Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

Kai Yuanqing Xiao

PhD Student, MIT CSAIL

[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

661	661	16	11
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.

98.6% independent of 644 classified citing papers

Citation type	Count
Independent	635
Self-citation	0
Co-author	9
Same-institution	0

17 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 produced a seminal technical report on GPT-4, establishing a foundational reference for large language model capabilities that has garnered nearly 28,000 citations.

CLAIM: The researcher's primary contribution is the publication of the "Gpt-4 technical report" in 2023, which serves as the core document for this line of work. This single publication stands alone without follow-up papers by the same author in the provided dataset, indicating its self-contained significance as a definitive technical overview.

ORIGINALITY: The title suggests the work addresses the need for comprehensive documentation of advanced generative pre-trained transformer capabilities. By releasing a detailed technical report, the researcher appears to have filled a critical gap in transparently communicating the architecture, performance, and limitations of state-of-the-art language models to the broader scientific community.

SIGNIFICANCE: The work has achieved extraordinary impact, evidenced by 27,979 citations. Analysis of 644 citing papers reveals that 98.6% originate from independent researchers, demonstrating that the contribution has been widely adopted and validated by the global academic community rather than relying on self-citation or institutional bias.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 635 **· 70 flagged influential by Semantic Scholar**

CORE PAPER

Gpt-4 technical report

2023 · 27,979 citations (GS)

Field-normalised: 23,913 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	The rise and potential of large language model based agents: A survey	Fudan University, Meituan	China	—
2	Diffusion models: A comprehensive survey of methods and applications	Beijing University of Posts and Telecommunications, Carnegie Mellon University, Peking University	China, United States	—
3	Visual autoregressive modeling: Scalable image generation via next-scale prediction	ByteDance, ByteDance (China), ByteDance Inc	China, United States	Influential
4	Instruction tuning for large language models: A survey	Alibaba Group, Nanyang Technological University, Peking University	China, Singapore, United States	—
5	Visual cot: Advancing multi-modal language models with a comprehensive dataset and benchmark for chain-of-thought reasoning	SenseTime, Sensetime (China), SenseTime Research	Canada, China, Hong Kong	—
6	Structured 3d latents for scalable and versatile 3d generation	Microsoft Research, Tsinghua University, University of Science and Technology of China	China	—
7	How to build the virtual cell with artificial intelligence: Priorities and opportunities	Broad Institute, Chan Zuckerberg Initiative (United States), Enzo Life Sciences	Canada, France, Germany	—

No.	Citing paper	Citing institution(s)	Country	S2
8	The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update	AARNet (Australia), AstraZeneca, Center for Advanced Studies Research and Development in Sardinia	Australia, Belgium, Czech Republic	—
9	An agentic system for rare disease diagnosis with traceable reasoning	Harvard Medical School, Shanghai Artificial Intelligence Laboratory, Shanghai Jiao Tong University	China, United States	—
10	Measuring implicit bias in explicitly unbiased large language models	Cornell Tech	United States	—
11	Roles, users, benefits, and limitations of chatbots in health care: rapid review	McGill University, McGill University Health Centre, Polytechnique Montréal	Canada	—
12	Sapiens: Foundation for human vision models	Meta	United States	—
13	Artificial intelligence for science in quantum, atomistic, and continuum systems	—	—	—
14	Advances and challenges in meta-learning: A technical review	Eindhoven University of Technology, Halmstad University, University of South Dakota	Netherlands, Sweden, United States	—
15	Unleashing the potential of prompt engineering for large language models	Beijing Normal University, BNU-HKBU United International College	China	—
16	A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions	Harbin Institute of Technology, Huawei Inc., Zhejiang University	China	—
17	A survey on multimodal large language models	Nanjing University, Skywork AI, Tencent	China	**Influential**
18	A comprehensive overview of large language models	Australian National University, Commonwealth Scientific and Industrial Research Organisation, King Fahd University of Petroleum and Minerals	Australia, China, Pakistan	—
19	Model merging in llms, mllms, and beyond: Methods, theories, applications, and opportunities	Nanyang Technological University, Northeastern University, Sun Yat-sen University	China, Singapore, United States	—
20	: A Vision-Language-Action Flow Model for General Robot Control	—	—	—
21	Deep time series models: A comprehensive survey and benchmark	Tsinghua University	China	—
22	Understanding transformer reasoning capabilities via graph algorithms	Columbia University, Google, Google DeepMind (United Kingdom)	Canada, United Kingdom, United States	—

No.	Citing paper	Citing institution(s)	Country	S2
23	A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges	Seoul National University	South Korea	—
24	Flipattack: Jailbreak llms via flipping	National University of Singapore	Singapore	Influential
25	From llm reasoning to autonomous ai agents: A comprehensive review	Khalifa University, Technology Innovation Institute, United Arab Emirates University	United Arab Emirates	—
26	Dynamic-llava: Efficient multimodal large language models via dynamic vision-language context sparsification	East China Normal University, Nanjing University, Xiamen University	China	—
27	Wavchat: A survey of spoken dialogue models	Alibaba Group, Microsoft, Tencent	China, United States	—
28	Deep learning for cross-domain data fusion in urban computing: Taxonomy, advances, and outlook	Carnegie Mellon University, JD Technology, Southwest Jiaotong University	China, Hong Kong, United States	—
29	Large language model inference acceleration: A comprehensive hardware perspective	Infinigence-AI, Shanghai Jiao Tong University, Tsinghua University	China	—
30	Can large language models transform computational social science?	Dartmouth College, Georgia Institute of Technology, Georgia Tech	United States	—

Showing the 30 most-cited of 635 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 a rigorous framework for evaluating neural network robustness using mixed integer programming, creating a foundational benchmark for formal verification in deep learning.

CLAIM: The researcher's seminal 2017 paper, 'Evaluating robustness of neural networks with mixed integer programming,' represents a core contribution to the field of secure machine learning. This work appears to have introduced a method for formally assessing the reliability of neural networks against adversarial perturbations.

ORIGINALITY: By applying mixed integer programming to neural network evaluation, this line of work addresses the critical gap in verifying model robustness. The titles suggest a shift from heuristic testing to mathematically rigorous verification, offering a novel approach to quantifying security vulnerabilities in deep learning systems.

SIGNIFICANCE: With 1,335 citations, the paper is highly influential. Notably, 98.6% of citing papers originate from independent researchers, indicating broad adoption across the global academic community. This widespread independent uptake underscores the work's status as a standard reference in robustness evaluation.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 0

CORE PAPER

[Evaluating robustness of neural networks with mixed integer programming]

2017 · 1,335 citations (GS)

Field-normalised: 940 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2017 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 developed a training method to induce ReLU stability, enabling faster adversarial robustness verification in neural networks.

The researcher's core contribution rests on the 2018 paper 'Training for faster adversarial robustness verification via inducing relu stability.' This work appears to address the computational bottleneck of verifying neural network robustness by introducing a training strategy that induces stability in ReLU activations. By focusing on this specific architectural property, the research suggests a novel approach to making verification processes more efficient without compromising security guarantees. The significance of this line of work is evidenced by its substantial uptake in the field, with 263 citations recorded for the core paper. Notably, 98.6% of the citing papers originate from independent researchers, indicating that the methodology has been widely adopted and validated by the broader scientific community rather than just the researcher's immediate circle. This high degree of independent engagement underscores the work's foundational role in advancing efficient adversarial robustness verification.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 0

CORE PAPER

[Training for faster adversarial robustness verification via inducing relu stability]

2018 · 263 citations (GS)

Field-normalised: 211 Semantic Scholar citations place it in the top 5% of Computer Science papers from 2018 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	67
Peking University	China	SCImago #11 · THE 13 · QS 14	34
University of California, Irvine Medical Center	United States	—	32
Shanghai Jiao Tong University	China	SCImago #10 · THE 40 · QS =47	29
Shanghai Artificial Intelligence Laboratory	China	SCImago #563	27
The Chinese University of Hong Kong	Hong Kong	SCImago #163 · THE =41 · QS =32	27
Carnegie Mellon University	United States	SCImago #266 · THE 24 · QS 52	27
Stanford University	United States	SCImago #18 · THE =5 · QS 3	27

Institution	Country	World ranking	Citing papers
Shanghai AI Laboratory	China	—	26
Nanyang Technological University	Singapore	SCImago #137	24
University of Science and Technology of China	China	SCImago #77 · THE 51 · QS =132	22
The University of Hong Kong	Hong Kong	SCImago #195 · THE 33 · QS 11	21
National University of Singapore	Singapore	SCImago #59 · THE 17 · QS 8	20
Zhejiang University	China	SCImago #6 · THE 39 · QS 49	18
University of Washington	United States	SCImago #45 · THE 25 · QS 81	18

Geographic distribution of citing authors

Country	Citing papers
China	326
United States	312
United Kingdom	62
Hong Kong	56
Canada	46
Singapore	43
Australia	30
Germany	26
South Korea	21
United Arab Emirates	15
France	14
Japan	12

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	Gpt-4 technical report	635	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	Evaluating robustness of neural networks with mixed integer programming	0	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 3	Training for faster adversarial robustness verification via inducing relu stability	0	8 CFR 204.5(h)(3)(v) — Criterion 5