi, University of Chicago, Chicago, Illinois, USA, giochoi@uchicago.edu; Donghyun Kang, University of Chicago, Chicago, Illinois, USA, kangd@uchicago.edu; Yingrong Mao, University of Chicago, Chicago, Illinois, USA, yingrong@uchicago.edu; James Evans, University of Chicago, Chicago, Illinois, USA, jevans@uchicago.edu.

Manuscript submitted to ACM

CCS Concepts: • **Computing methodologies → Natural language generation**; **Discourse, dialogue and pragmatics**; **Reasoning about belief and knowledge**; **Agent / discrete models**.

Additional Key Words and Phrases: Simulated Scholarship, Large Language Models, Collective Intelligence

ACM Reference Format:
Jiwoong Choi, Donghyun Kang, Yingrong Mao, and James Evans. 2025. Academic Simulacra: Forecasting Research Ideas through Multi-Agent LLM Simulations. In *Proceedings of ACM COLLECTIVE INTELLIGENCE (CI '2025).* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The recent development of Large Language Models (LLMs) has opened new possibilities for automated knowledge generation while simultaneously challenging traditional research paradigms. The scientific research process is typically viewed as the dynamic interplay between established knowledge and innovative reasoning. Since Don Swanson's classic work on Literature-Based Discovery [6], the prospect of automated systems that may mimic scholarly intuition has intrigued researchers—a process we now seek to simulate computationally. This research explores two fundamental questions: Can we create an agent that authentically replicates the thinking process of real-world scholars? And can such simulations reveal patterns that emerge from underlying social and institutional factors?

By employing iterative refinement and novelty optimization, researchers have begun to harness the vast pool of existing literature in new ways. Notable contributions include Baek et al. [1], who developed LLM-based tools for generating novel research ideas, and Wang et al. [8], whose frameworks directly optimize for novelty by combining retrieval with iterative hypothesis refinement. These advances suggest that LLMs may play an essential role in expanding scientific horizons.

In this study, we propose and apply a multi-agent framework powered by state-of-the-art LLMs to predict future research ideas appearing in major computer science conferences.

2 METHODOLOGY AND EXPERIMENT

2.1 Dataset

We initially collected all papers published in seven major computer science conferences in the field of artificial intelligence and related areas, sourced primarily from Semantic Scholar [2]. These conferences include ICML, ICLR, AAAI, CVPR, EMNLP, ECCV, and ACL (see Table A in the Appendix for full names). The resulting dataset comprises **435,967 papers** published up to 2024 and **51,396 authors** with corresponding IDs. This allows us to compile each author's publication history, including paper titles, abstracts, author names, and affiliations.

One of the major challenges we aim to tackle in our multi-agent discussions is preventing future data leakage—that is, keeping future information out of the generation pipeline. To ensure a strict causal forecasting setup, we include only papers formally published in 2024 as prediction targets. In other words, we exclude any papers that appeared as a preprint before 2024, given the common practice in computer science of posting work on repositories like arXiv prior to official conference presentations.

2.2 Large Language Models Setup

To prevent data leakage, we restricted ourselves to language models with a knowledge cutoff before 2024. All multi-agent idea generation was conducted on a local machine equipped with an NVIDIA RTX 3090 GPU, 48 CPU cores, and 128

GB of RAM. Given this resource-constrained environment, we employed Ollama[3], which allows users to run large language models (LLMs) locally.

We tested various combinations of model configurations, including quantization formats (e.g., FP16, Q4, Q8), parameter sizes (e.g., 3B, 8B, 70B), and the number of agents involved in the simulations. We also examined the effect of longer context windows on performance, which came at the cost of significantly slower inference. For example, when using a context window of 128k tokens with the LLaMA 3.1 8B model, we found that it took more than 30 minutes to generate an initial response from a single agent.

Based on this exploration, we selected LLaMA 3.1 8B, quantized to Q4_K_M, with a knowledge cutoff in December 2023. To strike a practical balance between prompt completeness and computational efficiency, we limited the context length to 8,192 tokens—or roughly 6,000 to 6,500 words—which was sufficient for our use case. This configuration also ensured that the model would not be aware of any publications made public in any form after the cutoff date. We set the temperature to 0.3 to balance creativity and reproducibility.

For the ethnicity inference task, which involved less text but remained central to our workflow, we used a more advanced model, LLaMA 3.3, to infer authors' ethnicity based solely on their given and surnames.

2.3 Multi-Agent Simulation Framework

Figure 2 shows the framework of our simulation. We built a multi-agent environment with Autogen [9], a framework that allows multiple LLM-powered agents to communicate, critique, and refine ideas under different roles. In our experiment, 'Conversable Agent' class of Autogen was customized to implement each scholar agent. Specifically, agents were instantiated based on authorship positions: for papers with more than three authors, only the first and last authors were instantiated, reflecting that the first and last authors typically contribute the most to a research paper [4]. For papers with three or fewer authors, all authors were instantiated. Each agent's "system prompt" was enriched with information from up to 10 of their most recent publications (titles and abstracts) in which they appeared as either the first or last author. Additionally, we implemented a Research Assistant Agent responsible for summarizing discussions among scholar agents and exporting the results in JSON format.

In addition, we designed role-specific system prompts to guide a collaborative research process. (See Appendix B) The first author is tasked with brainstorming a variety of research directions and topics based on their prior work, and is instructed to iteratively refine their ideas through feedback (from the last author) until they feel satisfied with the results. The last author, serving as corresponding author, is responsible for critically evaluating, selecting, and synthesizing these ideas while also generating appropriate paper titles. These agents are also asked to continuously revise their proposals until they are satisfied with the final collection of key ideas. Additionally, we deploy a research assistant agent that follows the scholars' discussion and extracts and summarizes the essential content.

We sampled 200 papers per conference published in 2024 for the target set, leading to 1,400 total simulations.

2.4 Evaluation

2.4.1 Key Idea Generation and Benchmark Extraction . As previously outlined, we implemented multi-scholarly agent discussions using Llama 3.1, with model specifications detailed in the previous section. Each simulation output was prompted to generate key research ideas, typically yielding one sentence, and up to two. This choice was inspired by a recent framework proposed by Zhang et al.[10], in which paper abstracts are mapped onto five aspects—context, key idea, method, outcome, and projected impact—allowing analysts to focus on specific dimensions of research. In our case, we naturally focus on the "Key Idea," which underpins a given research project. This focus also reflects our computing

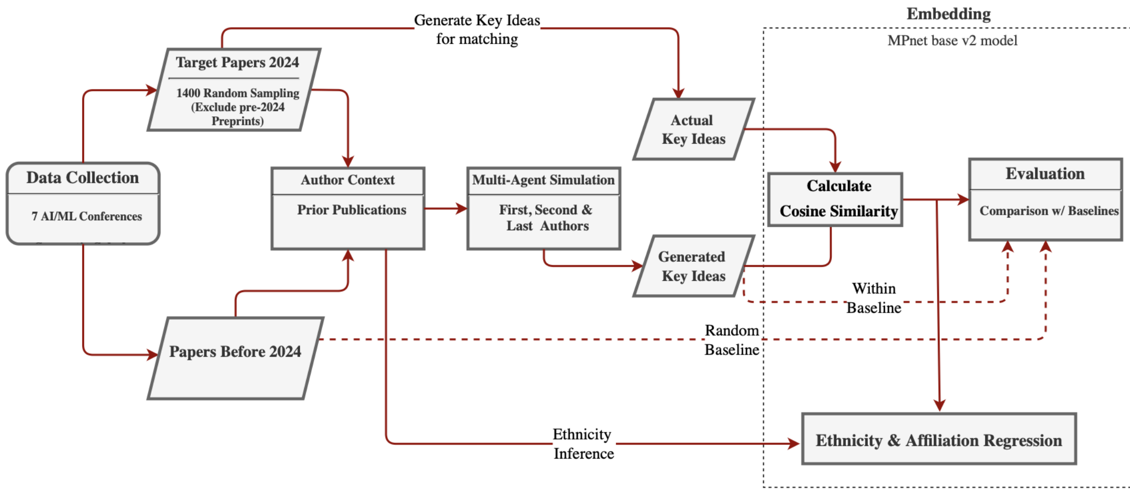

Fig. 2. Multi-Agent Simulation Framework

resource constraints, as described earlier. To evaluate generated key ideas, we compare them with those extracted from target abstracts. The system prompt used for key idea extraction is provided in Appendix C; a slightly modified version of the one proposed by Zhang et al.[10]

2.4.2 Semantic Representation through Vector Embeddings. To assess the similarity between generated and actual paper key ideas, we transformed both into a semantic embedding space and computed the cosine similarity between the associated vectors. We employed the 'all-mpnet-base-v2' model [5] for this task, as it generally achieves high accuracy in semantic similarity evaluations and provides more robust embeddings for nuanced text comparisons than the lighter all-MiniLM-L6-v2 model. After mapping the texts into a 768-dimensional space, the cosine similarity between embeddings was calculated.

2.4.3 Baseline Comparisons. We established two baselines to validate our model's performance.

1. Random Baseline: This baseline (See Table 1) involves calculating the cosine similarity between a randomly selected paper published before 2024 (within the same conference) and an actual paper published in 2024. We experimented with sampling one, two, and three random papers, averaging the cosine similarities for the two- and three-sample cases. The results were as follows: One sample mean = 0.198, std = 0.102; Two samples: mean = 0.211, std = 0.082; Three samples: mean = 0.218, std = 0.080.

Based on these findings, we decided to use three random samples per actual paper for the baseline. Notably, domain-specific conference papers (e.g., CVPR, ACL) exhibited higher similarity scores, whereas broader computer science conferences (e.g., AAAI) did not. (See Table 1)

2. Within-Actual Baseline: This baseline (see Table 2) measures the pairwise cosine similarity among all 200 actual key ideas (hence 19,900 unique pairs) from the same conference. It allows us to evaluate, on average, how key ideas from a given conference are situated within the semantic embedding space. More importantly, we posit that this baseline can help assess whether multi-agent discussions generate key ideas similar to the actual ones in context while remaining distinguishable from those in other papers within the same conference.

2.5 Ethnicity & Affiliation Regression Analysis

We aimed to investigate the relationship between surprising research outcomes and the ethnicity and affiliation of scholars. Llama3.3 inferred that 59.2% of the 51,396 scholars in our dataset are of Chinese ethnicity (note that Chinese Americans are classified as Chinese; see Appendix D.

Given the substantial representation of Chinese ethnicity in computer science academia, we conducted a Chinese-specific analysis. To this end, we estimated the following regression model:

$$\begin{aligned}\text{CosineSimilarity}_i &= \alpha + \beta_1\,\text{RatioNonChinaEthnicity}_i + \beta_2\,\text{RatioChinaAffiliation}_i \\ &+ \beta_3\,(\text{RatioNonChinaEthnicity}_i \times \text{RatioChinaAffiliation}_i) + \sum_c \gamma_c\,\mathbf{1}\{\text{conference} = c\} + \varepsilon_i. \quad (1)\end{aligned}$$

In this model, $\text{RatioNonChinaEthnicity}_i$ represents the proportion of coauthors for paper i inferred by Llama3.3 to be of non-Chinese ethnicity regardless of nationality, and $\text{RatioChinaAffiliation}_i$ denotes the proportion of coauthors whose current institutional affiliations are in China. This specification allows us to determine whether variations in ethnicity or affiliation are associated with unexpected research outcomes and hence potentially novel and surprising.

3 RESULTS

On average, simulating a single paper took approximately 1.5 minutes and involved 15 conversational turns. The research assistant agent then summarized the discussion, extracting an average of 7.44 key ideas (SD = 3.31). Both generated key ideas and those from actual papers were embedded into the embedding space. Given that scholars often generate multiple ideas while a paper typically represents only one, we measured similarity by selecting the maximum cosine similarity among all pairs between generated and actual key ideas for each paper.

3.1 Semantic Clustering

To assess the robustness of our embedding approach, we projected the embeddings of every generated key idea using t-SNE [7]. The visualization (see Figure 3) confirms that the embeddings reflect underlying semantic relationships. Notably, the left side appears to represent Computer Vision (See ECCV and CVPR), while the right side corresponds to Natural Language Processing (See EMNLP and ACL).

3.2 Comparison with Baselines

Table 3 shows the cosine similarity between the generated key ideas and the actual papers ones. For example, the result for CVPR showed a mean of 0.403 with a standard deviation of 0.138. Overall, compared to the two baselines (See Table 1, Table 2), the mean cosine similarity between the generated key ideas and the actual ones is 74.5% higher than that of the random baseline and 76.9% higher than that of the within baseline.

Figure 4 compares how closely our generated key ideas match the actual ones relative to a random paper from the same venue. The histogram for cosine similarity of generated ideas between actual key ideas (red) is consistently shifted toward higher values across all conferences. Notably, the mean similarity of our generated ideas often exceeds the 95th percentile of the random baseline. This result suggests that our simulation is not merely repeating past publication data but instead generates novel ideas from the past.

Figure 5 shows that the mean dispersion of cosine similarity across actual key ideas varies within the same conference. The rationale for this baseline is that if multi-agent discussion generates meaningful and distinguishable key ideas relative to other articles published in the same venue, the distribution of generated key ideas would shift to the right

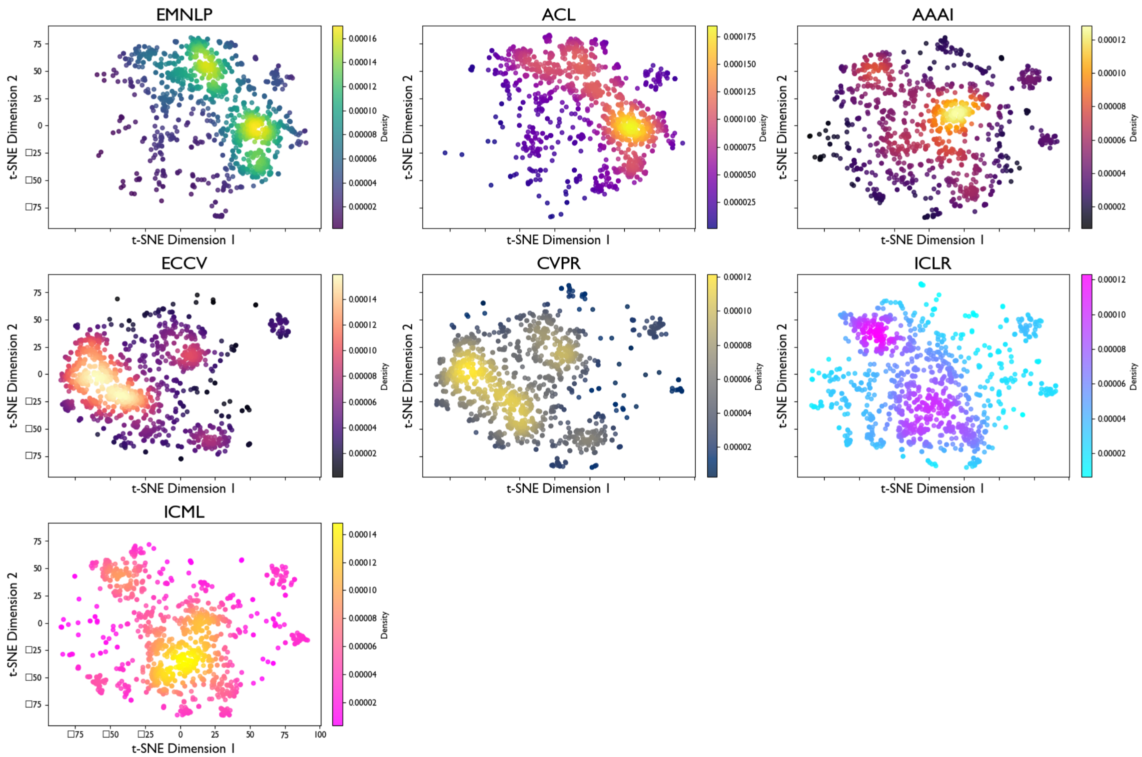

Fig. 3. t-SNE clustering of the generated key ideas embeddings by conference

compared to the within-conference baseline—indicating higher similarity. This shift would suggest that each paper's simulation retains distinct characteristics, which is confirmed.

Table 1. Random Baseline

Conference	Mean	Std
AAAI	0.139	0.068
ACL	0.245	0.075
CVPR	0.247	0.067
EMNLP	0.220	0.071
ECCV	0.225	0.076
ICLR	0.224	0.084
ICML	0.222	0.062

Table 2. Within-Actual Baseline

Conference	Mean	Std
AAAI	0.173	0.107
ACL	0.239	0.115
CVPR	0.240	0.106
EMNLP	0.249	0.123
ECCV	0.223	0.110
ICLR	0.204	0.109
ICML	0.194	0.108

Table 3. Gen. vs. Actual

Conference	Mean	Std
AAAI	0.336	0.136
ACL	0.401	0.147
CVPR	0.403	0.138
EMNLP	0.371	0.133
ECCV	0.379	0.128
ICLR	0.391	0.141
ICML	0.375	0.141

Qualitative inspection of the *top-5* simulation matches suggests that the agents sometimes predicted strikingly specific topics (e.g., event-based camera, neural radiance field, geospatial synthetic data, etc.). (See Appendix E)

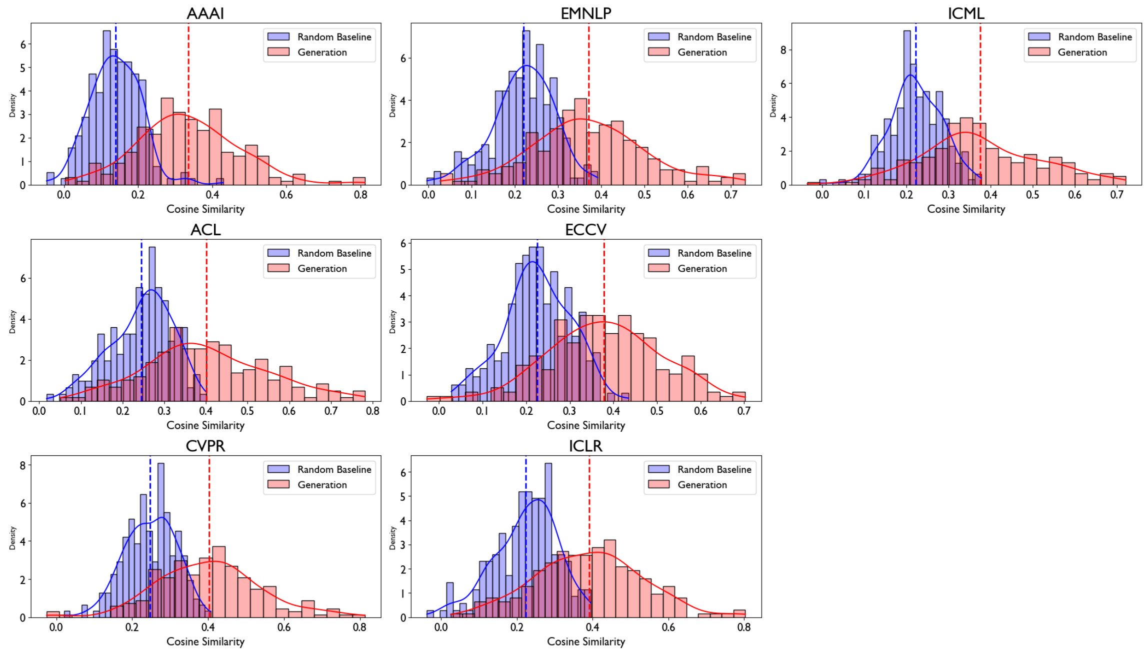

Fig. 4. Random Baseline v.s. Generation

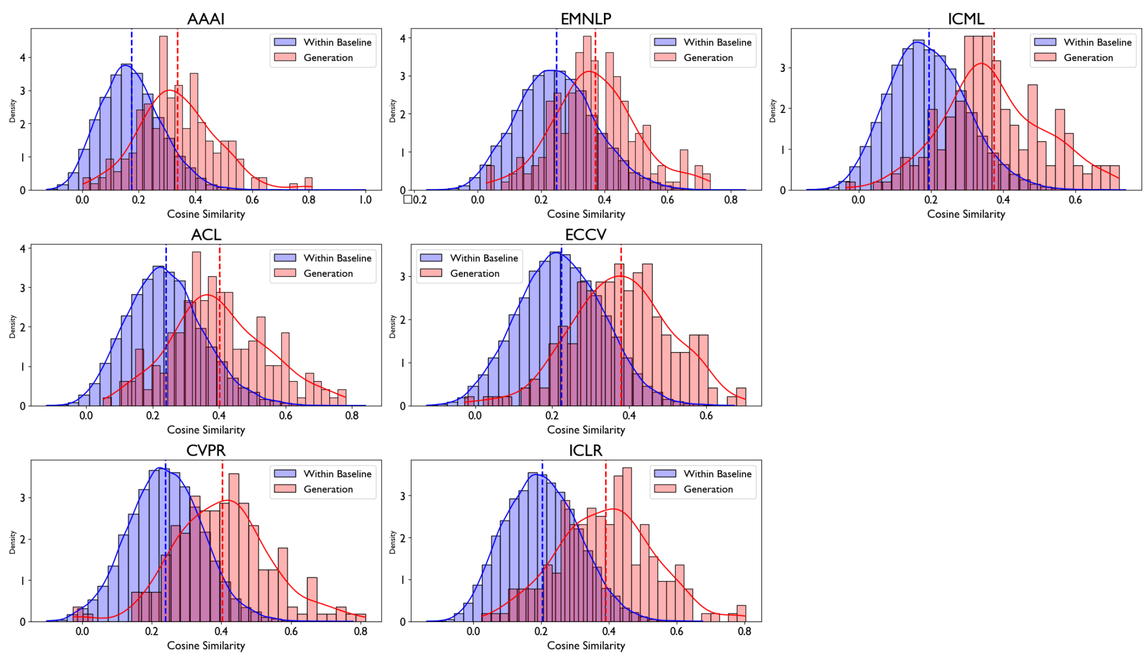

Fig. 5. Within-Actual Baseline v.s. Generation

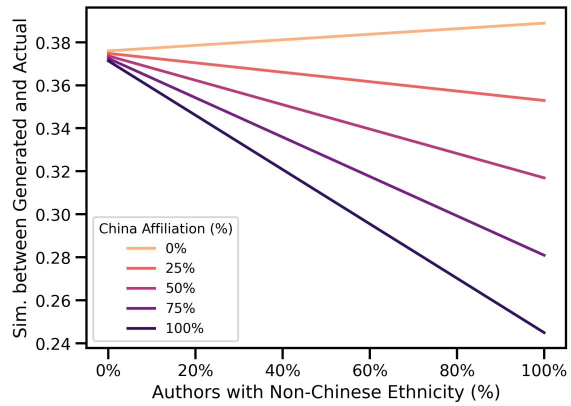

Fig. 6. Interaction between Author Ethnicity and Institutional Affiliation

3.3 Ethnicity & Affiliation Regression Results

We posit that the more similar the key idea generated from our multi-agent simulation is to the actual target key idea, the more predictable the research idea becomes. By contrast, if the generated key idea does not approximate the actual one, we regard it as less predictable—more surprising—and thus potentially more novel relative to the authors' prior work.

As outlined above, we obtain vector representations of both the generated key ideas and key ideas extracted from actual abstracts using the *all-mpnet-base-v2* model [5], which produces 768-dimensional embeddings. We compute the cosine similarity between these two vector representations. In other words, a higher cosine similarity indicates greater alignment between generated and actual key ideas—hence, higher predictability—or vice versa.

Then, as described in Eq. (1), we regress the cosine similarity score on the proportion of authors affiliated with China-based research institutions, the proportion inferred to be of Chinese ethnicity, and their interaction. The analysis suggests that the least predictable—and thus potentially most surprising—research ideas emerge from teams affiliated with Chinese institutions but composed of ethnically non-Chinese researchers. Conversely, among teams based outside China, the presence of ethnically Chinese researchers is associated with more surprising research.

These findings offer preliminary evidence that academic simulacra can anticipate the direction of scientific discovery, but to varying degrees depending on the influence of social and institutional factors on innovation.

4 CONCLUSION

In this work, we propose a framework in which multi-agent simulations are instantiated through scholar personas. This approach demonstrates the potential to computationally simulate the generation of research ideas. In experiments simulating 1,400 papers, we find that our generated key idea forecasts exhibit stronger alignment with key ideas extracted from actual 2024 conference publications than two plausible baselines: (1) the similarity between a random set of past papers from that conference and the actual one, and (2) the pairwise similarity among actual 2024 papers within the same conference. These results suggest that LLM-based collective academic simulacra can meaningfully emulate core elements of future scholarly contributions.

Our preliminary analysis of the relationship between sociocultural team composition and prediction further highlights the role of diversity in generating unexpected or surprising ideas. Specifically, we find that papers involving non-Chinese ethnic authors affiliated with Chinese institutions—or conversely, Chinese ethnic researchers based outside of China—are more likely to produce less predictable, and potentially more innovative research. These findings suggest that diversity can be a potential driver toward intellectual innovation.

Taken together, this work offers a path toward how multi-agent simulations can be used to model and even anticipate the evolution of scientific knowledge. Rather than simply mirroring existing patterns, these systems hold promise for actively exploring the generative frontiers of scholarship.

REFERENCES

[1] Jinheon Baek, Sujay Kumar Jauhar, Silviu Cucerzan, and Sung Ju Hwang. 2024. Researchagent: Iterative research idea generation over scientific literature with large language models. *arXiv preprint arXiv:2404.07738* (2024).

[2] Suzanne Fricke. 2018. Semantic scholar. *Journal of the Medical Library Association: JMLA* 106, 1 (2018), 145.

[3] Ollama. 2025. Ollama. https://ollama.com.

[4] Henry Sauermann and Carolin Haeussler. 2017. Authorship and contribution disclosures. *Science advances* 3, 11 (2017), e1700404.

[5] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. Mpnet: Masked and permuted pre-training for language understanding. *Advances in neural information processing systems* 33 (2020), 16857–16867.

[6] Don R Swanson. 1986. Undiscovered public knowledge. *The Library Quarterly* 56, 2 (1986), 103–118.

[7] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).

[8] Qingyun Wang, Doug Downey, Heng Ji, and Tom Hope. 2023. Scimon: Scientific inspiration machines optimized for novelty. *arXiv preprint arXiv:2305.14259* (2023).

[9] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Shaokun Zhang, Erkang Zhu, Beibin Li, Li Jiang, Xiaoyun Zhang, and Chi Wang. 2023. Autogen: Enabling next-gen llm applications via multi-agent conversation framework. *arXiv preprint arXiv:2308.08155* 3, 4 (2023).

[10] Xingjian Zhang, Yutong Xie, Jin Huang, Jinge Ma, Zhaoying Pan, Qijia Liu, Ziyang Xiong, Tolga Ergen, Dongsub Shim, Honglak Lee, et al. 2024. MASSW: A new dataset and benchmark tasks for ai-assisted scientific workflows. *arXiv preprint arXiv:2406.06357* (2024).

APPENDIX

A FULL NAMES OF CONFERENCES

Abbreviation	Full Name
ICML	International Conference on Machine Learning
ICLR	International Conference on Learning Representations
AAAI	AAAI Conference on Artificial Intelligence
CVPR	Computer Vision and Pattern Recognition
EMNLP	Conference on Empirical Methods in Natural Language Processing
ECCV	European Conference on Computer Vision
ACL	Annual Meeting of the Association for Computational Linguistics

Table 4. Full names of selected AI and Computer Science conferences.

B SYSTEM PROMPT EXAMPLES

```
first_author_sysmessage = (
    f"You are a researcher named {first_author_name}. "
    f"You have done these works before: {history_text}"
    "You're cooperating with other researchers to produce a new paper. "
    "You should come up with multiple research directions and topics, "
```

```
    "and describe your ideas towards others. "
    "After obtaining others' feedback, you will revise your answer "
    "based on their suggestions. "
    "When you feel satisfied about the results and want to end the "
    "discussion, say 'bye' to others"
)
last_author_sysmessage = (
    f"You are a researcher named {last_author_name}. "
    f"You have done these works before: {history_text}"
    "You're cooperating with other researchers to produce a set of "
    "paper topics and ideas. "
    "You will select, evaluate, criticize, and revise others' ideas. "
    "You should be responsible for generating titles and modifying ideas. "
    "You will keep suggesting until you're satisfied with the titles "
    "and key ideas. When you're satisfied and want to end the "
    "conversation, say 'bye'. "
)
ra_message = (
    "You are an AI research assistant. "
    "You should extract the results from the current conversation. "
    "The results should be the titles and the full abstracts for the new papers. "
    "Use the following format for each paper: "
    "<paper> paper number </paper>"
    "<title> paper title </title>"
    "<abstracts> paper abstracts </abstracts>"
)
```

C KEY IDEA EXTRACTION PROMPT

```
SYSTEM_PROMPT = r"""
You are an expert in computer science. Your task is to summarize the following five aspects
of the papers given the definitions below.

Definitions of Aspects

Context
- The status quo of related literature or reality which motivated this study. This could normally
be a problem, a research question, or a research gap that has not been successfully addressed
by previous work.
- Anything happened before this study.
```

```
Key Idea
- The main intellectual merit of this paper, often in comparison to the context. This could
normally be a novel idea or solution proposed in this paper that distinguishes it from what’s
already done in literature.
- Proposed in this study

Method (Validation Methodology)
- The specific experiment or proof that investigates and validates the key idea.
- CS papers often refer "Method" as algorithm or model, but our definition here is **different**.
- Performed in this study.

Outcome
- The factual statement about the study output. This could be the experiment results and any
other measurable outcome that has occurred. It marks whether the key hypothesis is testified
or not.
- Produced in this study.

Future Impact
- The impact of the work on the field explicitly anticipated by the authors, and potential
further research explicitly identified by the author that may improve or extend this study.

Notes
- If an aspect is NOT mentioned in the abstract, mark it as "N/A" (not applicable). DO NOT
come up with your own interpretation.
- Each aspect should be summarized in 1-2 sentences in most cases.
- Each aspect should be self-contained and should not contain references including other
aspects (cross-reference).
- Including specific names of proposed models, datasets, etc., in the summary is acceptable.
- If the problem definition is novel (e.g., proposing a new task), classify it as a Key Idea.
- Non-measurable outcomes should be categorized as Future Impact.
- Impacts that have already occurred should be considered as Outcome.
- A new observation that motivates the proposal of a key idea should be classified under Key Idea.
- Future Impact should not account for real impacts, such as the number of citations a paper
has received.

Response Format
The response should be a JSON object in the following format:
{
  "Context": "...",
  "Key Idea": "...",
  "Method": "...",
```

```
  "Outcome": "...",
  "Future Impact": "..."
}
"""
```

D INFERRED ETHNICITY OF ALL AUTHORS

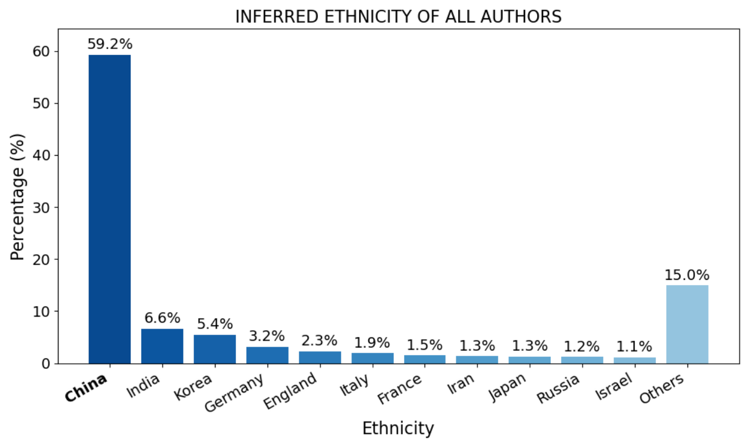

Fig. 7. Inferred Ethnicity Distribution (Percentage)

E TOP 10 GENERATED KEY IDEAS BY COSINE SIMILARITY ACROSS ALL CONFERENCES

Received 1 April 2025

Table 5. Top 10 Generated Key Ideas by Cosine Similarity Across All Conferences

Generated Key Ideas	**Actual Key Ideas**	**Cos. Similarity**	**Conf.**
This paper presents a novel approach to photometric stereo using event cameras. The authors discuss the potential benefits of using event cameras in photometric stereo, including increased accuracy and robustness under dynamic lighting conditions. They also provide examples of applications where event-based photometric stereo can be useful.	A novel approach called EventPS uses an event-based camera to reconstruct scenes with high temporal resolution.	0.813	CVPR
We propose a neural radiance field approach for real-time physics-based rendering, incorporating physical laws and constraints into the rendering process. We also explore the use of neural rendering techniques that incorporate physics-based models, such as radiosity or photon mapping.	The authors propose Neural Illumination Fields for synthesizing high-quality images under varying lighting conditions.	0.812	AAAI
This paper conducts a comparative study on the use of gradient-based methods for training normalizing flows. The authors investigate how to adapt path-gradient estimators for different types of normalizing flows and explore the benefits of using gradient-based methods.	Proposing a fast path gradient estimator that improves optimization performance in deep networks.	0.803	ICLR
This paper explores a scenario where agents have different preferences or objectives in dynamic matching markets. It designs algorithms that can handle this heterogeneity and provides regret bounds for these algorithms.	Proposing two new algorithms (AETDA and ODA) that enhance multi-agent decision-making in dynamic environments.	0.795	AAAI
Further investigate the robustness of quantum algorithms for finding approximate second-order stationary points in non-convex optimization problems and explore new methods for improving robustness.	Developing quantum algorithms for minimizing the error in quantum computations, leading to more robust results.	0.789	ICLR
Data augmentation techniques specifically designed for geospatial tasks can be used to generate synthetic data. This approach improves model robustness and generalizability, reducing overfitting.	Proposing a large-scale augmentation method for generating high-quality synthetic data for new environments using readily available geospatial data to improve text-based geospatial reasoning.	0.781	ACL
This study applies contrastive learning to implicit discourse relation recognition, exploring its effectiveness in improving model performance on this task. The research evaluates how the approach affects the model's ability to generalize across different types of implicit discourse relations and domains.	Exploring multi-label classification frameworks to handle implicit discourse relation recognition, showing that these methods do not depress performance for single-label prediction.	0.781	ACL
This paper proposes a meta-learning framework for online reinforcement learning that enables agents to quickly adapt to new tasks and environments. The framework combines techniques from meta-safe reinforcement learning, robust control theory, and stochastic optimization to ensure safety and stability in uncertain environments.	This paper proposes meta-safe reinforcement learning (Meta-SRL) through a CMDP-within-online framework to provide provable guarantees for reward maximization and constraint satisfaction.	0.769	ICLR
This paper explores the practical applications of text-to-3D generation in real-world scenarios such as architecture, product design, and interior decoration. The results demonstrate improved performance and efficiency in these domains.	Proposing Sherpa3D, a new text-to-3D framework that achieves high-fidelity, generalizability, and geometric consistency simultaneously by fully exploiting coarse 3D knowledge to enhance prompts and guide 2D lifting optimization for refinement.	0.765	CVPR
This paper presents a transfer learning approach for hateful memes detection in low-resource languages. We use multimodal fusion techniques to combine text and image features, and evaluate the effectiveness of different pre-trained language models.	The authors introduce a novel multimodal dataset for Bengali (BHM) and propose DORA, a multimodal deep neural network that extracts significant modality features from memes to detect hateful content and its targeted entities.	0.757	ACL