Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

Sheshera Mysore

Senior Applied Scientist, Microsoft

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to Criterion 5 (original contributions of major significance). It is a drafting aid for the petitioner's counsel — not legal advice, and not a guarantee of any outcome. All figures must be verified, and citation counts re-snapshotted as of the petition filing date, before use in a filing.

A. Overview & Filtering Statement

1,209	1,384	26	14
Citing papers mapped	Citation edges	Home papers mapped	h-index (GS)

Filtering statement — methodology & limits

Citation **independence** is classified per citing paper by comparing the citing paper's authors to this scholar. *Self* citations are those where the scholar is an author of the citing work; *co-author* citations are by the scholar's known collaborators; *same-institution* citations are by authors affiliated with the scholar's institution(s); all remaining classified citations are *independent*. Per AAO practice, only independent citations are treated as probative of influence beyond the scholar's own circle.

Known limitations — counsel must verify. (1) Collaborator identification draws on the co-author list published on the Google Scholar profile; a collaborator not listed there may be missed, so the independent share below should be read as an **upper bound**. (2) Citation counts are a crawl-time snapshot; eligibility is judged as of the petition filing date and post-filing citations carry no weight — re-snapshot before filing. (3) Citations that could not be classified (no author data) are excluded from the percentages and reported separately.

B. Citation Independence

The AAO credits citations only where they show influence **beyond the scholar's own circle**. Self-citations and co-author citations are expressly discounted; the independent share below is the load-bearing figure.

91.0% independent of 776 classified citing papers

Citation type	Count
Independent	706
Self-citation	14
Co-author	56
Same-institution	0

433 citing papers could not be classified (no author data) and are excluded from the percentages above.

C. Significant Contributions & Their Citation Evidence

Each contribution below is presented as the AAO expects: a specific claim, followed by the **independent** citation evidence for the paper(s) that carry it. Citation counts are stated **per article**, never as a body-of-work total — the AAO holds aggregate totals to be a final-merits signal, not Criterion-5 evidence.

Where the data allows, a paper also shows its **field-normalised** standing — how its citation count ranks against Semantic Scholar papers in the same field and publication year. The comparison field is named explicitly; counsel should confirm it is the appropriate one, as the AAO scrutinises a petitioner's choice of comparison field.

Contribution 1

Claim — Contribution 1

The researcher pioneered the integration of large language models with personalization, establishing a foundational framework that has significantly influenced independent research in adaptive AI systems.

The researcher's core contribution centers on the seminal 2024 paper 'Lamp: When large language models meet personalization,' which appears to establish a foundational approach for combining LLM capabilities with personalized user experiences. This work serves as the anchor for a broader research line exploring the nuances of human-AI interaction and control.

Originality in this line of work is suggested by the progression from the core theoretical framework to applied investigations in subsequent years. The 2025 follow-up papers indicate an expansion into specific domains, such as scientific search and human-AI collaboration behaviors, suggesting the researcher is actively refining how personalization balances with user control and real-world writing assistance.

The significance of this contribution is evidenced by the core paper's 502 citations, indicating strong adoption within the field. Notably, 92.4% of the scholar's total citing papers originate from independent researchers, demonstrating that this work has resonated beyond the immediate academic circle and influenced a broad, external community of scholars.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 348 · **50 flagged influential by Semantic Scholar**

CORE PAPER

Lamp: When large language models meet personalization

2024 · Proceedings of the 62nd Annual Meeting of the Association for Computational ..., 2024 · 502 citations (GS)

Field-normalised: 413 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2024 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	When large language models meet personalization: Perspectives of challenges and opportunities	Huawei Technologies, University of Electronic Science and Technology of China, University of Science and Technology of China	China	—
2	The benefits, risks and bounds of personalizing the alignment of large language models to individuals	Bocconi University, University of Oxford	Italy, United Kingdom	—
3	Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models	Alibaba Group, Beihang University, Beijing University of Posts and Telecommunications	China, Switzerland	Methodology
4	Retrieval-augmented generation with graphs (graphrag)	Adobe Research, Amazon, Meta	United States	—
5	From matching to generation: A survey on generative information retrieval	Renmin University of China, Tsinghua University, University of Montreal	Canada, China	—
6	Large language models as zero-shot conversational recommenders	Allen Institute of AI, Cornell University, Netflix Inc.	United States	Background
7	From persona to personalization: A survey on role-playing language agents	Fudan University, Shanghai University, System, Inc.	China, United States	—

No.	Citing paper	Citing institution(s)	Country	S2
8	Rewarded soups: towards pareto-optimal alignment by interpolating weights fine-tuned on diverse rewards	Sorbonne Université	France	—
9	Personalized soups: Personalized large language model alignment via post-hoc parameter merging	Allen Institute for AI, Carnegie Mellon University, University of Washington	United States	Background
10	Limitations of the llm-as-a-judge approach for evaluating llm outputs in expert knowledge tasks	Purdue University, University of Notre Dame	United States	—
11	From individual to society: A survey on social simulation driven by large language model-based agents	East China Normal University, Fudan University, Harbin Institute of Technology, Shenzhen	China	—
12	Democratizing large language models via personalized parameter-efficient fine-tuning	Amazon.com Inc., University of Notre Dame	United States	—
13	Once: Boosting content-based recommendation with both open-and closed-source large language models	The Hong Kong Polytechnic University, Waseda University	China, Japan	—
14	Can llm be a personalized judge?	University of Cambridge	United Kingdom	Background
15	Flask: Fine-grained language model evaluation based on alignment skill sets	Carnegie Mellon University, KAIST, Naver	France, South Korea, United States	Background
16	A survey on large language model-based game agents	Cisco, Georgia Institute of Technology	United States	—
17	Personalized language modeling from personalized human feedback	Carnegie Mellon University, University of Texas at Austin	United States	—
18	A survey of personalized large language models: Progress and future directions	Huawei, Huawei Technologies Co., Ltd, The Chinese University of Hong Kong	China	**Influential**
19	Multimodal pretraining, adaptation, and generation for recommendation: A survey	Huawei, Huazhong University of Science and Technology, The Hong Kong Polytechnic University	China	Background
20	Personalllm: Tailoring llms to individual preferences	Columbia University	United States	—
21	Personalized pieces: Efficient personalized large language models through collaborative efforts	University of Notre Dame	United States	**Methodology**
22	A survey of personalization: From rag to agent	City University of Hong Kong, Huawei	China, Hong Kong, Singapore	—
23	Llm discussion: Enhancing the creativity of large language models via discussion framework and role-play	National Taiwan University	Taiwan	**Methodology**
24	Personality alignment of large language models	University College London, Westlake University, Zhejiang University	China, United Kingdom	—

No.	Citing paper	Citing institution(s)	Country	S2
25	Teaching language models to evolve with users: Dynamic profile modeling for personalized alignment	Du Xiaoman Financial, Harbin Institute of Technology	China	—
26	How ai processing delays foster creativity: Exploring research question co-creation with an llm-based agent	University of Illinois Urbana-Champaign, University of Notre Dame	United States	Influential
27	Large language models empowered personalized web agents	Eastern Institute of Technology, National University of Singapore, The Hong Kong Polytechnic University	China, Singapore	Influential
28	Personagym: Evaluating persona agents and llms	Georgia Tech, Princeton University, University of Illinois Chicago	United States	—
29	Multimodal llms for health grounded in individual-specific data	Google Research	United States	Background
30	Measuring what makes you unique: Difference-aware user modeling for enhancing llm personalization	National University of Singapore, The Chinese University of Hong Kong, University of Science and Technology of China	China, Singapore	—

Showing the 30 most-cited of 330 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models

"Recent advances in the LLM community have showcased the potential of LLM customization and role-playing (Wei et al., 2023; Shanahan et al., 2023; Li et al., 2023a; Salemi et al., 2023; Maas et al., 2023; Li et al., 2023b; Chen et al., 2023a; Park et al., 2023)."

METHODOLOGY Personalized pieces: Efficient personalized large language models through collaborative efforts

"P CS with the non-personalized baseline, prompt-based methods (retrieval-augmented (Salemi et al., 2023) and profile-augmented personalization (Richard-son et al., 2023)), and PEFT-based personalization methods (PEFT retrieval (Zhao et al., 2024) and OPPU (Tan et al., 2024))."

METHODOLOGY Llm discussion: Enhancing the creativity of large language models via discussion framework and role-play

"To encourage LLMs to discuss with others and inspire others actively to engender collective creativity, we devise a three-phase discussion framework that explicitly requires each LLM to build upon others' responses."

FOLLOW-UP WORK

Bridging Personalization and Control in Scientific Personalized Search

2025 · Proceedings of the 48th International ACM SIGIR Conference on Research and ..., 2025 · 2 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	PARK: Personalized academic retrieval with knowledge-graphs	ISTI-CNR, University of Milan-Bicocca	Italy	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

FOLLOW-UP WORK

Prototypical Human-AI Collaboration Behaviors from LLM-Assisted Writing in the Wild

2025 · arXiv preprint arXiv:2505.16023, 2025 · 18 citations (GS)

Field-normalised: 12 Semantic Scholar citations place it in the top 10% of Computer Science papers from 2025 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Ai4research: A survey of artificial intelligence for scientific research	ByteDance, Central South University, Chinese University of Hong Kong	China, United States	—
2	The AI Memory Gap: Users Misremember What They Created With AI or Without	Aalto University, University of Bayreuth	Finland, Germany	—
3	Feedback by design: Understanding and overcoming user feedback barriers in conversational agents	Adobe Inc., Johns Hopkins University	United States	—
4	A Framework to Characterize Reporting on Generative AI Use	Microsoft Research, Princeton University, University of Washington	United States	—
5	Authorship Drift: How Self-Efficacy and Trust Evolve During LLM-Assisted Writing	KAIST	South Korea	—
6	Plotania: Exploring Transparency Trade-offs in AI Co-Writing Through Virtual Readers and Transparent Attribution	City University of Hong Kong, RMIT University, Tsinghua University	Australia, China	—
7	Show or Tell? Modeling the evolution of request-making in Human-LLM conversations	Cornell University	United States	—
8	Programming by Chat: A Large-Scale Behavioral Analysis of 11,579 Real-World AI-Assisted IDE Sessions	University of Notre Dame, Vanderbilt University	United States	—
9	Can You Make It Sound Like You? Post-Editing LLM-Generated Text for Personal Style	University of Maryland	United States	—
10	From Words to Widgets for Controllable LLM Generation	Allen Institute for AI, Cornell University, The University of Texas at Austin	United States	—
11	Co-Data: Cultivating Effective Human-LLM Collaboration for Collaborative Data Processing	Carnegie Mellon University Africa, Delft University of Technology, Google DeepMind	France, Netherlands, Rwanda	—
12	Priming, Path-dependence, and Plasticity: Understanding the molding of user-LLM interaction and its implications from (many) chat logs in the wild	Cornell University	United States	—
13	From Planning to Revision: How AI Writing Support at Different Stages Alters Ownership	Columbia University, Johns Hopkins University, University of Michigan	Australia, United States	—
14	RECAP: An End-to-End Platform for Capturing, Replaying, and Analyzing AI-Assisted Programming Interactions	Carnegie Mellon University	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
15	Tinker Tales: Supporting Child-AI Collaboration through Co-Creative Storytelling with Educational Scaffolding	Carnegie Mellon University, Emory University	United States	—
16	Componentization: Decomposing Monolithic LLM Responses into Manipulable Semantic Units	Honda Research Institute	United States	**Influential**
17	Can You Make It Sound Like You? Post-Editing LLM-Generated Text for Personal Style	University of Maryland	United States	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Contribution 2

Claim — Contribution 2

The researcher developed CSFCube, a test collection for faceted query by example, and extended this framework to evaluate instructed retrieval models and LLM-augmented narrative recommendations.

The researcher's contribution centers on advancing information retrieval evaluation through the creation of CSFCube, a test collection of computer science research articles designed for faceted query by example. This core work, published in 2021, established a foundational resource for assessing how users can explore complex information spaces using example-based queries. The titles indicate that this line of work addresses the challenge of supporting effective exploration in retrieval systems, moving beyond simple keyword matching to more nuanced, example-driven interaction models.

Originality in this trajectory is suggested by the progression from establishing a specific test collection to investigating the capabilities of newer technologies within that context. The follow-up work appears to examine whether instructed retrieval models can genuinely support exploration, questioning the efficacy of current approaches. Furthermore, the researcher extended these concepts to large language model-augmented narrative-driven recommendations, indicating a shift toward integrating generative AI into recommendation and retrieval workflows. This chronological development suggests a sustained effort to refine how users interact with and navigate large corpora of scientific literature.

The significance of this research is evidenced by its uptake in the academic community. The core paper has accumulated 40 citations, while the subsequent work on LLM-augmented recommendations has garnered 100 citations, indicating growing interest in these methods. Notably, analysis of the researcher's broader citation record reveals that 92.4% of citations come from independent researchers, suggesting that this line of work has influenced peers outside the researcher's immediate institution and collaboration network. This high degree of independent citation underscores the broader impact and relevance of the contributions to the field of information retrieval.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 81 · **4 flagged influential by Semantic Scholar**

CORE PAPER

CSFCube--A Test Collection of Computer Science Research Articles for Faceted Query by Example

2021 · arXiv preprint arXiv:2103.12906, 2021 · 40 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Taxonomy-guided semantic indexing for academic paper search	Korea University, Pohang University of Science and Technology, University of Illinois at Urbana Champaign	South Korea, United States	**Methodology**

No.	Citing paper	Citing institution(s)	Country	S2
2	Improving scientific document retrieval with concept coverage-based query set generation	Korea University, Pohang University of Science and Technology, University of Illinois at Urbana-Champaign	South Korea, United States	—
3	Retrieval for extremely long queries and documents with RPRS: a highly efficient and effective transformer-based re-ranker	Leiden University, University of Milano-Bicocca	Italy, Netherlands	—
4	Chain of Retrieval: Multi-Aspect Iterative Search Expansion and Post-Order Search Aggregation for Full Paper Retrieval	KAIST	South Korea	—
5	Multi-Facet Blending for Faceted Query-by-Example Retrieval	Pohang University of Science and Technology, POSTECH	South Korea	Influential
6	On the interpolation of contextualized term-based ranking with BM25 for query-by-example retrieval	Leiden University	Netherlands	Background
7	CASPER: Concept-integrated Sparse Representation for Scientific Retrieval	Aalto University, University of Illinois Urbana-Champaign	Finland, United States	—
8	Improving Scientific Document Retrieval with Academic Concept Index	Korea University, Pohang University of Science and Technology, University of Illinois at Urbana-Champaign	South Korea, United States	—
9	Large-scale evaluation of transformer-based article encoders on the task of citation recommendation	University of Zagreb	Croatia	Methodology
10	UniFAR: A Unified Facet-Aware Retrieval Framework for Scientific Documents	Beihang University, Shandong University	China	—
11	Scientific paper retrieval with llm-guided semantic-based ranking	Korea University, University of Illinois Urbana-Champaign	South Korea, United States	—
12	Improving BERT-based query-by-document retrieval with multi-task optimization	Leiden University, University of Strathclyde	Netherlands, United Kingdom	Background
13	CoRank: LLM-based compact reranking with document features for scientific retrieval	Korea University, University of Illinois at Urbana-Champaign, University of Illinois Urbana-Champaign	South Korea, United States	—
14	Measuring Risk of Bias in Biomedical Reports: The RoBBR Benchmark	UC San Diego, University of California, San Diego, University of Southern California	United States	—
15	Hierarchical transformer-based query by multiple documents	Amazon Inc., University of Massachusetts Amherst	United States	Background
16	Paperregister: Boosting flexible-grained paper search via hierarchical register indexing	Chinese Academy of Sciences, University of Chinese Academy of Sciences	China	—
17	Aspect-Aware Content-Based Recommendations for Mathematical Research Papers	FIZ Karlsruhe, National Institute of Informatics, University of Göttingen	Germany, Japan	—

No.	Citing paper	Citing institution(s)	Country	S2
18	Abstracts Embeddings Evaluation: A Case Study of Artificial Intelligence and Medical Imaging for the COVID-19 Infection	University of Bologna	Italy	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY Taxonomy-guided semantic indexing for academic paper search

"We select two recently published datasets: CSFCube (Mysore et al., 2021) and DORIS-MAE (Wang et al., 2023)."

METHODOLOGY Large-scale evaluation of transformer-based article encoders on the task of citation recommendation

"...such representations, recent work has proposed various transformer-based article encoders (TAEs), i.e., LMs that are finetuned using citation or co-citation information as a training signal, such as SPECTER (Cohan et al., 2020), ASPIRE (Mysore et al., 2021a), and SciNCL (Ostendorff et al., 2022)."

FOLLOW-UP WORK

Can Instructed Retrieval Models Really Support Exploration?

2026 · Proceedings of the 2026 Conference on Human Information Interaction and ..., 2026 · 0 citations (GS)

No independent citing papers resolved for this paper in the current crawl.

FOLLOW-UP WORK

Large language model augmented narrative driven recommendations

2023 · Proceedings of the 17th ACM Conference on Recommender Systems, 777-783, 2023 · 100 citations (GS)

Field-normalised: 76 Semantic Scholar citations place it in the top 5% of Computer Science papers from 2023 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	User modeling in the era of large language models: Current research and future directions	University of Notre Dame	United States	Methodology
2	A survey on personalized and pluralistic preference alignment in large language models	Adobe Research, Allen Institute of AI, The Ohio State University	United States	—
3	Recommender systems in the era of large language models (llms)	Michigan State University, National University of Defense Technology, The Hong Kong Polytechnic University	Australia, China, Hong Kong	—
4	A survey of GPT-3 family large language models including ChatGPT and GPT-4	Akmumus AI	India	—
5	A survey on large language models for recommendation	University of Science and Technology of China	China	Methodology
6	A survey on knowledge distillation of large language models	Tencent Hunyuan, The University of Hong Kong, The University of Sydney	Australia, Hong Kong, Singapore	Background
7	Towards open-world recommendation with knowledge augmentation from large language models	Huawei, Shanghai Jiao Tong University, Tencent	China	—

No.	Citing paper	Citing institution(s)	Country	S2
8	A review of modern recommender systems using generative models (gen-recsys)	Amazon, Polytechnic University of Bari, University of California	Canada, Germany, Italy	Influential
9	How can recommender systems benefit from large language models: A survey	Huawei, Shanghai Jiao Tong University, Tencent	China	—
10	Large language models are competitive near cold-start recommenders for language-and item-based preferences	Google, University of Toronto	Canada, Norway, United States	Background
11	Let me do it for you: Towards llm empowered recommendation via tool learning	University of Amsterdam, University of Science and Technology of China, Upwork	China, Netherlands, United States	Methodology
12	Cold-start recommendation towards the era of large language models (llms): A comprehensive survey and roadmap	Hong Kong University of Science and Technology (Guangzhou), Jinan University, Mohamed bin Zayed University of Artificial Intelligence	China, Hong Kong, United Arab Emirates	—
13	Towards next-generation llm-based recommender systems: A survey and beyond	Jilin University, Macquarie University, Meta AI	Australia, China, United States	—
14	Enhancing recommendation diversity by re-ranking with large language models	University College Cork	Ireland	—
15	Llmcdsr: Enhancing cross-domain sequential recommendation with large language models	The Hong Kong University of Science and Technology, The Hong Kong University of Science and Technology (Guangzhou), University of Science and Technology of China	China	—
16	Notellm: A retrievable large language model for note recommendation	University of Science and Technology of China, Xiaohongshu Inc.	China	Methodology
17	Negative sampling in recommendation: A survey and future directions	Nanjing University of Science and Technology, National University of Singapore, Shandong University	China, Singapore	—
18	Large language models enhanced collaborative filtering	Kuaishou Technology Co., Ltd., Renmin University of China	China	—
19	Rella: Retrieval-enhanced large language models for lifelong sequential behavior comprehension in recommendation	Huawei, Shanghai Jiao Tong University	China	Background
20	Openp5: An open-source platform for developing, training, and evaluating llm-based recommender systems	Rutgers University	United States	Methodology
21	Exploring the impact of large language models on recommender systems: An extensive review	Carnegie Mellon University, Santa Clara University, Stanford University	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
22	Generative news recommendation	Leiden University, Renmin University of China, Shandong University	China, Netherlands	Background
23	Stealthy attack on large language model based recommendation	Chinese Academy of Sciences, Institute of Automation, Chinese Academy of Sciences, Northeastern University	China, United States	Background
24	Tapping the potential of large language models as recommender systems: A comprehensive framework and empirical analysis	Meituan Group, Renmin University of China	China	—
25	Sinkt: A structure-aware inductive knowledge tracing model with large language model	Huawei, Shanghai Jiao Tong University	China	—
26	Transparent and scrutable recommendations using natural language user profiles	The University of Sheffield, University College London	United Kingdom	Background
27	Reindex-then-adapt: Improving large language models for conversational recommendation	Cornell University, Netflix Inc., UC San Diego	United States	—
28	Lifelong personalized low-rank adaptation of large language models for recommendation	Huawei, Shanghai Jiao Tong University	China	Background
29	Neighborhood-based collaborative filtering for conversational recommendation	Cornell University, Netflix Inc., University of California San Diego	United States	—
30	Cora: Collaborative information perception by large language model's weights for recommendation	Chinese Academy of Sciences, Northeastern University	China, United States	Background

Showing the 30 most-cited of 63 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY User modeling in the era of large language models: Current research and future directions

"Mysore et al. [154] augment narrative-driven recommendation using LLMs for author narrative query generation based on user-item interactions and train retrieval models with these LLM-augmented queries."

METHODOLOGY A survey on large language models for recommendation

"Meanwhile, to summarize the user's intention by prompt based on their interaction data, MINT [Mysore et al. , 2023] employed Instruct-GPT, a 175B parameter LLM, to generate a synthetic narrative query."

METHODOLOGY Let me do it for you: Towards llm empowered recommendation via tool learning

"• LLMs enhance RSs: Here, RSs are enhanced with world knowledge and reasoning abilities of LLMs [25, 27, 32, 41, 50, 59]."

METHODOLOGY Notellm: A retrievable large language model for note recommendation

"The first method is utilizing LLMs to augment data [21, 29, 51]."

METHODOLOGY Openp5: An open-source platform for developing, training, and evaluating llm-based recommender systems

"LLM can be used for feature engineering, which takes the original data as input and generates rich textual features as data augmentations [32, 47]."

Contribution 3

Claim — Contribution 3

The researcher pioneered automated extraction of action graphs from materials synthesis texts, establishing a foundational framework for semantic annotation in materials science procedural literature.

The researcher's core contribution rests on the 2017 paper 'Automatically extracting action graphs from materials science synthesis procedures,' which introduced a method for structuring unstructured synthesis data. This work appears to address the challenge of converting narrative experimental procedures into machine-readable formats, a critical gap in computational materials science.

Originality is suggested by the chronological progression from this core method to subsequent works. The 2019 follow-up, 'The materials science procedural text corpus,' indicates the development of annotated resources to support such extraction, while the 2021 paper 'MS-Mentions' suggests further refinement in entity recognition within these texts. This trajectory implies a systematic effort to build robust semantic structures for procedural analysis.

Significance is evidenced by strong independent uptake. The core paper has 53 citations, while the 2019 corpus paper has 158 citations, indicating growing reliance on these resources. With 92.4% of citations from independent researchers, the work demonstrates broad adoption beyond the researcher's immediate circle, validating its utility in the wider scientific community.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 120 · **13 flagged influential by Semantic Scholar**

CORE PAPER

Automatically extracting action graphs from materials science synthesis procedures

2017 · arXiv preprint arXiv:1711.06872, 2017 · 53 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Opportunities and challenges for machine learning in materials science	University of Wisconsin-Madison	United States	—
2	Automated extraction of chemical synthesis actions from experimental procedures	IBM Research Europe	Switzerland	—
3	Device fabrication knowledge extraction from materials science literature	Tata Consultancy Services Ltd	India	—
4	The SOFC-exp corpus and neural approaches to information extraction in the materials science domain	Bosch Center for Artificial Intelligence, Robert Bosch GmbH, University of Augsburg	Germany, United States	Result
5	Unleashing the power of knowledge extraction from scientific literature in catalysis	University of Delaware	United States	—
6	Annotating and extracting synthesis process of all-solid-state batteries from scientific literature	Panasonic Corporation, National Institute of Advanced Industrial Science and Technology, Toyota Technological Institute, Toyota Technological Institute, National Institute of Advanced Industrial Science and Technology	Japan, United States	—
7	Pcmsp: A dataset for scientific action graphs extraction from polycrystalline materials synthesis procedure text	University of California Santa Barbara, University of California, Santa Barbara	United States	Methodology

No.	Citing paper	Citing institution(s)	Country	S2
8	MatScIE: An automated tool for the generation of databases of methods and parameters used in the computational materials science literature	Indian Institute of Technology, Indian Institute of Technology Kharagpur	India	—
9	Made of steel? learning plausible materials for components in the vehicle repair domain	University of Stuttgart	Germany	Methodology
10	OSPAR: A corpus for extraction of organic synthesis procedures with argument roles	Hokkaido University	Japan	—
11	MuLMS: A multi-layer annotated text corpus for information extraction in the materials science domain	Bosch Center for Artificial Intelligence, Robert Bosch GmbH, University of Augsburg	Germany, United States	Methodology
12	[illegible]	[illegible]	[illegible]	—
13	Relation/Entity-Centric Reading Comprehension	Toyota Technological Institute at Chicago	United States	Methodology
14	Combinatorial synthesis for AI-driven materials discovery	California Institute of Technology	United States	—
15	Named entity recognition and normalization applied to large-scale information extraction from the materials science literature	—	—	—
16	Semi-supervised machine-learning classification of materials synthesis procedures	University of California, Berkeley	United States	—
17	Pipelines for procedural information extraction from scientific literature: towards recipes using machine learning and data science	Johns Hopkins University, Kansas State University, Lawrence Livermore National Laboratory	United States	—
18	Building open knowledge graph for metal-organic frameworks (mof-kg): Challenges and case studies	Colorado School of Mines, Drexel University, University of Central Florida	United States	Background
19	On the limits of learning to actively learn semantic representations	Hebrew University of Jerusalem, Tel-Aviv University, University of Utah	Israel, United States	Background
20	Text to insight: Accelerating organic materials knowledge extraction via deep learning	Drexel University, Kebotix, Inc., Northeastern University	United States	—
21	Information extraction and graph representation for the design of formulated products	TCS Research	—	Background

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology* / *Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

RESULT The SOFC-exp corpus and neural approaches to information extraction in the materials science domain

"Mysore et al. (2017) apply the generative model of Kiddon et al. (2015) to induce action graphs for synthesis procedures of materials from text."

METHODOLOGY Pcmsp: A dataset for scientific action graphs extraction from polycrystalline materials synthesis procedure text

"Previous research (Mysore et al., 2017, 2019) either annotates the whole synthesis paragraph in the general inorganic domain, ignoring the non-synthesis sentences and subdomain discrepancy or only focuses on entity mentions (Friedrich et al., 2020; O'Gorman et al., 2021)."

METHODOLOGY Made of steel? learning plausible materials for components in the vehicle repair domain

"...(Rani and Kumar, 2021) and LDA (Venugopal et al., 2021), as well as supervised methods for domain-specific NER, often focusing on a single material or material group (Mysore et al., 2017, 2019; Friedrich et al., 2020; Gupta et al., 2022; Nayak and Timmapathini, 2021; O'Gorman et al., 2021)."

METHODOLOGY MuLMS: A multi-layer annotated text corpus for information extraction in the materials science domain

"Mysore et al. (2017) exper-iment with unsupervised methods for extracting action graphs for synthesis procedures."

METHODOLOGY Relation/Entity-Centric Reading Comprehension

"[92] apply the generative model of Kiddon et al."

FOLLOW-UP WORK

The materials science procedural text corpus: Annotating materials synthesis procedures with shallow semantic structures

2019 · Proceedings of the 13th linguistic annotation workshop, 56-64, 2019 · 158 citations (GS)

Field-normalised: 105 Semantic Scholar citations place it in the top 5% of Materials Science papers from 2019 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Enabling large language models for real-world materials discovery	Intel	United States	—
2	Opportunities and challenges for machine learning in materials science	University of Wisconsin-Madison	United States	—
3	Language models for materials discovery and sustainability: Progress, challenges, and opportunities	New York University, Oak Ridge National Laboratory	United States	—
4	Automated extraction of chemical synthesis actions from experimental procedures	IBM Research Europe	Switzerland	—
5	MatSci-NLP: Evaluating scientific language models on materials science language tasks using text-to-schema modeling	Intel, University of Montreal	Canada, United States	**Methodology**
6	Analyzing research trends in inorganic materials literature using NLP	Panasonic Corporation, Toyota Technological Institute	Japan, United States	Background
7	Data augmentation techniques for process extraction from scientific publications	Fujitsu Limited	Japan	—
8	[illegible]	[illegible]	Japan	—
9	MatSciBERT: A materials domain language model for text mining and information extraction	Indian Institute of Technology Delhi, Intel	India, United States	—
10	Language models and protocol standardization guidelines for accelerating synthesis planning in heterogeneous catalysis	ETH Zurich, IBM Research Europe	Switzerland	—
11	A general-purpose material property data extraction pipeline from large polymer corpora using natural language processing	Georgia Institute of Technology, Indian Institute of Technology, Indore	India, United States	—
12	An analysis of simple data augmentation for named entity recognition	CSIRO	Australia	**Methodology**
13	Lifelong pretraining: Continually adapting language models to emerging corpora	AWS AI Labs, MIT, University of Southern California	Australia, United States	—

No.	Citing paper	Citing institution(s)	Country	S2
14	A family of large language models for materials research with insights into model adaptability in continued pretraining	Indian Institute of Technology Delhi	India	—
15	Foundational large language models for materials research	Cerebras Systems, Inc., Indian Institute of Technology Delhi, Intel	India, United States	**Influential**
16	A survey on cutting-edge relation extraction techniques based on language models	University of Granada	Spain	—
17	The SOFC-exp corpus and neural approaches to information extraction in the materials science domain	Bosch Center for Artificial Intelligence, Robert Bosch GmbH, University of Augsburg	Germany, United States	**Methodology**
18	Large language models for heterogeneous catalysis	National University of Singapore	Singapore	—
19	Annotated textual dataset PV600 of perovskite bandgaps for information extraction from literature	University of Turku	Finland	—
20	Few-shot named entity recognition: definition, taxonomy and research directions	University of Naples Federico II	Italy	**Methodology**
21	Schema: State changes matter for procedure planning in instructional videos	Columbia University, The Hong Kong University of Science and Technology	China, United States	**Methodology**
22	Causal reasoning of entities and events in procedural texts	Allen Institute for AI, Carnegie Mellon University, King's College London	United Kingdom, United States	—
23	SLM-MATRIX: a multi-agent trajectory reasoning and verification framework for enhancing language models in materials data extraction	Peking University	China	—
24	ORKG-Leaderboards: a systematic workflow for mining leaderboards as a knowledge graph	Leibniz University of Hannover, TIB, Leibniz Information Centre for Science and Technology	Germany	Background
25	Knowledge graph for solubility big data: construction and applications	Gannan Normal University	China	—
26	DiSCoMaT: distantly supervised composition extraction from tables in materials science articles	Indian Institute of Technology Delhi	India	Background
27	A dataset for tracking entities in open domain procedural text	Allen Institute for AI, Allen Institute for Artificial Intelligence, Carnegie Mellon University	United States	**Methodology**
28	Causal discovery from data assisted by large language models	Naval Research Laboratory, Oak Ridge National Laboratory, University of Maryland, College Park	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
29	Versatile Deep Learning Pipeline for Transferable Chemical Data Extraction	Cornell University	United States	—
30	Unleashing the power of knowledge extraction from scientific literature in catalysis	University of Delaware	United States	—

Showing the 30 most-cited of 84 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY MatSci-NLP: Evaluating scientific language models on materials science language tasks using text-to-schema modeling

"MatSci-NLP contains NER task data adapted from Weston et al. (2019); Friedrich et al. (2020); Mysore et al. (2019); Yamaguchi et al. (2020)."

METHODOLOGY An analysis of simple data augmentation for named entity recognition

"Second, applying all data augmentation methods together outperforms any single data augmentation on average, although, when the complete training set is used, applying single data augmentation may achieve better results (c.f., MaSciP-Recurrent and i2b2-2010-Transformer)."

METHODOLOGY The SOFC-exp corpus and neural approaches to information extraction in the materials science domain

"The work closest to ours is the one of Mysore et al. (2019) who annotate a corpus of 230 para-graphs describing synthesis procedures with operations and their arguments, e.g., "The resulting [solid products Material] were ... .""

METHODOLOGY Few-shot named entity recognition: definition, taxonomy and research directions

"MaSciP [100] 2019 Synthesis procedures annotated with synthesis operations and their typed arguments (e."

METHODOLOGY Schema: State changes matter for procedure planning in instructional videos

"5 generated descriptions using our chain-of-thought prompting. where only the action step annotations a 1: T are available (Figure 1 (Bollini et al., 2013; Yang & Nyberg, 2015; Mysore et al., 2019), we represent steps as their before-states and after-states."

FOLLOW-UP WORK

MS-Mentions: Consistently Annotating Entity Mentions in Materials Science Procedural Text

2021 · Proceedings of the 2021 Conference on Empirical Methods in Natural Language ..., 2021 · 24 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Alloy synthesis and processing by semi-supervised text mining	University of Science and Technology Beijing	China	—
2	Pcmsp: A dataset for scientific action graphs extraction from polycrystalline materials synthesis procedure text	University of California Santa Barbara, University of California, Santa Barbara	United States	Methodology
3	Made of steel? learning plausible materials for components in the vehicle repair domain	University of Stuttgart	Germany	Methodology
4	SynKB: Semantic search for synthetic procedures	Georgia Institute of Technology, Johns Hopkins University, SRI International	United States	—
5	OSPAR: A corpus for extraction of organic synthesis procedures with argument roles	Hokkaido University	Japan	—
6	Automatic knowledge acquisition from superconductivity information in literature	Nagoya University, Toyota Technological Institute	Japan, United States	—
7	MuLMS: A multi-layer annotated text corpus for information extraction in the materials science domain	Bosch Center for Artificial Intelligence, Robert Bosch GmbH, University of Augsburg	Germany, United States	Influential

No.	Citing paper	Citing institution(s)	Country	S2
8	MuLMS-AZ: an argumentative zoning dataset for the materials science domain	Bosch Center for Artificial Intelligence, LMU Munich, Robert Bosch GmbH	Germany, United States	—
9	Rapid adaptation of chemical named entity recognition using few-shot learning and llm distillation	University of Delaware	United States	—
10	Advancing sentiment analysis for low-resourced african languages using pre-trained language models	University of Johannesburg, University of the Witwatersrand	South Africa	**Influential**
11	POLYIE: A dataset of information extraction from polymer material scientific literature	Georgia Institute of Technology	United States	**Methodology**
12	Quokka: An open-source large language model chatbot for material science	University of California, Santa Barbara	United States	**Methodology**
13	Seed-guided fine-grained entity typing in science and engineering domains	IBM Almaden Research Center, IBM Thomas J. Watson Research Center, University of Illinois	United States	**Methodology**
14	Towards Fully-Automated Materials Discovery via Large-Scale Synthesis Dataset and Expert-Level LLM-as-a-Judge	Ajou University, Hankuk University of Foreign Studies, Hanyang University	South Korea	—
15	From the Rock Floor to the Cloud: A Systematic Survey of State-of-the-Art NLP in Battery Life Cycle	Luleå University of Technology, Lund University	Sweden	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology* / *Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY Pcmsp: A dataset for scientific action graphs extraction from polycrystalline materials synthesis procedure text

"Previous research (Mysore et al., 2017, 2019) either annotates the whole synthesis paragraph in the general inorganic domain, ignoring the non-synthesis sentences and subdomain discrepancy or only focuses on entity mentions (Friedrich et al., 2020; O'Gorman et al., 2021)."

METHODOLOGY Made of steel? learning plausible materials for components in the vehicle repair domain

", 2021), as well as supervised methods for domain-specific NER, often focusing on a single material or material group (Mysore et al., 2017, 2019; Friedrich et al., 2020; Gupta et al., 2022; Nayak and Timmapathini, 2021; O'Gorman et al., 2021)."

METHODOLOGY POLYIE: A dataset of information extraction from polymer material scientific literature

"More recent datasets are created manually for solid oxide fuel cells (Friedrich et al., 2020) and material science synthesis procedures (O'Gorman et al., 2021)."

METHODOLOGY Quokka: An open-source large language model chatbot for material science

"NLP techniques have been widely used for various materials science tasks, ranging from material action graph extraction [Yang et al., 2022, Friedrich et al., 2020, O'Gorman et al., 2021], intelligent knowledge search [Yang et al., 2023b] and instruction following [Song et al., 2023b]."

METHODOLOGY Seed-guided fine-grained entity typing in science and engineering domains

"Although we focus on software and security domain examples, our entity typing framework can be applied to other specialized domains including science (Wang et al. 2021) and engineering (O'Gorman et al. 2021)."

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
University of Massachusetts Amherst	United States	SCImago #788 · QS =247	48
Carnegie Mellon University	United States	SCImago #266 · THE 24 · QS 52	38
University of Washington	United States	SCImago #45 · THE 25 · QS 81	26
Massachusetts Institute of Technology	United States	SCImago #41 · THE 2 · QS 1	22
National University of Singapore	Singapore	SCImago #59 · THE 17 · QS 8	22
University of Science and Technology of China	China	SCImago #77 · THE 51 · QS =132	21
Georgia Institute of Technology	United States	SCImago #270 · THE =41 · QS =123	19
University of Illinois Urbana-Champaign	United States	QS =70	19
Alibaba Group	China	SCImago #226	18
Fudan University	China	SCImago #46 · THE 36 · QS 30	16
Seoul National University	South Korea	SCImago #135 · THE =58 · QS =38	15
Huawei	Singapore	—	15
Cornell University	United States	SCImago #61 · THE =18 · QS 16	15
Allen Institute for AI	United States	—	15
University of Michigan	United States	SCImago #43 · THE 23 · QS 45	15

Geographic distribution of citing authors

Country	Citing papers
United States	413
China	194
South Korea	54
United Kingdom	47
Germany	39
Singapore	32
Japan	28
India	26
Canada	25
Italy	23
Netherlands	22
Australia	15

Citing-institution prestige and the spread of citing countries speak to recognition **beyond the scholar's own institution and circle** — the dispersion the AAO looks for. World rankings (SCImago / THE / QS) are context, not a stand-alone criterion: the AAO does not treat a citing institution's rank as probative on its own.

F. AAO Precedent Considerations

Pre-filing self-check (AAO denial patterns)

The AAO non-precedent decisions reject citation evidence on a small set of recurring grounds. Confirm the petition addresses each before filing:

- ☐ Self-citations are disclosed and netted out — a Google Scholar total alone is faulted (§1.1).
- ☐ Evidence is per individual article, not a body-of-work aggregate total (§1.2).
- ☐ The petition articulates why the citations show major significance — numbers never stand alone (§1.5).
- ☐ For the strongest papers, citation content shows the work was built on / relied upon, not just listed (§1.6, §2.2).
- ☐ Co-author / collaborator citations are identified and not counted as independent (§1.7).
- ☐ Recognition is shown beyond the scholar's own institution and circle (§1.8).
- ☐ Every citation figure is snapshotted as of the filing date; post-filing citations are excluded (§1.9).
- ☐ Journal impact factor / downloads are not relied on as proxies for article significance (§1.10, §1.12).
- ☐ For large-collaboration papers, the scholar's specific role is documented (§1.13).
- ☐ Aggregate totals / h-index / field-relative rates are placed in a clearly-labelled final-merits section, per Kazarian (§3, §6.1.7).

Disclaimer

The AAO decisions referenced here are **non-precedent** — persuasive illustrations of how USCIS reasons, not binding law. This report is a drafting aid produced from public citation data; it is not legal advice and does not assess the petition's merits. All analysis must be reviewed by qualified immigration counsel.

G. Citation Evidence Index

Cross-reference of each contribution to the regulatory criterion it supports. Counsel should map these to the petition's exhibit numbers.

Contribution	Core paper	Indep. cites	Supports
Contribution 1	Lamp: When large language models meet personalization	348	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	CSFCube--A Test Collection of Computer Science Research Articles for Faceted Query by Example	81	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 3	Automatically extracting action graphs from materials science synthesis procedures	120	8 CFR 204.5(h)(3)(v) — Criterion 5