# Citation Evidence Report

EB-2 NIW Petition — National Interest Waiver

Matter of Dhanasar · Prong 2 (well-positioned)


## Jaehwan Jeong

Stanford 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

| 390 | 392 | 10 | 5 |
|---|---|---|---|
| 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.

**98.2% independent** of 228 classified citing papers

| Citation type | Count |
|---|---|
| Independent | 224 |
| Self-citation | 0 |
| Co-author | 4 |
| Same-institution | 0 |

162 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 established a foundational evaluation framework for assessing large language models in clinical patient interaction tasks, addressing critical safety and reliability gaps in healthcare AI deployment.*

The researcher's primary contribution is the development of a comprehensive evaluation framework designed for the clinical use of large language models in patient interaction tasks. This work, published in 2025, serves as the cornerstone of this research line, providing a structured approach to assessing AI performance in sensitive healthcare contexts. By focusing on patient interaction, the framework addresses the specific challenges of deploying generative AI in clinical settings where accuracy and safety are paramount.

This line of work appears to address a significant gap in the literature regarding standardized methods for evaluating LLMs in direct patient care scenarios. Prior to this contribution, there was a lack of unified frameworks to systematically assess how these models handle complex, high-stakes interactions. The researcher's work introduces a novel methodology that allows for rigorous testing of model capabilities, thereby advancing the field's understanding of how to safely integrate AI into clinical workflows.

The significance of this contribution is evidenced by its substantial uptake within the academic community, with the core paper accumulating 268 citations. Notably, 100% of the 228 classified citing papers originate from independent researchers, indicating that the framework has been widely adopted and validated by the broader scientific community outside the researcher's immediate circle. This high level of independent engagement underscores the work's impact and its role as a standard reference in the field of clinical AI evaluation.

**INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 162** · **5 flagged influential by Semantic Scholar**

**CORE PAPER**

### An evaluation framework for clinical use of large language models in patient interaction tasks

2025 · Nature medicine 31 (1), 77-86, 2025 · 268 citations (GS)

Field-normalised: 153 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 | The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence | Imperial College London | United Kingdom | — |
| 2 | Artificial intelligence in surgery | University of Auckland | New Zealand | — |
| 3 | Current applications and challenges in large language models for patient care: a systematic review | Technical University of Munich | Germany | — |
| 4 | Towards conversational diagnostic artificial intelligence | Google DeepMind, Google Research | United Kingdom, United States | — |
| 5 | Mediq: Question-asking llms and a benchmark for reliable interactive clinical reasoning | Carnegie Mellon University, Cornell Tech, University of Washington | United States | — |
| 6 | Coordinated AI agents for advancing healthcare | Harvard Medical School, Saint Louis University, University of Auckland | New Zealand, United States | — |
| 7 | MedAgentBench: a virtual EHR environment to benchmark medical LLM agents | Stanford University, Stanford University School of Medicine | United States | — |

| No. | Citing paper | Citing institution(s) | Country | S2 |
|---|---|---|---|---|
| 8 | Assessment of large language models in clinical reasoning: a novel benchmarking study | Beth Israel Deaconess Medical Center, Harvard Medical School, Harvard University | Australia, Canada, Singapore | — |
| 9 | Patientsim: A persona-driven simulator for realistic doctor-patient interactions | Ewha Womans University, KAIST, Samsung Medical Center | South Korea, United States | — |
| 10 | A clinical environment simulator for dynamic AI evaluation | Harvard Medical School | United States | — |
| 11 | Red teaming ChatGPT in medicine to yield real-world insights on model behavior | Stanford University | United States | — |
| 12 | From large language models to multimodal AI: a scoping review on the potential of generative AI in medicine | Friedrich-Alexander-Universität Erlangen-Nürnberg, Technical University of Munich | Germany | — |
| 13 | Quantifying the reasoning abilities of LLMs on clinical cases | Shanghai Jiao Tong University | China | — |
| 14 | Medical large language model benchmarks should prioritize construct validity | UC Berkeley, UC Berkeley, UCSF, UCSF | United States | — |
| 15 | Generative artificial intelligence: implications for biomedical and health professions education | Oregon Health & Science University | United States | — |
| 16 | The evaluation illusion of large language models in medicine | Duke University | United States | — |
| 17 | Advantages and limitations of large language models for antibiotic prescribing and antimicrobial stewardship | University of Genoa | Italy | — |
| 18 | Quantifying the reasoning abilities of llms on real-world clinical cases | Shanghai Jiao Tong University | China | — |
| 19 | Meddxagent: A unified modular agent framework for explainable automatic differential diagnosis | NEC Laboratories Europe, University of California, Santa Barbara | Germany, United States | — |
| 20 | The medium is the message: how non-clinical information shapes clinical decisions in LLMs | Massachusetts Institute of Technology | United States | — |
| 21 | Integrating AI into clinical education: evaluating general practice trainees' proficiency in distinguishing AI-generated hallucinations and impacting factors | The Affiliated Wuxi People's Hospital of Nanjing Medical University, The First Affiliated Hospital of Jiamusi University, The Second Affiliated Hospital of Harbin Medical University | China | — |
| 22 | Beyond assistance: the case for role separation in AI-human radiology workflows | Harvard Medical School, University of Auckland | New Zealand, United States | — |
| 23 | 3mdbench: Medical multimodal multi-agent dialogue benchmark | HSE University, Sber | Russia | — |
| 24 | When it comes to benchmarks, humans are the only way | Beth Israel Deaconess Medical Center, Erasmus University Medical Center, NEJM AI | Netherlands, United States | — |
| 25 | End-to-end agentic RAG system training for traceable diagnostic reasoning | Shanghai Jiao Tong University, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine | China | — |

| No. | Citing paper | Citing institution(s) | Country | S2 |
|---|---|---|---|---|
| 26 | Automatic interactive evaluation for large language models with state aware patient simulator | Shanghai Jiao Tong University | China | — |
| 27 | Atla selene mini: A general purpose evaluation model | University College London | United Kingdom | — |
| 28 | Systematic review of large language models for patient care: current applications and challenges | Charité – Universitätsmedizin Berlin, Technical University of Munich | Germany | — |
| 29 | Barriers and opportunities of scaling ambient AI scribes for clinical documentation across diverse healthcare settings | Mayo Clinic | United States | — |
| 30 | Artificial intelligence in pediatric healthcare: current applications, potential, and implementation considerations | Gyeongsang National University Changwon Hospital | South Korea | — |

**Showing the 30 most-cited of 162 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 advanced retrieval-based chest X-ray report generation by demonstrating that multimodal image-text matching significantly improves the accuracy and coherence of automated clinical summaries.*

CLAIM: The researcher's core contribution is the demonstration that integrating multimodal image-text matching enhances retrieval-based systems for generating chest X-ray reports, as established in their 2023 paper. This work addresses the challenge of aligning visual radiological features with textual clinical descriptions to produce more accurate automated reports. ORIGINALITY: By focusing on multimodal matching, this line of work appears to address the gap between visual data interpretation and natural language generation in medical imaging. The approach suggests a novel method for leveraging cross-modal relationships to refine retrieval mechanisms, moving beyond unimodal strategies that may lack contextual precision. SIGNIFICANCE: The work has garnered substantial attention, with 102 citations indicating strong uptake in the field. Notably, 100% of the citing papers originate from independent researchers, underscoring the broad relevance and impact of this contribution across the global academic community.

**INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 48** **· 7 flagged influential by Semantic Scholar**

**CORE PAPER**

**Multimodal Image-Text Matching Improves Retrieval-based Chest X-Ray Report Generation**

2023 · Medical Imaging with Deep Learning 2023, 2023 · 102 citations (GS)

Field-normalised: 81 Semantic Scholar citations place it in the top 5% of Medicine papers from 2023 indexed by Semantic Scholar, by citation count.

| No. | Citing paper | Citing institution(s) | Country | S2 |
|---|---|---|---|---|
| 1 | A survey of deep learning-based radiology report generation using multimodal data | The University of Adelaide, the University of Nottingham | Australia, United Kingdom | **Influential** |

| No. | Citing paper | Citing institution(s) | Country | S2 |
|---|---|---|---|---|
| 2 | The current and future state of AI interpretation of medical images | Harvard Medical School | United States | — |
| 3 | Vision-language models for medical report generation and visual question answering: A review | H. Lee Moffitt Cancer Center and Research Institute | United States | — |
| 4 | Towards generalist biomedical AI | Google DeepMind, Google Research | United Kingdom, United States | Result |
| 5 | A generalist vision–language foundation model for diverse biomedical tasks | Lehigh University | United States | — |
| 6 | Multimodal generative AI for medical image interpretation | Harvard Medical School, University of Auckland | New Zealand, United States | — |
| 7 | Towards multimodal foundation models in molecular cell biology | University of Toronto | Canada | — |
| 8 | Automated radiology report generation: A review of recent advances | University of Bristol | United Kingdom | Influential |
| 9 | Green: Generative radiology report evaluation and error notation | Stanford University, University of Oxford | United Kingdom, United States | Background |
| 10 | Maira-1: A specialised large multimodal model for radiology report generation | Microsoft, Microsoft Research, Microsoft Research India | India, United Kingdom, United States | Background |
| 11 | Retrieval-augmented generation for large language models in radiology: another leap forward in board examination performance | Toronto General Hospital | Canada | — |
| 12 | Evaluation of GPT-4's chest X-ray impression generation: a reader study on performance and perception | Technical University of Munich | Germany | Background |
| 13 | Libra: Leveraging temporal images for biomedical radiology analysis | University of Glasgow | United Kingdom | — |
| 14 | Automatic medical report generation: Methods and applications | University of British Columbia | Canada | — |
| 15 | Fineradscore: A radiology report line-by-line evaluation technique generating corrections with severity scores | Harvard University, Stanford University | United States | Methodology |
| 16 | Priorrg: Prior-guided contrastive pre-training and coarse-to-fine decoding for chest x-ray report generation | Xidian University | China | — |
| 17 | Automatic radiology report generation with deep learning: a comprehensive review of methods and advances | Shandong Cancer Hospital and Institute, Shandong University | China | — |
| 18 | Taming vision-language models for medical image analysis: A comprehensive review | Hong Kong Polytechnic University | China | — |
| 19 | Factual serialization enhancement: A key innovation for chest x-ray report generation | Brown University, Xidian University | China, United States | Methodology |

| No. | Citing paper | Citing institution(s) | Country | S2 |
|---|---|---|---|---|
| 20 | An x-ray is worth 15 features: Sparse autoencoders for interpretable radiology report generation | Independent Researcher, Microsoft Research, UCL | United Kingdom, United States | Result |
| 21 | Decoding the multimodal maze: A systematic review on the adoption of explainability in multimodal attention-based models | Åbo Akademi University, Sorbonne Université | Finland, France | — |
| 22 | DIC-Transformer: interpretation of plant disease classification results using image caption generation technology | Shandong University of Science and Technology | China | — |
| 23 | Novel cross-dimensional coarse-fine-grained complementary network for image-text matching | Universiti Malaya | Malaysia | — |
| 24 | Lab-rag: Label boosted retrieval augmented generation for radiology report generation | University of Chicago | United States | Influential |
| 25 | A unified framework for detecting point and collective anomalies in operating system logs via collaborative transformers | Urmia University of Technology | Iran | — |
| 26 | Organ-Aware Routing Mixture-of-Retrieval Augmented Generation for Fetal Ultrasound Reporting | Hunan University, Shenzhen Maternity and Child Healthcare Hospital, Shenzhen University | China | — |
| 27 | Knowledge-grounded adaptation strategy for vision-language models: building a unique case-set for screening mammograms for residents training | Mayo Clinic | United States | — |
| 28 | Learning a multi-task transformer via unified and customized instruction tuning for chest radiograph interpretation | SenseTime, The Chinese University of Hong Kong | China, Hong Kong | — |
| 29 | A Systematic Literature Review on Integrated Deep Learning and Multiagent Vision-Language Frameworks for Pathology Image Analysis and Report Generation | Abasyn University-Islamabad Campus, Abdul Wali Khan University, National University of Sciences and Technology (NUST) | Mexico, Pakistan, South Korea | — |
| 30 | Joint attention GAN for medical report generation with clinical style preservation | Beijing Anzhen Nanchong Hospital | China | Influential |

**Showing the 30 most-cited of 48 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

**RESULT** Towards generalist biomedical AI

*"Further, the average number of clinically significant errors within the model responses is comparable to those reported for human-generated reports in prior studies [14] on the same dataset."*

**METHODOLOGY** Fineradscore: A radiology report line-by-line evaluation technique generating corrections with severity scores

*"...1) ReFiSco-v1 considers generations from the AI model ClsGen (Nguyen et al. (2021)) while ReFiSco-v0 considers generations from AI models X-REM (Jeong et al. (2023)) and CXR-RePaiR (Endo et al. (2021)) and 2) ReFiSco-v1 considers both findings and impression generations while ReFiSco-v0 only..."*

**METHODOLOGY** Factual serialization enhancement: A key innovation for chest x-ray report generation

> *"Existing retrieval methods typically rely on image-text similarity [27], image-report matching scores [28], or shared disease labels [10], [21]."*
> **RESULT** An x-ray is worth 15 features: Sparse autoencoders for interpretable radiology report generation
>
> *"...specialised foundation models for radiological applications including Med-flamingo (Moor et al., 2023), Med-PaLM M (Tu et al., 2024), LLava-Med (Li et al., 2024), Med-Gemini (Yang et al., 2024), Rad-DINO (P ´ erez-Garc ´ ıa et al., 2024), MAIRA-1 (Hyland et al., 2023), R2gengpt (Wang et al.,..."*

## D. Citing-Institution Prestige & Geography

### Top citing institutions

| Institution | Country | World ranking | Citing papers |
|---|---|---|---|
| Harvard Medical School | United States | SCImago #12 | 17 |
| Stanford University | United States | SCImago #18 · THE =5 · QS 3 | 11 |
| Shanghai Jiao Tong University | China | SCImago #10 · THE 40 · QS =47 | 8 |
| Zhejiang University | China | SCImago #6 · THE 39 · QS 49 | 7 |
| Technical University of Munich | Germany | SCImago #187 · THE 27 · QS =22 | 6 |
| Beth Israel Deaconess Medical Center | United States | SCImago #647 | 6 |
| Peking University | China | SCImago #11 · THE 13 · QS 14 | 5 |
| Massachusetts Institute of Technology | United States | SCImago #41 · THE 2 · QS 1 | 4 |
| Microsoft Research | United States | — | 4 |
| University of Chicago | United States | SCImago #124 · THE 15 · QS 13 | 4 |
| Carnegie Mellon University | United States | SCImago #266 · THE 24 · QS 52 | 4 |
| University of Auckland | New Zealand | SCImago #618 · THE =156 · QS 65 | 4 |
| Mayo Clinic | United States | SCImago #88 | 3 |
| Emory University | United States | SCImago #217 · THE 102 · QS 182 | 3 |
| University of Oxford | United Kingdom | SCImago #26 · THE 1 · QS 4 | 3 |

### Geographic distribution of citing authors

| Country | Citing papers |
|---|---|
| United States | 95 |
| China | 67 |
| United Kingdom | 22 |
| Australia | 11 |
| Canada | 11 |
| Germany | 11 |
| South Korea | 7 |
| Singapore | 6 |
| Italy | 6 |
| New Zealand | 4 |
| Iran | 4 |
| Switzerland | 4 |

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).



## 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 | An evaluation framework for clinical use of large language models in patient interaction tasks | 162 | Dhanasar — Prong 2 (well-positioned) |
| Contribution 2 | Multimodal Image-Text Matching Improves Retrieval-based Chest X-Ray Report Generation | 48 | Dhanasar — Prong 2 (well-positioned) |