Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

Xuechen Li

Unknown affiliation

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

29	29	5	21
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.

100.0% independent of 29 classified citing papers

Citation type	Count
Independent	29
Self-citation	0
Co-author	0
Same-institution	0

0 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 provided a seminal framework for evaluating the dual implications of foundation models, establishing a critical reference point for assessing their societal impact and technical potential.

CLAIM: The researcher's primary contribution is the publication of "On the opportunities and risks of foundation models" (2021), which serves as a foundational text for understanding the broader implications of this technology class. This work stands alone as the core contribution in this specific line of inquiry, without direct follow-up papers by the same author listed in the provided data.

ORIGINALITY: The title suggests the work addresses a critical gap by simultaneously examining both the potential benefits and the inherent dangers of foundation models. By framing the discussion around "opportunities and risks," the researcher appears to have introduced a balanced, comprehensive perspective that was likely novel at the time of publication, moving beyond purely technical assessments to include broader societal considerations.

SIGNIFICANCE: The work has achieved substantial recognition, evidenced by 9,571 citations. Notably, analysis of 29 citing papers reveals that 100% are from independent researchers, indicating that the contribution has resonated widely across the global academic community rather than being confined to the researcher's immediate circle. This high level of independent uptake underscores the paper's status as a widely accepted reference point in the field.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 10

CORE PAPER

On the opportunities and risks of foundation models

2021 · 9,571 citations (GS)

Field-normalised: 6,284 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	Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data (2024)	The Chinese University of Hong Kong, The University of Hong Kong, TikTok	Hong Kong	Background
2	A Survey on Evaluation of Large Language Models (2024)	Carnegie Mellon University, Hong Kong University of Science and Technology, Institute of Automation, Chinese Academy of Sciences	China, Hong Kong, United States	Background
3	A Survey on Large Language Models for Code Generation (2026)	NAVER Cloud, The Hong Kong University of Science and Technology, The Hong Kong University of Science and Technology (Guangzhou)	China, South Korea	—
4	The Rise and Potential of Large Language Model Based Agents: A Survey (2025)	Alibaba Group, ByteDance, Fudan University	China	—
5	Scientific discovery in the age of artificial intelligence (2023)	BioMap, Boehringer Ingelheim, Broad Institute of MIT and Harvard	Canada, China, Germany	—
6	AI models collapse when trained on recursively generated data (2024)	Imperial College London, University of Cambridge, University of Edinburgh	Canada, United Kingdom	—

No.	Citing paper	Citing institution(s)	Country	S2
7	Towards a general-purpose foundation model for computational pathology (2024)	Brigham and Women's Hospital, Brigham and Women's Hospital, Harvard Medical School, Brigham and Women’s Hospital, Harvard Medical School	United States	—
8	MathVista: Evaluating Mathematical Reasoning of Foundation Models in Visual Contexts (2024)	Microsoft, Stanford University, University of California, Los Angeles	United States	Background
9	A foundation model for clinical-grade computational pathology and rare cancers detection (2024)	Memorial Sloan Kettering Cancer Center, Microsoft Research, NSW Health Pathology, St George Hospital	Australia, United States	—
10	Qwen Technical Report (2023)	—	—	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).

Contribution 2

Claim — Contribution 2

The researcher advanced variational autoencoder theory by isolating specific sources of disentanglement, a foundational contribution evidenced by nearly 1,900 citations from independent scholars.

The researcher’s core contribution centers on the 2018 paper 'Isolating sources of disentanglement in variational autoencoders.' This work stands as a seminal piece in the field, establishing a critical framework for understanding how latent variables in generative models can be effectively separated. The titles indicate a focus on the theoretical underpinnings of disentanglement, addressing a fundamental challenge in interpreting and controlling complex neural network representations.

This line of work appears to address the difficulty of ensuring that learned latent factors correspond to meaningful, independent data characteristics. By isolating these sources, the researcher provided a methodological advance that clarifies the mechanisms behind disentanglement. The absence of follow-up papers by the same researcher suggests this single publication serves as a definitive, standalone theoretical contribution rather than part of an extended iterative series.

The significance of this work is underscored by its substantial citation count of 1,895, indicating widespread adoption and influence within the academic community. Notably, 100% of the classified citing papers originate from independent researchers, demonstrating that the contribution has resonated beyond the researcher’s immediate circle and has become a standard reference for independent scholars exploring variational autoencoders and disentanglement.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 4

CORE PAPER

Isolating sources of disentanglement in variational autoencoders

2018 · 1,895 citations (GS)

Field-normalised: 999 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2018 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Interpretable machine learning: Fundamental principles and 10 grand challenges (2022)	Duke University	United States	—
2	Algorithmic fairness in artificial intelligence for medicine and healthcare (2023)	Boston University, Brigham and Women's Hospital, Harvard Medical School, Broad Institute of Harvard and Massachusetts Institute of Technology	United States	—
3	A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions (2024)	American International University-Bangladesh, International Islamic University Chittagong, Mälardalen University	Bangladesh, Sweden, United States	—
4	Learning deep representations by mutual information estimation and maximization (2019)	MILA, MRN, MSR	Canada	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).

Contribution 3

Claim — Contribution 3

The researcher established a holistic evaluation framework for language models, providing a comprehensive benchmark that has become a standard reference in the field.

The researcher's primary contribution is the development of a holistic evaluation framework for language models, as detailed in the 2022 paper titled 'Holistic evaluation of language models.' This work serves as the foundational piece for this line of inquiry, standing alone without direct follow-up publications by the same author in the provided dataset.

This line of work appears to address the need for comprehensive assessment methodologies in natural language processing. By focusing on 'holistic' evaluation, the research suggests a move beyond narrow metrics to capture broader model capabilities. The absence of immediate follow-up papers by the researcher indicates that the core framework was sufficiently robust to stand as a definitive contribution in its own right.

The significance of this work is evidenced by its substantial citation count of 2,688, indicating widespread adoption and influence within the academic community. Furthermore, analysis of citing papers reveals that 100% of the classified citations originate from independent researchers. This high degree of independent uptake underscores the framework's utility and acceptance across diverse institutions, confirming its status as a seminal contribution to the field.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

Holistic evaluation of language models

2022 · 2,688 citations (GS)

Field-normalised: 1,527 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	MMLU-Pro: A More Robust and Challenging Multi-Task Language Understanding Benchmark (2024)	Carnegie Mellon University, University of Toronto, University of Waterloo	Canada, United States	Background
2	xLSTM: Extended Long Short-Term Memory (2024)	Johannes Kepler University Linz, NXAI	Austria	—
3	Solving olympiad geometry without human demonstrations (2024)	Google DeepMind, New York University	United States	—
4	A Survey on Large Language Model (LLM) Security and Privacy: The Good, the Bad, and the Ugly (2024)	Drexel University	United States	—
5	Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference (2024)	—	—	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).

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Microsoft Research	United States	—	4
University of Cambridge	United Kingdom	SCImago #63 · THE =3 · QS 6	3
Massachusetts Institute of Technology	United States	SCImago #41 · THE 2 · QS 1	3
New York University	United States	SCImago #116 · THE =31 · QS 55	3
The Chinese University of Hong Kong	Hong Kong	SCImago #163 · THE =41 · QS =32	3
Stanford University	United States	SCImago #18 · THE =5 · QS 3	3
University of Toronto	Canada	SCImago #39 · THE 21 · QS 29	3
The University of Hong Kong	Hong Kong	SCImago #195 · THE 33 · QS 11	2
Google DeepMind	United States	SCImago #90	2
Peking University	China	SCImago #11 · THE 13 · QS 14	2
Broad Institute of MIT and Harvard	United States	SCImago #112	2
Harvard Medical School	United States	SCImago #12	2
Imperial College London	United Kingdom	SCImago #69 · THE 8 · QS 2	2
Carnegie Mellon University	United States	SCImago #266 · THE 24 · QS 52	2
University of Oxford	United Kingdom	SCImago #26 · THE 1 · QS 4	2

Geographic distribution of citing authors

Country	Citing papers
United States	14
China	8
United Kingdom	4

Country	Citing papers
Canada	4
Hong Kong	3
Australia	2
Germany	2
Singapore	2
Pakistan	1
Saudi Arabia	1
Austria	1
South Korea	1

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.

E. Citation Growth Over Time

Distinct citing papers by publication year. Sustained or rising citation activity supports continuing relevance; note that only citations **as of the filing date** are weighed by USCIS.

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	On the opportunities and risks of foundation models	10	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	Isolating sources of disentanglement in variational autoencoders	4	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 3	Holistic evaluation of language models	5	8 CFR 204.5(h)(3)(v) — Criterion 5