

Selene Báez Santamaría

 selbaez.github.io |  [selBaez](#) |  [selene-báez-santamaría-53a1b496](#) |  0000-0003-3808-2614
 selene.baez.santamaria@gmail.com |  +31 (6) 4286-1744

PROFILE

I am a computational scientist on Conversational AI, specializing in Knowledge Representation (KR) for Natural Language Processing (NLP). Through neuro-symbolic approaches, I research dialogue representation and evaluation. I am determined and thrive in collaborative workplaces that encourage self-growth through challenging projects. I appreciate data-driven decisions and the value these bring to stakeholders.

TECHNICAL SKILLS

Programming	Python (HuggingFace, nltk, spacy, pytorch, tensorflow, keras, pandas, PySpark, scikit learn, seaborn, Flask), Java, Matlab, Prolog, Haskell, ROS.
Databases	SQL, SPARQL/RDF, Neo4j, NoSQL.
Robots	Softbank Robotics (Pepper, Nao).
Version Control	Git (9 packages published on PyPi as CLTL).
DevOps	AWS(EC2, RDS, Lambda, Cognito)
Languages	Spanish (native), English (fluent), French (beginner), Dutch (beginner)

RESEARCH EXPERIENCE

[Dynamic and Distributed Information Systems Group, UZH](#) — Zurich, CHE

PostDoctorate

Feb 2025 – *Present*

Work with [Prof. Abraham Bernstein](#) with a focus on rare disease diagnosis via patient similarity. Co-wrote grant proposals for conversational AI projects on digital democracy and recommender systems.

[Instituto de Investigaciones en Matematicas Aplicadas y en Sistemas, UNAM](#) — Mexico City, MEX

Visiting Researcher

Mar 2024 – Jun 2024

Work on Hate Speech and automatic Counter-Narrative generation using graph representations of dialogue. Collaboration with [Dr. Helena Gomez](#) and [Dr. Ilia Markov](#).

[Computational Linguistics and Text-Mining Lab, VU Amsterdam](#) — Amsterdam, NLD

PhD

Mar 2020 – Dec 2024

My research focuses on knowledge exchange via dialogue between AI agents and humans in collaborative tasks. Supervised by [Prof. Piek Vossen](#), as part of the [Hybrid Intelligence Centre](#).

University Research Fellow

Sep 2017 – Aug 2019

In the Spinoza project: “[Understanding Language by Machines](#)”, I enabled a Pepper robot to learn from open-domain dialogue. Using Speech/Object recognition, NLP and KR techniques, I approach artificial cognition aspects such as provenance, theory of mind, relevance and permanence.

[College of Metropolitan Transportation, BJUT](#) — Beijing, CHN

Visiting Researcher

Feb 2017 - Jun 2017

Work on comparing supervised and unsupervised machine learning techniques to uncover travel pattern behaviour. Collaboration with [Prof. Jiancheng Weng](#) and [Prof. Zhisheng Huang](#).

[Laboratory of Computational Intelligence, UBC](#) — British Columbia, CAN

Lab Assistant

Nov 2013 – May 2014

On the “CanWheel” project, which enhances the navigation system of a semi-autonomous robotic wheelchair, my role involved conducting WOZ user experiments, and analyzing user interviews.

PROFESSIONAL EXPERIENCE

myTomorrows — Amsterdam, NLD

Artificial Intelligence Researcher

Jun 2019 – Feb 2020

I was responsible for transforming medical unstructured language to generate a semantic knowledge graph. I supervised interns’ master theses resulting in three paper publications.

Data Scientist

Oct 2017 – Jun 2019

I implemented an ETL pipeline for big data ingestion of approved and pre-approval medical interventions. I developed a search engine and their corresponding APIs for data retrieval.

Center for Advanced Studies, IBM — Amsterdam, NLD

Watson Demo Intern

Jul 2016 – Jan 2017

I built multimodal AI applications using the Watson Developer Cloud and Nao robots as interface.

Learning and Technology Services (LTS), Sauder School of Business, UBC — Vancouver, BC

Faculty Liaison

Jan 2015 – May 2015

I managed relations with faculty by analyzing in-classroom needs and incorporating learning tools.

EDUCATION

Vrije Universiteit Amsterdam (VU Amsterdam)

Sep 2015 – Aug 2017

MSc in Artificial Intelligence: Intelligent Systems Design. (Cum Laude)

Thesis: Big public transportation data mining using machine learning for pattern recognition.

Techniques: ensemble classifiers, autoencoders, clustering, cluster computing, multi-threading tools.

University of British Columbia (UBC)

Sep 2011 – Aug 2014

B.Sc in Cognitive Systems: Computational Intelligence and Design. (w/distinction)

Thesis: Prediction of opponent team’s behaviour in the soccer RoboCup SSL tournament.

Techniques: Markov Chains

AWARDS

Research highlights

1st Place at the LM-KBC Challenge

ISWC 2022

1st Place at the Argument Mining Challenge

COLING 2022

Best Paper Award at Workshop on Customized Chat Grounding

COLING 2022

2nd Place at the Hackathon on eXplainable AI

500 EUR, TNO 2018

Scholarships and grants

Network Institute Research Voucher (with [Lea Krause](#), [Jan Kalo](#))

1,500 EUR, VU Amsterdam 2023

Holland Scholarship Programme

5,000 EUR, VU Amsterdam 2016

VU Fellowship Programme Scholarship

15,000 EUR, VU Amsterdam 2015

Faculty of Science International Students Scholarship

12,000 CAD, UBC 2014

Trek Excellence Scholarship for Continuing Students

4,000 CAD, UBC 2013

President’s Entrance Scholarship

5,000 CAD, UBC 2012

CONFERENCES

Attendance [ACL 2025](#), [EMNLP 2024](#), [NAACL 2024](#), [BNAIC 2023](#), [SIGDIAL 2023](#), [COLING 2022](#), [HHAI 2022](#), [ISWC 2019](#), [ACL 2019](#), [TSD 2018](#), [ECAI 2016](#), [ECCV 2016](#)

Reviewing [WMT@EMNLP 2025](#), [MMSR@IWCS 2025](#), [ARR February 2025](#), [COLING 2025](#), [ARR December 2024](#), [EMNLP 2024 \(ARR June 2024\)](#), [NAACL 2024 \(ARR December 2023\)](#), [CRUM@UMAP 2023](#), [EACL 2023](#), [MMAI@HHAI 2022](#)

TALKS

Interviews [Blog Post @myTomorrows \(2018\)](#), [Guest speaker @”So you want to be a Data Scientist” Podcast \(2020\)](#), [Promotional video for VU Amsterdam \(2021\)](#)

Presentations [Speaker @FoMo: Foundation models in Amsterdam \(2023\)](#), [Demo for Robert Dijkgraaf, Minister of Education, Culture and Science of Netherlands \(2022\)](#), [Keynote \(representing Robert-Jan Sips\) @2nd Annual Robotics and Artificial Intelligence Summit \(2018\)](#), [Presentation \(helping Piek Vossen\) @Paradisolezing \(2018\)](#)

PUBLICATIONS

- [1] David Salvador Márquez et al. “NLP@IIMAS-CLTL at Multilingual Counterspeech Generation: Combating Hate Speech Using Contextualized Knowledge Graph Representations and LLMs”. In: *Proceedings of the First Workshop on Multilingual Counterspeech Generation*. Ed. by Helena Bonaldi et al. Abu Dhabi, UAE: Association for Computational Linguistics, Jan. 2025, pp. 29–36. URL: <https://aclanthology.org/2025.mcg-1.4/>.
- [2] Piek Vossen et al. “Extracting triples from dialogues for conversational social agents”. In: *HHAI 2025: Hybrid Human AI Systems for the Social Good*. Frontiers in Artificial Intelligence and Applications. IOS Press, 2025.
- [3] Selene Baez Santamaria et al. “Contextualized Graph Representations for Generating Counter-Narratives against Hate Speech”. In: *Findings of the Association for Computational Linguistics: EMNLP 2024*. Ed. by Yaser Al-Onaizan et al. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 7664–7674. DOI: [10.18653/v1/2024.findings-emnlp.450](https://doi.org/10.18653/v1/2024.findings-emnlp.450). URL: <https://aclanthology.org/2024.findings-emnlp.450/>.
- [4] Selene Báez Santamaría. “Knowledge-centered conversational agents with a drive to learn”. In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 4: Student Research Workshop)*. Ed. by Yang (Trista) Cao et al. Mexico City, Mexico: Association for Computational Linguistics, June 2024, pp. 83–92. URL: <https://aclanthology.org/2024.naacl-srw.10>.
- [5] Selene Báez Santamaría et al. “Knowledge acquisition for dialogue agents using reinforcement learning on graph representations”. In: *arXiv preprint arXiv:2406.19500* (2024). DOI: <https://doi.org/10.48550/arXiv.2406.19500>.
- [6] Pei-Yu Chen et al. “Intelligent Support Systems for Lifestyle Change: Integrating Dialogue, Information Extraction, and Reasoning”. In: *HHAI 2024: Hybrid Human AI Systems for the Social Good* (2024), pp. 457–459. DOI: <https://doi.org/10.3233/FAIA240223>.
- [7] Bernd JW Dudzik et al. “Hybrid Intelligence Supports Application Development for Diabetes Lifestyle Management”. In: *Journal of Artificial Intelligence Research* 80 (2024), pp. 919–929. DOI: <https://doi.org/10.1613/jair.1.15916>.

- [8] Lea Krause et al. “Graph Representations for Machine Translation in Dialogue Settings”. In: *Proceedings of the Ninth Conference on Machine Translation*. Ed. by Barry Haddow et al. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 1038–1046. DOI: [10.18653/v1/2024.wmt-1.106](https://doi.org/10.18653/v1/2024.wmt-1.106). URL: <https://aclanthology.org/2024.wmt-1.106/>.
- [9] Piek Vossen et al. “A Conversational Agent for Structured Diary Construction Enabling Monitoring of Functioning & Well-Being”. In: *HHA1 2024: Hybrid Human AI Systems for the Social Good* (2024), pp. 315–324. DOI: <https://doi.org/10.3233/FAIA240204>.
- [10] Selene Báez Santamaría et al. “The Role of Personal Perspectives in Open-Domain Dialogue”. In: *Artificial Intelligence and Machine Learning, BNAIC/Benelearn*. Springer Nature Switzerland, 2023. URL: https://bnaic2023.tudelft.nl/static/media/BNAICBENELEARN_2023_paper_23.a7a48c6078e92bcb7550.pdf.
- [11] Lea Krause et al. “Confidently Wrong: Exploring the Calibration and Expression of (Un)Certainty of Large Language Models in a Multilingual Setting”. In: *Proceedings of the Workshop on Multimodal, Multilingual Natural Language Generation and Multilingual WebNLG Challenge (MM-NLG 2023)*. Ed. by Albert Gatt et al. Prague, Czech Republic: Association for Computational Linguistics, Sept. 2023, pp. 1–9. URL: <https://aclanthology.org/2023.mmnlg-1.1>.
- [12] Lea Krause et al. “Leveraging Few-Shot Data Augmentation and Waterfall Prompting for Response Generation”. In: *Proceedings of The Eleventh Dialog System Technology Challenge*. Ed. by Yun-Nung Chen et al. Prague, Czech Republic: Association for Computational Linguistics, Sept. 2023, pp. 193–205. URL: <https://aclanthology.org/2023.dstc-1.22>.
- [13] Dimitrios Alivanistos et al. “Prompting as Probing: Using Language Models for Knowledge Base Construction”. In: (2022). URL: <https://ceur-ws.org/Vol-3274/paper2.pdf>.
- [14] Selene Báez Santamaría et al. “Access to care: analysis of the geographical distribution of healthcare using Linked Open Data”. In: *4th workshop on Semantic Web solutions for large-scale biomedical data analytics* (2022). DOI: <https://doi.org/10.48550/arXiv.2204.05206>.
- [15] Selene Báez Santamaría et al. “Evaluating Agent Interactions Through Episodic Knowledge Graphs”. In: *Proceedings of the 1st Workshop on Customized Chat Grounding Persona and Knowledge*. 2022, pp. 15–28. URL: <https://aclanthology.org/2022.ccgpk-1.3>.
- [16] Thomas Baier et al. “A modular architecture for creating multimodal agents”. In: *Workshop on the representation, sharing and evaluation of multimodal agent interaction* (2022). DOI: <https://doi.org/10.48550/arXiv.2206.00636>.
- [17] Michiel van der Meer et al. “Will It Blend? Mixing Training Paradigms & Prompting for Argument Quality Prediction”. In: *Proceedings of the 9th Workshop on Argument Mining*. 2022, pp. 95–103. URL: <https://aclanthology.org/2022.argmining-1.8>.
- [18] Selene Báez Santamaría et al. “EMISSOR: A platform for capturing multimodal interactions as Episodic Memories and Interpretations with Situated Scenario-based Ontological References”. In: *Proceedings of the 1st Workshop on Multimodal Semantic Representations (MMSR)*. 2021, pp. 56–77. URL: <https://aclanthology.org/2021.mmsr-1.6>.
- [19] Guusje Boomgaarda et al. “Learning Profile-Based Recommendations for Medical Search Auto-Complete”. In: (2021). URL: <http://ceur-ws.org/Vol-2846/paper34.pdf>.
- [20] Erick Velazquez Godinez et al. “Language Identification for Short Medical Texts.” In: *HEALTHINF*. 2020, pp. 399–406. DOI: <https://doi.org/10.5220/0008950903990406>.

- [21] Emmanouil Manousogiannis et al. “Normalization of Long-tail Adverse Drug Reactions in Social Media”. In: *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*. 2020, pp. 49–58. DOI: [10.18653/v1/2020.louhi-1.6](https://doi.org/10.18653/v1/2020.louhi-1.6).
- [22] Emmanouil Manousogiannis et al. “Give it a shot: Few-shot learning to normalize ADR mentions in Social Media posts”. In: *Proceedings of the Fourth Social Media Mining for Health Applications (# SMM4H) Workshop & Shared Task*. 2019, pp. 114–116. DOI: [10.18653/v1/W19-3219](https://doi.org/10.18653/v1/W19-3219).
- [23] Piek Vossen et al. “A communicative robot to learn about us and the world”. In: *2019 Annual International Conference on Computational Linguistics and Intellectual Technologies, Dialogue 2019*. 2019, pp. 728–743. URL: http://www.dialog-21.ru/media/4872/_-dialog2019scopus_rev2.pdf.
- [24] Piek Vossen et al. “Leolani: A robot that communicates and learns about the shared world”. In: *2019 ISWC Satellite Tracks (Posters and Demonstrations, Industry, and Outrageous Ideas), ISWC 2019-Satellites*. CEUR-WS. 2019, pp. 181–184. URL: <https://ceur-ws.org/Vol-2456/paper47.pdf>.
- [25] Piek Vossen et al. “Modelling context awareness for a situated semantic agent”. In: *International and Interdisciplinary Conference on Modeling and Using Context*. Springer. 2019, pp. 238–252. DOI: https://doi.org/10.1007/978-3-030-34974-5_20.
- [26] Quan Liang et al. “Individual travel behavior modeling of public transport passenger based on graph construction”. In: *Journal of Advanced Transportation* 2018 (2018). DOI: [10.1155/2018/3859830](https://doi.org/10.1155/2018/3859830).
- [27] Piek Vossen et al. “Leolani: a reference machine with a theory of mind for social communication”. In: *International conference on text, speech, and dialogue*. Springer. 2018, pp. 15–25. DOI: https://doi.org/10.1007/978-3-030-00794-2_2.
- [28] Selene Báez Santamaría. “Predicting opponent team activity in a RoboCup environment”. In: *arXiv preprint arXiv:1503.01446* (2015). DOI: <https://doi.org/10.48550/arXiv.1503.01446>.