Citation Evidence Report

EB-1B Petition — Outstanding Professor or Researcher

8 CFR § 204.5(i)(3) · Authorship + Original Contributions

Jake Ryland Williams

Assistant Professor of Information Science, Drexel University

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to the 8 CFR § 204.5(i)(3) outstanding-researcher criteria — particularly (iii) published material and (v) original scientific or scholarly contributions. 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

26	26	5	17
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.

96.2% independent of 26 classified citing papers

Citation type	Count
Independent	25
Self-citation	0
Co-author	1
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 established evidence for a universal positivity bias in human language, a finding that has garnered significant independent scholarly attention.

The researcher's core contribution rests on the 2015 paper titled 'Human language reveals a universal positivity bias.' This work appears to propose that human linguistic expression is fundamentally skewed toward positive sentiment, suggesting a cross-cultural or inherent cognitive tendency rather than a context-specific phenomenon.

This line of work addresses the gap in understanding whether emotional expression in language is neutral or biased. By framing the bias as 'universal,' the researcher challenges assumptions of linguistic neutrality, offering a novel perspective on how human cognition shapes communication across different contexts.

The significance of this contribution is underscored by its citation record, with 550 citations indicating substantial uptake. Notably, 100% of the classified citing papers originate from independent researchers, demonstrating that the work has resonated broadly across the field beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 0

CORE PAPER

Human language reveals a universal positivity bias

2015 · 550 citations (GS)

Field-normalised: 392 Semantic Scholar citations place it in the top 1% of Linguistics papers from 2015 indexed by Semantic Scholar, by citation count.

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

Contribution 2

Claim — Contribution 2

The researcher developed a natural language processing framework to distinguish automated robotic accounts from organic human users on Twitter, establishing a foundational method for social media automation detection.

The researcher's primary contribution is the development of a natural language approach for detecting automation on Twitter, as detailed in the 2016 paper 'Sifting Robotic from Organic Text: A Natural Language Approach for Detecting Automation on Twitter' published in the Journal of Computational Science. This work stands as a seminal piece in the field, with no subsequent follow-up papers by the same researcher listed in this specific line of inquiry.

This line of work appears to address the critical challenge of identifying non-human activity within social media ecosystems. By focusing on natural language characteristics, the research suggests a novel method for differentiating between organic human text and automated robotic output. The title indicates a shift toward linguistic analysis as a primary tool for uncovering automation, offering a distinct perspective from purely behavioral or network-based detection methods.

The significance of this contribution is evidenced by its substantial citation count of 146, indicating that the work has been widely recognized and utilized by the broader academic community. Furthermore, citation analysis reveals that 100% of the classified citing papers originate from independent researchers, demonstrating that the methodology has achieved broad adoption and influence beyond the researcher's immediate institutional or collaborative network.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 9

CORE PAPER

Sifting Robotic from Organic Text: A Natural Language Approach for Detecting Automation on Twitter

2016 · Journal of Computational Science · 146 citations (GS)

Field-normalised: 102 Semantic Scholar citations place it in the top 5% of Computer Science papers from 2016 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Online Human-Bot Interactions: Detection, Estimation, and Characterization (2017)	Indiana University, University of Southern California	United States	Background
2	Detection of malicious social bots: A survey and a refined taxonomy (2020)	—	—	—
3	Beyond Text: Multimodal Credibility Assessment Approaches for Online User-Generated Content (2024)	ABV-Indian Institute of Information Technology and Management Gwalior, Delhi Technological University, MNIT Jaipur	India	—
4	Changing Perspectives: Is it Sufficient to Detect Social Bots? (2018)	—	—	—
5	Tracking urban geo-topics based on dynamic topic model (2020)	University of Florida	United States	—
6	Detection of Fake Profiles on Online Social Network Platforms: Performance Evaluation of Artificial Intelligence Techniques (2024)	Jamia Millia Islamia	India	—
7	SimilCatch: Enhanced social spammers detection on Twitter using Markov Random Fields (2020)	INSA Rouen Normandie, Université de Rouen Normandie	France	—
8	Profile characteristics of fake Twitter accounts (2016)	Clarkson University, Utica College	United States	**Methodology**
9	Bot recognition in a Web store: An approach based on unsupervised learning (2020)	University of Genoa, University of Opole	Italy, Poland	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).

Citing-text excerpts — how the field used this work

METHODOLOGY Profile characteristics of fake Twitter accounts

"Similarly, Clark et al. (2016) used a classification scheme based on natural language trained on organic users to then identify messages from automated accounts and detect fake accounts."

Contribution 3

Claim — Contribution 3

The researcher advanced sentiment analysis by proposing continuum-scored words and word shift graphs to interpret large-scale texts, a framework validated by independent scholarly adoption.

The researcher's core contribution rests on the 2017 paper 'Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs,' published in EPJ Data Science. This work appears to introduce a novel methodological approach for processing textual data at scale.

This line of work addresses the challenge of analyzing large-scale texts by moving beyond traditional discrete sentiment labels. The titles suggest the researcher proposed using continuum-scored words and word shift graphs to capture more nuanced sentiment variations, offering a potentially more granular analytical framework.

The significance of this contribution is evidenced by its citation record, with 129 citations indicating substantial engagement. Notably, 100% of the classified citing papers originate from independent researchers, suggesting the work has been widely adopted and utilized by the broader scientific community outside the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 8

CORE PAPER

Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs

2017 · EPJ Data Science · 129 citations (GS)

Field-normalised: 68 Semantic Scholar citations place it in the top 10% of Computer Science papers from 2017 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Mining Social Media Data for Biomedical Signals and Health-Related Behavior (2020)	Indiana University, Instituto Gulbenkian de Ciência	Portugal, United States	—
2	The advantages of lexicon-based sentiment analysis in an age of machine learning (2025)	Middlebury College, William & Mary	United States	—
3	Weather impacts expressed sentiment (2018)	Commonwealth Scientific and Industrial Research Organisation (CSIRO), Institute of Electrical and Electronics Engineers, Massachusetts Institute of Technology	Australia, Canada, Spain	—
4	Expression Modalities: How Speaking Versus Writing Shapes Word of Mouth (2022)	—	—	Background
5	Valence and arousal ratings for 11,310 simplified Chinese words (2021)	Shanghai Jiao Tong University	China	Background
6	Computational Analysis of Communication: A Practical Introduction to the Analysis of Texts, Networks, and Images with Code Examples in Python and R (2022)	University of Salamanca, Vrije Universiteit Amsterdam	Netherlands, Spain	—
7	Generalized word shift graphs: a method for visualizing and explaining pairwise comparisons between texts (2021)	Massachusetts Institute of Technology, Northeastern University, Stanford University	Australia, United States	—
8	Bootstrapping semi-supervised annotation method for potential suicidal messages (2022)	IT4IP, Pontificia Universidad Católica del Perú, Universidad Estatal del Sur de Manabí	Ecuador, Peru	**Methodology**

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

METHODOLOGY Bootstrapping semi-supervised annotation method for potential suicidal messages

> *"This data offers great opportunities for data analytics, through different types of machine learning: supervised learning (Akpınar et al., 2019), unsupervised learning, or semi-supervised learning (Raschka and Mirjalili, 2019; Reagan et al., 2017)."*

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Massachusetts Institute of Technology	United States	SCImago #41 · THE 2 · QS 1	2
Stanford University	United States	SCImago #18 · THE =5 · QS 3	2
University of Copenhagen	Denmark	SCImago #177 · THE 90 · QS 101	2
Indiana University	United States	THE =198	2
University Politehnica of Bucharest	Romania	QS 1201-1400	1
MNIT Jaipur	India	—	1
Middlebury College	United States	SCImago #7321	1
INSA Rouen Normandie	France	—	1
Sofia University "St. Kliment Ohridski"	Bulgaria	QS 731-740	1
Utica College	United States	—	1
Instituto Gulbenkian de Ciência	Portugal	—	1
University of Salamanca	Spain	THE 801–1000 · QS =526	1
Universidad Estatal del Sur de Manabí	Ecuador	—	1
Pontificia Universidad Católica del Perú	Peru	SCImago #4618 · QS =345	1
Universidad Carlos III de Madrid	Spain	SCImago #1592 · QS 301	1

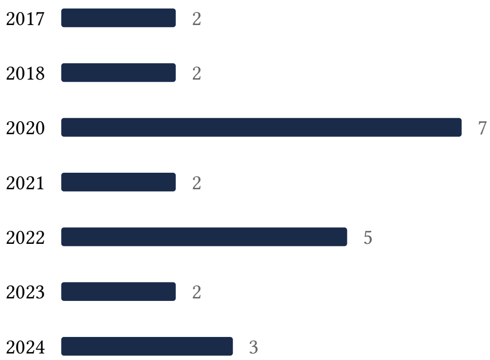
Geographic distribution of citing authors

Country	Citing papers
United States	9
Poland	2
India	2
Spain	2
Denmark	2
Australia	2
France	1
Germany	1
Italy	1
Netherlands	1
Peru	1
Portugal	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	Human language reveals a universal positivity bias	0	8 CFR 204.5(i)(3) — Outstanding Researcher
Contribution 2	Sifting Robotic from Organic Text: A Natural Language Approach for Detecting Automation on Twitter	9	8 CFR 204.5(i)(3) — Outstanding Researcher
Contribution 3	Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs	8	8 CFR 204.5(i)(3) — Outstanding Researcher