Citation Evidence Report

EB-2 NIW Petition — National Interest Waiver

Matter of Dhanasar · Prong 2 (well-positioned)

Ziyu Liu

Shanghai Jiao Tong University

[Google Scholar profile]

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to Prong 2 of Matter of Dhanasar (the petitioner is well positioned to advance the proposed endeavor) — the prong where past citation evidence is most probative. 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

146	177	16	8
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.

73.4% independent of 124 classified citing papers

Citation type	Count
Independent	91
Self-citation	7
Co-author	20
Same-institution	6

22 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 large vision-language model alignment by introducing multi-image augmented direct preference optimization and subsequent visual reinforcement fine-tuning methods.

The researcher established a foundational contribution in aligning large vision-language models through the core paper 'MIA-DPO: Multi-Image Augmented Direct Preference Optimization For Large Vision-Language Models,' published at ICLR 2025. This work appears to address the challenge of optimizing model preferences using augmented multi-image inputs, providing a novel framework for direct preference optimization in this domain.

Building on this foundation, the researcher extended the methodology with follow-up work titled 'Visual-rft: Visual reinforcement fine-tuning' in 2025. The chronological progression from direct preference optimization to reinforcement fine-tuning suggests a comprehensive approach to refining visual-language capabilities, indicating that the initial framework served as a critical stepping stone for more advanced training techniques.

The significance of this line of work is evidenced by substantial citation activity. The follow-up papers have accumulated over 450 citations each, while the core paper has received 42 citations. Notably, 79.0% of the citing papers originate from independent researchers, demonstrating that the broader academic community has adopted and built upon these methods beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 27 **· 2 flagged influential by Semantic Scholar**

CORE PAPER

MIA-DPO: Multi-Image Augmented Direct Preference Optimization For Large Vision-Language Models

2024 · ICLR 2025 · 42 citations (GS)

Field-normalised: 30 Semantic Scholar citations place it in the top 10% of Computer Science papers from 2024 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	MM-RLHF: The Next Step Forward in Multimodal LLM Alignment (2025)	—	—	—
2	Mitigating Hallucination in Multimodal Large Language Model via Hallucination-targeted Direct Preference Optimization (2025)	Renmin University of China, Tencent	China	—
3	Aligning Multimodal LLM with Human Preference: A Survey (2025)	Nanyang Technological University, National University of Singapore, Shenzhen International Graduate School, Tsinghua University	China, Singapore	—
4	Probing Visual Language Priors in VLMs (2024)	LG AI Research, University of Michigan	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).

FOLLOW-UP WORK

Visual-rft: Visual reinforcement fine-tuning

2025 · 471 citations (GS)

> Field-normalised: 430 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2025 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	From System 1 to System 2: A Survey of Reasoning Large Language Models	AiShiWeiLai AI Research, Chinese Academy of Sciences, City University of Hong Kong and the Hong Kong University of Science and Technology (Guangzhou)	China, United Arab Emirates, United Kingdom	—
2	Towards Reasoning Era: A Survey of Long Chain-of-Thought for Reasoning Large Language Models	Central South University, Fudan University, The University of Hong Kong	China	—
3	Remote Sensing Spatiotemporal Vision–Language Models: A Comprehensive Survey	Inner Mongolia University	China	—
4	A Survey of State of the Art Large Vision Language Models: Alignment, Benchmark, Evaluations and Challenges	University of Maryland	—	—
5	DeepEyes: Incentivizing "Thinking with Images" via Reinforcement Learning	—	—	—
6	Vlm-r1: A stable and generalizable r1-style large vision-language model	Carnegie Mellon University, Zhejiang University	China, United States	—
7	Videorft: Incentivizing video reasoning capability in mllms via reinforced fine-tuning	Beijing Institute of Technology, Shenzhen University	China	—
8	Thinkact: Vision-language-action reasoning via reinforced visual latent planning	National Taiwan University, NVIDIA	Taiwan, United States	—
9	More thinking, less seeing? assessing amplified hallucination in multimodal reasoning models	Stanford University, UC Santa Cruz, UC Santa Cruz, UC Santa Barbara	United States	—
10	Q-insight: Understanding image quality via visual reinforcement learning	Bytedance Inc., Peking University	China, United States	—
11	Ui-r1: Enhancing efficient action prediction of gui agents by reinforcement learning	The Chinese University of Hong Kong, VIVO	China, Hong Kong	**Influential**
12	Time-r1: Post-training large vision language model for temporal video grounding	Independent Researcher, Renmin University of China	China, United States	—
13	Vrag-rl: Empower vision-perception-based rag for visually rich information understanding via iterative reasoning with reinforcement learning	Alibaba Group	China	—
14	Perception-r1: Pioneering perception policy with reinforcement learning	Beijing University of Posts and Telecommunications, Huazhong University of Science and Technology, Johns Hopkins University	China, United States	**Influential**
15	Noisyrollout: Reinforcing visual reasoning with data augmentation	National University of Singapore, Sea AI Lab	Singapore	—
16	Delving into rl for image generation with cot: A study on dpo vs. grpo	Peking University, The Chinese University of Hong Kong	China, Hong Kong	—

No.	Citing paper	Citing institution(s)	Country	S2
17	Beyond chemical qa: Evaluating llm's chemical reasoning with modular chemical operations	International Digital Economy Academy, Peking University, Yale University	China, 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).

FOLLOW-UP WORK

Visual-RFT: Visual Reinforcement Fine-Tuning

2025 · arXiv.org · 459 citations (GS)

Field-normalised: 430 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2025 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	From System 1 to System 2: A Survey of Reasoning Large Language Models (2025)	AiShiWeiLai AI Research, Chinese Academy of Sciences, City University of Hong Kong and the Hong Kong University of Science and Technology (Guangzhou)	China, United Arab Emirates, United Kingdom	—
2	Towards Reasoning Era: A Survey of Long Chain-of-Thought for Reasoning Large Language Models (2025)	Central South University, Fudan University, The University of Hong Kong	China	—
3	Remote Sensing Spatiotemporal Vision–Language Models: A Comprehensive Survey (2025)	Inner Mongolia University	China	—
4	A Survey of State of the Art Large Vision Language Models: Alignment, Benchmark, Evaluations and Challenges (2025)	University of Maryland	—	—
5	DeepEyes: Incentivizing "Thinking with Images" via Reinforcement Learning (2026)	—	—	—
6	VisuLogic: A Benchmark for Evaluating Visual Reasoning in Multi-modal Large Language Models (2025)	SenseTime Research, University of Science and Technology of China, Xi'an Jiaotong University	China	—

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 benchmark for evaluating long-context document understanding capabilities, specifically addressing the critical challenge of integrating visualizations within extended textual contexts.

The researcher's primary contribution centers on the development of MMLONGBENCH-DOC, a seminal work published in Neural Information Processing Systems in 2024. This paper introduces a specialized framework for benchmarking how artificial intelligence systems comprehend documents that combine extensive text with complex visual elements. By focusing on the intersection of long-context processing and visual understanding, the work defines a new standard for evaluating multimodal document analysis.

This line of work appears to address a significant gap in the evaluation of large language models, which often struggle with the coherence and integration of visual data within long documents. The title suggests a novel approach to measuring performance not just on text length, but on the nuanced understanding of visualizations embedded within that text. As a standalone core contribution without immediate follow-up papers listed, it represents a foundational effort to standardize testing in this emerging subfield of multimodal AI.

The significance of this contribution is evidenced by its rapid uptake in the academic community, with 158 citations recorded shortly after publication. Notably, 79.0% of the citing papers originate from independent researchers, indicating that the benchmark has been widely adopted by the broader scientific community rather than just the researcher's immediate circle. This high degree of independent citation suggests that MMLONGBENCH-DOC has become a recognized and essential tool for validating advancements in long-context document understanding.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 3

CORE PAPER

MMLONGBENCH-DOC: Benchmarking Long-context Document Understanding with Visualizations

2024 · Neural Information Processing Systems · 158 citations (GS)

Field-normalised: 122 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	M3DocRAG: Multi-modal Retrieval is What You Need for Multi-page Multi-document Understanding (2024)	UNC Chapel Hill	United States	—
2	ViDoRAG: Visual Document Retrieval-Augmented Generation via Dynamic Iterative Reasoning Agents (2025)	University of Science and Technology of China	China	—
3	A Survey of Multimodal Retrieval-Augmented Generation (2025)	Huawei Cloud BU	—	—

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 developed Rar, a framework for retrieving and ranking augmented multimodal large language models to enhance visual recognition capabilities.

The researcher's core contribution is the development of Rar, a system for retrieving and ranking augmented multimodal large language models for visual recognition, as detailed in their 2024 publication. This work appears to address the challenge of effectively leveraging augmented MLLMs within visual recognition tasks by introducing a structured approach to retrieval and ranking.

The originality of this line of work lies in its specific focus on augmenting MLLMs for visual recognition, suggesting a novel methodological gap in how these models are accessed and prioritized for such applications. The titles indicate a targeted solution to optimizing model utility in this domain.

The significance of this contribution is evidenced by its rapid uptake, with 29 citations recorded since 2024. Notably, 79.0% of the citing papers originate from independent researchers, indicating that the broader academic community recognizes the value and applicability of this framework beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 11

CORE PAPER

Rar: Retrieving and ranking augmented mllms for visual recognition

2024 · 29 citations (GS)

Field-normalised: 23 Semantic Scholar citations place it in the top 10% of Computer Science papers from 2024 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Retrieval augmented generation and understanding in vision: A survey and new outlook	HKUST(GZ), HKUST(GZ), Sichuan University, HKUST(GZ), Tinghua University	Bulgaria, China, Italy	—
2	Unified generative and discriminative training for multi-modal large language models	Nanyang Technological University, National University of Singapore, Singapore Management University	China, Singapore	Background
3	When machine unlearning meets retrieval-augmented generation (rag): Keep secret or forget knowledge?	City University of Macau, University of Technology Sydney	Australia, China	—
4	Deepmmsearch-r1: Empowering multimodal llms in multimodal web search	Apple, Johns Hopkins University	United States	—
5	Can multimodal large language models be guided to improve industrial anomaly detection?	University of California Irvine, University of Connecticut	United States	—
6	A multi-expert framework for enhancing multimodal large language models in industrial anomaly detection	University of Connecticut	United States	—
7	RAG-Adapter: A Plug-and-Play RAG-enhanced Framework for Long Video Understanding	Central China Normal University, Central South University of Forestry and Technology, Hunan University	China	—
8	Mitigating behavioral hallucination in multimodal large language models for sequential images	King Abdullah University of Science and Technology, University of Copenhagen, University of Electronic Science and Technology of China	China, Denmark, Saudi Arabia	—
9	M2ConceptBase: A Fine-Grained Aligned Concept-Centric Multimodal Knowledge Base	Fudan University, Soochow University, Zhejiang Lab	China	—
10	Efficient Vocabulary-Free Fine-Grained Visual Recognition in the Age of Multimodal LLMs	Helmholtz Center for Information Security, Indian Institute of Technology Hyderabad	Germany, India	—

No.	Citing paper	Citing institution(s)	Country	S2
11	Group-Relative Visual Discrimination Enhancement for Unlocking Intrinsic Capability of MLLMs	Peng Cheng Laboratory, Shandong Institute of Automation	China	—

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
The Chinese University of Hong Kong	Hong Kong	SCImago #163 · THE =41 · QS =32	15
Shanghai AI Laboratory	China	—	15
Peking University	China	SCImago #11 · THE 13 · QS 14	15
University of Science and Technology of China	China	SCImago #77 · THE 51 · QS =132	12
Shanghai Innovation Institute	China	—	12
Shanghai Jiao Tong University	China	SCImago #10 · THE 40 · QS =47	11
Zhejiang University	China	SCImago #6 · THE 39 · QS 49	10
Fudan University	China	SCImago #46 · THE 36 · QS 30	9
Tsinghua University	China	SCImago #8 · THE 12 · QS =17	8
Shanghai AI Lab	China	—	8
Shanghai Artificial Intelligence Laboratory	China	SCImago #563	8
Nanyang Technological University	Singapore	SCImago #137	7
National University of Singapore	Singapore	SCImago #59 · THE 17 · QS 8	7
The University of Hong Kong	Hong Kong	SCImago #195 · THE 33 · QS 11	7
Nanjing University	China	SCImago #178 · THE =62 · QS =103	5

Geographic distribution of citing authors

Country	Citing papers
China	96
United States	28
Hong Kong	18
Singapore	14
United Kingdom	4
United Arab Emirates	3
Saudi Arabia	3
Canada	2
India	2

Country	Citing papers
Australia	2
United States; Hong Kong	1
Bulgaria	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	MIA-DPO: Multi-Image Augmented Direct Preference Optimization For Large Vision-Language Models	27	Dhanasar — Prong 2 (well-positioned)
Contribution 2	MMLONGBENCH-DOC: Benchmarking Long-context Document Understanding with Visualizations	3	Dhanasar — Prong 2 (well-positioned)
Contribution 3	Rar: Retrieving and ranking augmented mllms for visual recognition	11	Dhanasar — Prong 2 (well-positioned)