Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

NAN DU

OpenAI

Google Scholar profile

Generated 2026-06-08 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

2,375	2,621	30	52
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.

94.4% independent of 2,307 classified citing papers

Citation type	Count
Independent	2,177
Self-citation	6
Co-author	124
Same-institution	0

68 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 advanced language model capabilities by establishing foundational fine-tuning methods and extending them into reasoning, acting, and large-scale architecture designs.

The researcher's contribution centers on a seminal 2021 paper on fine-tuned language models as zero-shot learners, which serves as the foundation for subsequent work. This line of research appears to address the challenge of enhancing model versatility and scale, evolving from initial fine-tuning strategies to broader architectural scaling and synergistic reasoning frameworks.

The progression from the core paper to follow-up works on scaling pathways and reasoning-acting synergy suggests a deliberate effort to overcome limitations in generalization and complex task execution. By building upon the initial zero-shot learning framework, the researcher appears to have expanded the functional scope of language models, moving from basic adaptation to sophisticated, multi-modal operational capabilities.

The significance of this work is evidenced by substantial citation counts, with the core paper and follow-ups accumulating thousands of citations each. Notably, 94.4% of classified citations originate from independent researchers, indicating that this line of inquiry has been widely adopted and validated by the broader scientific community beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 2,398 · **264 flagged influential by Semantic Scholar**

CORE PAPER

Finetuned language models are zero-shot learners

2021 · 5,933 citations (GS)

Field-normalised: 4,941 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2021 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	The rise and potential of large language model based agents: A survey	Fudan University, Meituan	China	—
2	A survey of large language models	Alibaba Group, Renmin University of China, Université de Montréal	Canada, China, United States	Influential
3	Large language models are zero-shot time series forecasters	Carnegie Mellon University, New York University	United States	—
4	A comprehensive survey of small language models in the era of large language models: Techniques, enhancements, applications, collaboration with llms, and ...	Amazon, Penn State University, The Pennsylvania State University	United States	—
5	Mm-llms: Recent advances in multimodal large language models	Kyoto University, Mohamed bin Zayed University of Artificial Intelligence, Tencent	China, Japan, United Arab Emirates	—
6	Unified-io 2: Scaling autoregressive multimodal models with vision language audio and action	Allen Institute for AI, Allen Institute for Artificial Intelligence, University of Illinois at Urbana-Champaign	United States	—
7	Meshlrm: Large reconstruction model for high-quality meshes	Adobe Research, University of California San Diego	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
8	Hyenadna: Long-range genomic sequence modeling at single nucleotide resolution	Harvard University, Mila and Université de Montréal, Stanford University	Canada, United States	—
9	Empowering biomedical discovery with AI agents	Broad Institute, Harvard Medical School, Harvard University	United Kingdom, United States	—
10	Advances and challenges in foundation agents: From brain-inspired intelligence to evolutionary, collaborative, and safe systems	Nanyang Technological University, The Hong Kong University of Science and Technology (Guangzhou), University of California, Irvine Medical Center	Canada, China, Singapore	—
11	The dawn of lmms: Preliminary explorations with gpt-4v (ision)	Microsoft, Microsoft Corporation, National University of Singapore	Singapore, United States	—
12	An empirical study of catastrophic forgetting in large language models during continual fine-tuning	Tencent (China), Tencent Inc, The Chinese University of Hong Kong	China	—
13	Continual learning of large language models: A comprehensive survey	Google, Rutgers University	United States	—
14	Foundational challenges in assuring alignment and safety of large language models	—	—	—
15	A survey on large language models for code generation	NAVER Cloud, The Hong Kong University of Science and Technology, The Hong Kong University of Science and Technology (Guangzhou)	China, South Korea	—
16	Large language models: A survey	Cologne University of Applied Sciences, Czech Technical University in Prague, Expedia Group (United States)	Czech Republic, Germany, Singapore	—
17	Unifying large language models and knowledge graphs: A roadmap	Beijing University of Technology, Griffith University, Hefei University of Technology	Australia, China, Singapore	—
18	Camel: Communicative agents for" mind" exploration of large language model society	King Abdullah University of Science and Technology (KAUST)	Saudi Arabia	—
19	The prompt report: A systematic survey of prompt engineering techniques	University of Maryland	United States	—
20	Glm-130b: An open bilingual pre-trained model	Tsinghua University, Zhipu AI, Zhipu.AI	China	**Influential**
21	A survey on LLM-based multi-agent systems: workflow, infrastructure, and challenges	University of Technology Sydney, Wuhan University	Australia, China	—
22	A survey on multimodal large language models	Nanjing University, Skywork AI, Tencent	China	**Influential**
23	The llama 3 herd of models	—	—	—

No.	Citing paper	Citing institution(s)	Country	S2
24	Safety alignment should be made more than just a few tokens deep	Google DeepMind, Princeton University, Tsinghua University	China, United Kingdom, United States	Influential
25	Weak-to-strong generalization: Eliciting strong capabilities with weak supervision	Anthropic, EleutherAI, OpenAI	United States	—
26	Editing models with task arithmetic	Allen Institute for AI, Microsoft Research, University of Washington	United States	—
27	: A Vision-Language-Action Flow Model for General Robot Control	—	—	—
28	Deep time series models: A comprehensive survey and benchmark	Tsinghua University	China	—
29	Function vectors in large language models	Northeastern University	United States	—
30	Large language models meet text-centric multimodal sentiment analysis: A survey	Chinese Academy of Sciences, Harbin Institute of Technology	China	—

Showing the 30 most-cited of 893 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).

FOLLOW-UP WORK

Palm: Scaling language modeling with pathways

2023 · 9,424 citations (GS)

Field-normalised: 7,900 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2023 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Risk taxonomy, mitigation, and assessment benchmarks of large language model systems	Ant Group, Institute of Information Engineering, Chinese Academy of Sciences, Tsinghua University	China	—
2	Tree of thoughts: Deliberate problem solving with large language models	Google DeepMind, Princeton University, University of California, Irvine Medical Center	United Kingdom, United States	—
3	A survey of large language models	Alibaba Group, Renmin University of China, Université de Montréal	Canada, China, United States	Influential
4	Mamba: Linear-time sequence modeling with selective state spaces	Princeton University	United States	—
5	A comprehensive survey of small language models in the era of large language models: Techniques, enhancements, applications, collaboration with llms, and ...	Amazon, Penn State University, The Pennsylvania State University	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
6	Mm-llms: Recent advances in multimodal large language models	Kyoto University, Mohamed bin Zayed University of Artificial Intelligence, Tencent	China, Japan, United Arab Emirates	—
7	Voxposer: Composable 3d value maps for robotic manipulation with language models	MIT, Stanford University, University of Pennsylvania	United States	—
8	Visual autoregressive modeling: Scalable image generation via next-scale prediction	ByteDance, ByteDance (China), ByteDance Inc	China, United States	—
9	MobileNetV4: Universal models for the mobile ecosystem	Google, National University of Singapore	Singapore, United States	—
10	Nucleotide transformer: building and evaluating robust foundation models for human genomics	InstaDeep	United Kingdom	—
11	Sequential modeling enables scalable learning for large vision models	Indian Institute of Technology Kanpur, Johns Hopkins University, UC Berkeley	India, Israel, United States	—
12	Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness	Massachusetts Institute of Technology	United States	—
13	Vision-language models for medical report generation and visual question answering: A review	H. Lee Moffitt Cancer Center and Research Institute	United States	—
14	AI generates covertly racist decisions about people based on their dialect	Allen Institute for Artificial Intelligence, Stanford University	United States	—
15	The dawn of lmms: Preliminary explorations with gpt-4v (ision)	Microsoft, Microsoft Corporation, National University of Singapore	Singapore, United States	—
16	Inductive biases for deep learning of higher-level cognition	Université de Montréal, University of Montreal	Canada	—
17	Continual learning of large language models: A comprehensive survey	Google, Rutgers University	United States	—
18	Advances and challenges in meta-learning: A technical review	Eindhoven University of Technology, Halmstad University, University of South Dakota	Netherlands, Sweden, United States	—
19	A survey on large language models for code generation	NAVER Cloud, The Hong Kong University of Science and Technology, The Hong Kong University of Science and Technology (Guangzhou)	China, South Korea	—
20	Large language models: A survey	Cologne University of Applied Sciences, Czech Technical University in Prague, Expedia Group (United States)	Czech Republic, Germany, Singapore	—
21	Autonomous chemical research with large language models	Carnegie Mellon University	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
22	A review on large language models: Architectures, applications, taxonomies, open issues and challenges	Bangladesh University of Engineering and Technology, Charles Darwin University, Lappeenranta-Lahti University of Technology	Australia, Bangladesh, Finland	—
23	Unleashing the potential of prompt engineering for large language models	Beijing Normal University, BNU-HKBU United International College	China	—
24	Camel: Communicative agents for" mind" exploration of large language model society	King Abdullah University of Science and Technology (KAUST)	Saudi Arabia	—
25	A survey on deep neural network pruning: Taxonomy, comparison, analysis, and recommendations	Harbin Institute of Technology (Shenzhen), The University of Adelaide, University of Adelaide	Australia, China	—
26	A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions	Harbin Institute of Technology, Huawei Inc., Zhejiang University	China	—
27	Glm-130b: An open bilingual pre-trained model	Tsinghua University, Zhipu AI, Zhipu.AI	China	**Influential**
28	Deja vu: Contextual sparsity for efficient llms at inference time	Adobe Research, Carnegie Mellon University, ETH Zurich	China, Switzerland, United States	—
29	A survey on LLM-based multi-agent systems: workflow, infrastructure, and challenges	University of Technology Sydney, Wuhan University	Australia, China	—
30	The llama 3 herd of models	—	—	—

Showing the 30 most-cited of 752 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).

FOLLOW-UP WORK

React: Synergizing reasoning and acting in language models

2022 · 10,393 citations (GS)

Field-normalised: 7,235 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2022 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Risk taxonomy, mitigation, and assessment benchmarks of large language model systems	Ant Group, Institute of Information Engineering, Chinese Academy of Sciences, Tsinghua University	China	—
2	The rise and potential of large language model based agents: A survey	Fudan University, Meituan	China	—

No.	Citing paper	Citing institution(s)	Country	S2
3	Tree of thoughts: Deliberate problem solving with large language models	Google DeepMind, Princeton University, University of California, Irvine Medical Center	United Kingdom, United States	—
4	A survey of large language models	Alibaba Group, Renmin University of China, Université de Montréal	Canada, China, United States	—
5	Meddxagent: A unified modular agent framework for explainable automatic differential diagnosis	NEC Laboratories Europe, University of California, Irvine Medical Center	Germany, United States	—
6	The dawn of lmms: Preliminary explorations with gpt-4v (ision)	Microsoft, Microsoft Corporation, National University of Singapore	Singapore, United States	—
7	A survey on large language models for code generation	NAVER Cloud, The Hong Kong University of Science and Technology, The Hong Kong University of Science and Technology (Guangzhou)	China, South Korea	—
8	Large language models: A survey	Cologne University of Applied Sciences, Czech Technical University in Prague, Expedia Group (United States)	Czech Republic, Germany, Singapore	—
9	Unleashing the potential of prompt engineering for large language models	Beijing Normal University, BNU-HKBU United International College	China	—
10	Retrieval-augmented generation for ai-generated content: A survey	Peking University, Shanghai Jiao Tong University	China	—
11	Advancing transformer architecture in long-context large language models: A comprehensive survey	Baidu (China), Nanjing University, State Key Laboratory of New Technology of Computer Software	China	—
12	A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions	Harbin Institute of Technology, Huawei Inc., Zhejiang University	China	—
13	Toward edge general intelligence with agentic AI and agentification: Concepts, technologies, and future directions	Auburn University, Guangdong University of Technology, Institute for Infocomm Research	Australia, Canada, China	—
14	A survey on agentic multimodal large language models	Chinese University of Hong Kong, City University of Hong Kong, Communication University of China	China, Singapore	—
15	Function calling in large language models: Industrial practices, challenges, and future directions	Ant Group, City University of Hong Kong	China, Hong Kong	**Influential**
16	Beyond self-talk: A communication-centric survey of llm-based multi-agent systems	Beihang University, Beijing University of Posts and Telecommunications	China	—

No.	Citing paper	Citing institution(s)	Country	S2
17	From llm reasoning to autonomous ai agents: A comprehensive review	Khalifa University, Technology Innovation Institute, United Arab Emirates University	United Arab Emirates	—
18	Natural-language agent harnesses	Harbin Institute of Technology (Shenzhen), Tsinghua University	China	—
19	Skill-SD: Skill-Conditioned Self-Distillation for Multi-turn LLM Agents	The Chinese University of Hong Kong, University of Chinese Academy of Sciences, University of Science and Technology of China	China, Hong Kong	—
20	Exploring large language model based intelligent agents: Definitions, methods, and prospects	Peking University, Tencent (China), The Chinese University of Hong Kong, Shenzhen	China	—
21	Core: Benchmarking LLMs' code reasoning capabilities through static analysis tasks	Purdue University	United States	—
22	A Survey on Vision–Language–Action Models for Embodied AI	Chinese University of Hong Kong, Huawei, The Chinese University of Hong Kong	China	—
23	Deep learning in single-cell and spatial transcriptomics data analysis: advances and challenges from a data science perspective	Anhui University of Science and Technology, Peng Cheng Laboratory, Tsinghua University	China, United States	—
24	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	Influential
25	From matching to generation: A survey on generative information retrieval	Renmin University of China, Tsinghua University, University of Montreal	Canada, China	—
26	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	China	—
27	A survey on large language model-based game agents	Cisco, Georgia Institute of Technology	United States	Influential
28	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	—
29	Social skill training with large language models	Georgia Institute of Technology, Georgia Tech, Stanford University	United States	—
30	Trism for agentic ai: A review of trust, risk, and security management in llm-based agentic multi-agent systems	Cornell University, University of Groningen, Vector Institute	Canada, Netherlands, United States	—

Showing the 30 most-cited of 753 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).

Contribution 2

Claim — Contribution 2

The researcher established foundational methods for community detection in large-scale social networks, a seminal contribution evidenced by high independent citation rates.

The researcher's core contribution rests on the 2007 paper 'Community detection in large-scale social networks.' This work appears to address the challenge of identifying structural groups within massive network datasets, a problem that was becoming increasingly critical as social data grew in volume and complexity during that period.

The titles suggest this line of work provided a scalable approach to analyzing network topology. By focusing on 'large-scale' systems, the research likely bridged the gap between theoretical community detection models and practical applications on real-world social graphs, offering a framework that could handle the computational demands of extensive datasets.

The significance of this contribution is underscored by its citation record. With 344 citations, the paper is well-cited within its field. Notably, 94.4% of the citing papers originate from independent researchers, indicating that the work has been widely adopted and validated by the broader scientific community rather than relying on self-citation or institutional bias.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 0

CORE PAPER

Community detection in large-scale social networks

2007 · 344 citations (GS)

Field-normalised: 248 Semantic Scholar citations place it in the top 5% of Computer Science papers from 2007 indexed by Semantic Scholar, by citation count.

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

Contribution 3

Claim — Contribution 3

The researcher advanced collaborative filtering by integrating tagging behaviors into recommendation systems, establishing a foundational approach widely adopted by independent scholars.

The researcher's core contribution rests on the 2008 paper 'Improved recommendation based on collaborative tagging behaviors.' This work appears to represent a significant step in refining recommendation algorithms by leveraging user-generated tags alongside traditional collaborative filtering methods. The title suggests a focus on enhancing system accuracy or relevance through this hybrid behavioral data.

This line of work addresses the limitations of early collaborative filtering systems, which often relied solely on explicit ratings or item similarities. By incorporating tagging behaviors, the researcher likely provided a novel mechanism to capture implicit user preferences and semantic context. The absence of follow-up papers by the same author indicates that this single publication served as a definitive, self-contained contribution to the field rather than the start of a prolonged personal research series.

The significance of this contribution is evidenced by its substantial citation count of 138. More importantly, analysis of the broader citation landscape reveals that 94.4% of citations to the researcher's work originate from independent researchers. This high degree of independent uptake suggests that the method proposed in the 2008 paper has been widely recognized, validated, and utilized by the broader academic community as a reliable and influential advancement in recommendation technology.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 0

CORE PAPER

Improved recommendation based on collaborative tagging behaviors

2008 · 138 citations (GS)

Field-normalised: 105 Semantic Scholar citations place it in the top 5% of Computer Science papers from 2008 indexed by Semantic Scholar, by citation count.

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

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Tsinghua University	China	SCImago #8 · THE 12 · QS =17	175
University of California, Irvine Medical Center	United States	—	143
Stanford University	United States	SCImago #18 · THE =5 · QS 3	138
Carnegie Mellon University	United States	SCImago #266 · THE 24 · QS 52	129
Google Research	United States	—	97
Peking University	China	SCImago #11 · THE 13 · QS 14	96
Microsoft	United States	—	82
Google DeepMind	United States	SCImago #90	82
National University of Singapore	Singapore	SCImago #59 · THE 17 · QS 8	80
Microsoft Research	United States	—	79
Nanyang Technological University	Singapore	SCImago #137	78
Google	United States	—	75
University of Washington	United States	SCImago #45 · THE 25 · QS 81	68
Princeton University	United States	SCImago #386 · THE =3 · QS =25	68
Zhejiang University	China	SCImago #6 · THE 39 · QS 49	67

Geographic distribution of citing authors

Country	Citing papers
United States	1,328
China	947
United Kingdom	241
Singapore	181
Canada	141
Hong Kong	139
Australia	108
Germany	85
South Korea	59
Switzerland	45

Country	Citing papers
France	44
India	41

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	Finetuned language models are zero-shot learners	2,398	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	Community detection in large-scale social networks	0	8 CFR 204.5(h)(3)(v) — Criterion 5

Contribution	Core paper	Indep. cites	Supports
Contribution 3	Improved recommendation based on collaborative tagging behaviors	0	8 CFR 204.5(h)(3)(v) — Criterion 5