HÉCTOR J. VÁZQUEZ MARTÍNEZ

✉ hjvm@upenn.edu hjvm.github.io linkedin.com/in/hjvm/ github.com/hjvm

SUMMARY

Applied AI scientist and final-year Ph.D. candidate in Linguistics at UPenn (B.S./M.Eng. EECS, MIT). My research asks how language models represent language, how to evaluate them rigorously, and how to make speech and language AI work with far less data. In industry I have led production NLP for the U.S. Air Force, co-invented three U.S. patents in explainable AI and natural-language interfaces, brought in $1.25M+ in AI contracts, and built agentic AI products and an AI strategy roadmap for a fabrication-tech startup. I also organize academic conferences and research incubators that seed societally impactful AI research.

EDUCATION

University of Pennsylvania **Fall 2022 - May 2027 (expected)**
PhD in Linguistics *GPA: 3.94/4.0*
Dissertation: Sample-efficient speech representation learning and speech segmentation.

Massachusetts Institute of Technology **Jun. 2015 - Jun. 2021**
M.Eng., B.Sc. in Electrical Engineering and Computer Science *GPA: 5.0/5.0*
M.Eng. thesis: Novel linguistic methods for evaluating neural language models.

ACADEMIC PUBLICATIONS

H. J. Vázquez Martínez, "findsylls: A Language-Agnostic Toolkit for Syllable-Level Speech Tokenization and Embedding," *Proc. Interspeech 2026*, 6568–6573, doi: 10.21437/Interspeech.2026-820. [**Best Student Paper nominee**] Code: github.com/hjvm/findsylls.

H. J. Vázquez Martínez & C. Yang, "Measuring Form and Function in Language Models," *Proceedings of the Second BabyLM Workshop* (ACL), 2026, 438–458, doi: 10.18653/2026.babylm-main.35.

H. J. Vázquez Martínez, A. Heuser, C. Yang & J. Kodner "Evaluating the Existence Proof: LLMs as Cognitive Models of Language Acquisition," in Mendívil-Giró, J-L. (Ed.), Artificial Knowledge of Language: A Linguist's Perspective on Its Nature, Origins and Use, Vernon Press, 2026.

Daultani, Vijay, **H. J. Vázquez Martínez**, and Naoaki Okazaki. "Acceptability Evaluation of Naturally Written Sentences." Journal of Information Processing 32 (2024): 652-666. https://doi.org/10.2197/ipsjjip.32.652

H. J. Vázquez Martínez, A. Heuser, C. Yang & J. Kodner "Evaluating Neural Language Models as Cognitive Models of Language Acquisition," *GenBench2023 workshop: The first workshop on (benchmarking) generalisation in NLP*, Singapore, 2023. https://doi.org/10.18653/v1/2023.genbench-1.4

H. J. Vázquez Martínez & B. Santorini, "A variational model of the loss of English OV," *24th Conference on Diachronic Generative Syntax*, Paris, France, 2023.

H. J. Vázquez Martínez, "The Acceptability Delta Criterion: Testing Knowledge of Language using the Gradience of Sentence Acceptability," *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, Punta Cana, Dominican Republic, 2021. https://doi.org/10.18653/v1/2021.blackboxnlp-1.38

S. Indurkhya, **H. J. Vázquez Martínez**, A. Indurkhya, C. Donalek, "Using Natural Language Processing to Construct a Knowledge Graph of Test Incident Reports," *Proceedings of the 38th International Test & Evaluation Symposium*, Online, 2021.

H. J. Vázquez Martínez, S. Indurkhya, G. Zanfardino, A. Indurkhya, S. Sahu, C. Donalek, M. Amori, "XAI Methods in VIP bring us closer to interpretable Network Graphs," *Caltech Explainable AI (XAI) Workshop*, Pasadena, California, 2021.

H. J. Vázquez Martínez (2021). The acceptability delta criterion: memorization is not enough (Master's Thesis, Massachusetts Institute of Technology). DSpace@MIT. https://hdl.handle.net/1721.1/130703

PATENTS

S. Indurkhya, **H. J. Vázquez Martínez**, A. Salimov, A. Indurkhya, G. Zanfardino, E. Sloan, C. Donalek, M. Amori, (2023). Systems and Methods for Numeric Network Extraction (U.S. Patent No. US-20230306044-A1). U.S. Patent and Trademark Office.

H. J. Vázquez Martínez, S. Indurkhya, G. Zanfardino, A. Indurkhya, S. Sahu, C. Donalek, M. Amori, (2023). Systems and Methods for Network Explainability (U.S. Patent No. US-20230004557-A1). U.S. Patent and Trademark Office.

S. Indurkhya, **H. J. Vázquez Martínez**, G. Zanfardino, C. Donalek, (2023). Systems and Methods for Natural Language Querying (U.S. Patent No. US-20220342873-A1). U.S. Patent and Trademark Office.

INDUSTRY EXPERIENCE

Jeevy Fabrication **Contract, part-time**
AI Strategy Consultant & Lead AI Engineer *Aug. 2026 – Nov. 2026*

- Built agentic AI features into the Jeevy Gnosis platform, cutting the time project managers and estimators spend on project planning, quoting, and resource allocation by an order of magnitude (10x).
- Wrote the company's AI strategy roadmap and secured buy-in from the CEO, CTO, and COO.
- Won admission to NVIDIA Inception and Google for Startups, securing over $400K in AI compute credits and access to their startup and venture-capital networks.

Virtualitics, Inc **Pasadena, CA**
Artificial Intelligence (AI) Engineer *Jun. 2024 – Aug. 2024*
Natural Language Processing (NLP) Engineer *Feb. 2021 – Aug. 2022*

- Lead engineer and project manager on a U.S. Air Force (AFRL) SBIR Phase II contract (*No. FA864922P0579*), held with a U.S. Secret clearance (now inactive). Built an NLP pipeline that analyzes B-52 software deficiency reports, collaborated with Air Force analysts to deliver all milestones, and deployed it to the production platform; $750K in revenue for Virtualitics.
- Brought in $500K in new client contracts with patented invention of explainable AI capability for graph analytics (US Patent No. US-20230004557-A1) that accelerated adoption of Virtualitics' core network-analysis routines.
- Reduced time to first insight for non-technical users of Virtualitics' software from more than 1 hour to 10–20 minutes with a patented natural language recommendation system (US Patent No. US-20220342873-A1) that suggested relevant workflows from live data.
- Returned to lead a 5-person team on generative AI integration, automating key parts of the analytics pipeline and delivering demos to executives, investors, and clients that secured buy-in for a 12-month roadmap to scale client-facing AI features.

AWARDS & HONORS

- UPenn Harrison Graduate Fellow *2025–2026*
- UPenn Fontaine Fellow *2022–2026*
- Hispanic Scholarship Fund (HSF) *2016-2026*
- Google PhD Fellowship UPenn Nominee *May 2025*
- National GEM Consortium Employer Fellow *Nov. 2024*
- MIT Charles and Jennifer Johnson Artificial Intelligence and Decision Thesis Award *Jun. 2022*
- MIT Harold L. Hazen Teaching Award *May 2021*
- Zeno Karl Schindler Summerschool Grant *May 2019*
- MIT Excellence Award: Albert G. Hill Prize *May 2018*
- MIT EECS Research & Innovation Scholar *2017–2018*
- Palantir Future Scholarship *Oct. 2017*

CERTIFICATIONS & CLEARANCE

- Technical AI Safety Project, BlueDot Impact *June 2026*
- Technical AI Safety, BlueDot Impact *May 2026*
- RAG & Agentic AI Specialization, IBM
- Certificate in Language and Communication Sciences, University of Pennsylvania *May 2025*
- Certificate in College and University Teaching, University of Pennsylvania *May 2025*
- U.S. Government Secret clearance, held for U.S. Air Force work (inactive)

SKILLS

- **Programming:** Python, C#, R, MATLAB
- **Applied AI / GenAI:** Agentic systems, LangGraph, LangChain, RAG, vector DBs, MCP, LLM evaluation & benchmarking
- **ML & Deep Learning:** PyTorch, Hugging Face Transformers, scikit-learn, TensorFlow (basic)
- **Speech & NLP:** Representation learning, fine-tuning, interpretability, data-efficient modeling, speech segmentation, TextGrid/Praat
- **Data & Scientific Computing:** NumPy, Pandas, Matplotlib, Seaborn, Librosa
- **Tools & Workflow:** Git/GitHub, Docker (basic), REST APIs, Jupyter, VSCode
- **Communication:** AI strategy & roadmaps, executive & client demos, technical, grant & patent writing, stakeholder coordination
- **Languages:** Spanish (native), English (native/bilingual), German (professional)

LEADERSHIP & SERVICE

AI4GOOD Research Incubator **Philadelphia, PA**
Co-founder, Chair *Dec. 2024 - May 2025*

- Hosted first interdisciplinary AI workshop series and research incubator leading to 9 interdisciplinary research proposals for socially impactful applications of AI. One of 12 projects selected from 230 proposals university-wide to be funded by UPenn's inaugural *Draw Down the Lightning Grant* ($25K USD).

47th & 48th annual Penn Linguistics Conference **Philadelphia, PA**
Conference Chair *Sep. 2022 - May 2024*
- Across both years, won over $30K USD in grants, managed a budget of over $50K and a team of 30 graduate students. Brought 4 academic leaders as keynote speakers and attracted over 200 attendees in total. Generated over $10K in profit, guaranteeing sufficient funding for PLC49 and PLC50.

TEACHING EXPERIENCE

Certification in College and University Teaching University of Pennsylvania

Graduate TA (PhD) **Philadelphia, PA**
LING 0001 - Intro to Linguistics *Spring 2025*
LING 4000 - Tutorial in Linguistics *Fall 2024*
LING 0500 - Intro. to Formal Linguistics *Spring 2024*
LING 0001 - Intro to Linguistics *Fall 2023*

Head TA (M.Eng.) **Cambridge, MA**
6.034 - Artificial Intelligence *Fall 2020*

Graduate TA (M.Eng.) **Cambridge, MA**
6.08 - Intro. to Embedded Systems *Spring 2020*
6.034 - Artificial Intelligence *Fall 2019*
6.833 - The Human Intelligence Enterprise *Spring 2019*

Instructor (B.S.) **Santiago, Chile**
Intro. to Programming *January 2016*

RELEVANT COURSEWORK*

PhD coursework
CIS 6300 - Advanced NLP *Spring 2025*
CIS 6200 - Advanced Deep Learning *Spring 2025*
CIS 6250 - Theory of Machine Learning *Fall 2024*
CIS 5300 - Natural Language Processing *Fall 2024*
LING 5900 - Deep Learning and Large Language Models in Linguistic Research *Spring 2024*
LING 5700 - Distributional Learning *Fall 2023*
LING 5210 - Phonetics I: Experimental *Fall 2023*
LING 5250 - Computer Analysis and Modeling of Biological Signals and Systems *Spring 2023*
LING 5220 - Phonetics II: Data Science *Spring 2023*
LING 5150 - Dynamics of Language *Fall 2022*

M.Eng. coursework
6.884 – Neuro-symbolic Models for NLP *Fall 2020*
9.190 – Computational Psycholinguistics *Spring 2020*
9.530 – Emergent Computations within Distributed Neural Circuits *Spring 2020*
9.S916 – Deep-learning and symbolic models of human language *Spring 2020*
9.520/6.860 – Statistical Learning Theory and Applications *Fall 2019*

B.S. coursework
6.S081 – Human Computational Intelligence *Fall 2018*
227-0046-10L – Signals and Systems II *Spring 2018*
9.66/6.804 – Computational Cognitive Science *Fall 2017*
6.036 – Introduction to Machine Learning *Fall 2017*
6.003 – Signals & Systems *Fall 2017*
6.803 – The Human Intelligence Enterprise *Spring 2017*
6.141 – Robotics: Science and Systems I *Spring 2017*
6.034 – Introduction to Artificial Intelligence *Fall 2016*

PROFESSIONAL & EXTRACURRICULAR AFFILIATIONS

- Member, Association for Computational Linguistics *2021-present*
- Member, Linguistics Society of America *2024-present*
- Member, Fairmount Rowing Association *2023–present*
- Member, Wharton Crew Rowing Team *2022–present*
- Member, Treinta y Tres Delaware (Rueda de Casino) Dance and Performance Team *2022–present*
- Curriculum Director and Dance Instructor, MIT Casino Rueda *2018–2020*
- Team Member, NCAA Division 1 MIT Men's Lightweight Rowing Team *2015–2019*
- Member, Association of Puerto Rican Students at MIT *2015–2021*