

# Selene Báez Santamaría

 [selbaez.github.io](https://github.com/selbaez) |  [selBaez](https://twitter.com/selBaez) |  [selene-baez-santamaría-53a1b496](https://www.linkedin.com/in/selene-baez-santamaría-53a1b496)  
 [selene.baez.santamaria@gmail.com](mailto:selene.baez.santamaria@gmail.com) |  +31 (6) 4286-1744

## PROFILE

---

I am an Artificial Intelligence PhD researcher in the neuro-symbolic domain. My main area of interest is Knowledge Representation for Natural Language Processing. I work in Conversational AI, particularly in dialogue representation and evaluation. I am a determined and goal-oriented person with strong analytical and critical thinking. I favour collaborative workplaces that promote self-growth through challenging projects. I appreciate data-driven decisions and the value that data insights can bring to stakeholders.

## TECHNICAL SKILLS

---

Programming languages	Python (pandas, PySpark, tensorflow, keras, theano, scikit learn, nltk, spacy, seaborn), Java, Matlab, Prolog, Haskell, ROS.
Databases	SQL, SPARQL/RDF, Neo4j.
Robots	Softbank Robotics (Pepper, Nao).
Version Control	Git (Github: <a href="https://github.com/selBaez">selBaez</a> ).
Languages	Spanish (native), English (fluent), French (beginner), Dutch (beginner)

## RESEARCH EXPERIENCE

---

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

- *PhD student*

Mar 2020 - **Present**

- I study trust as a relationship between social agents in a multimodal world. My research aims to create and evaluate a computational model of trust, from a robot's perspective towards trusting humans in collaborative tasks. Supervised by Prof. Piek Vossen, this project is part of the [Hybrid Intelligence Centre](#).

[Computational Lexicology and Terminology Lab \(VU Amsterdam\)](#) — Amsterdam, NLD

- *University Research Fellow*

Sep 2017 - Aug 2019

- I participated in the Spinoza project: "[Understanding Language by Machines](#)", thus enabling a Pepper robot to learn from open-domain language. Using Speech/Object recognition, NLP and Knowledge Representation techniques, it tackles problems of artificial cognition such as provenance, theory of mind, relevance and permanence.

[Laboratory of Computational Intelligence, UBC](#) — Vancouver, BC

- *Lab Assistant*

Nov 2013 - May 2014

- I participated in the "[CanWheel](#)" project centered around improving the navigation system of a semi-autonomous robotic wheelchair. My main role was data collection via Wizard-of-Oz type of experiments with end-users, as well as the quantitative and qualitative analysis of user interviews.

## PROFESSIONAL EXPERIENCE

---

[myTomorrows](#) — Amsterdam, NLD

- *Artificial Intelligence Researcher*

Jun 2019 - Feb 2020

- I was responsible for processing medical unstructured language to generate structured data in the form of a semantic knowledge graph. I supervised interns' master theses resulting in three paper publications.

myTomorrows — Amsterdam, NLD

- *Data Scientist*

Oct 2017 - Jun 2019

- I implemented an ETL pipeline for big data ingestion of approved and pre-approval medical interventions. I developed of 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 AI applications using the Watson Developer Cloud for backend services with Nao robots as interface. I was in charge of the end-to-end POC development combining cognitive tools related to natural language, vision and speech.

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

- *Faculty Liaison*

Jan 2015 - May 2015

- I helped in incorporating interactive learning tools for students, and built course infrastructure and analysis of in-classroom needs. I was integral in managing relations with key faculty.

## EDUCATION

---

Vrije Universiteit Amsterdam (VU Amsterdam) — Amsterdam, NLD

- *MSc in Artificial Intelligence: Intelligent Systems Design.*

*Cum Laude*

Sep 2015 – Aug 2017

- Courses: Machine Learning, Deep Learning, Data Mining, Computer Vision, NLP, Multi-Agent Systems.

Beijing University of Technology (BJUT) — Beijing, CHN

- *Master Thesis*

Feb 2017 - Jun 2017

- Thesis: Big public transportation data mining using supervised and unsupervised machine learning for feature engineering and pattern recognition. Techniques: ensemble classifiers, autoencoders, clustering, cluster computing, multi-threading tools.

University of British Columbia (UBC) — Vancouver, BC

- *B.Sc in Cognitive Systems: Computational Intelligence and Design.*

*Graduated with distinction*

Sep 2011 – Aug 2014

- Thesis: Prediction of the 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

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

## PUBLICATIONS

---

- Alivanistos, Dimitrios et al. (2022). “Prompting as Probing: Using Language Models for Knowledge Base Construction”. In: URL: <https://ceur-ws.org/Vol-3274/paper2.pdf>.
- Baier, Thomas et al. (2022). “A modular architecture for creating multimodal agents”. In: *arXiv preprint arXiv:2206.00636*. DOI: <https://doi.org/10.48550/arXiv.2206.00636>.
- Meer, Michiel van der et al. (2022). “Will It Blend? Mixing Training Paradigms & Prompting for Argument Quality Prediction”. In: *Proceedings of the 9th Workshop on Argument Mining*, pp. 95–103. URL: <https://aclanthology.org/2022.argmining-1.8>.
- Santamaria, Selene Baez, Emmanouil Manousogiannis, et al. (2022). “Access to care: analysis of the geographical distribution of healthcare using Linked Open Data”. In: *arXiv preprint arXiv:2204.05206*. DOI: <https://doi.org/10.48550/arXiv.2204.05206>.
- Santamaria, Selene Baez et al. (2022). “Evaluating Agent Interactions Through Episodic Knowledge Graphs”. In: *Proceedings of the 1st Workshop on Customized Chat Grounding Persona and Knowledge*, pp. 15–28. URL: <https://aclanthology.org/2022.ccgpk-1.3>.
- Boomgaard, Guusje et al. (2021). “Learning Profile-Based Recommendations for Medical Search Auto-Complete”. In: URL: <http://ceur-ws.org/Vol-2846/paper34.pdf>.
- Santamaria, Selene Baez, Thomas Baier, et al. (2021). “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)*, pp. 56–77. URL: <https://aclanthology.org/2021.mmsr-1.6>.
- Godinez, Erick Velazquez et al. (2020). “Language Identification for Short Medical Texts.” In: *HEALTH-INF*, pp. 399–406. DOI: <https://doi.org/10.5220/0008950903990406>.
- Manousogiannis, Emmanouil, Sepideh Mesbah, Alessandro Bozzon, Robert-Jan Sips, et al. (2020). “Normalization of Long-tail Adverse Drug Reactions in Social Media”. In: *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*, pp. 49–58. DOI: [10.18653/v1/2020.louhi-1.6](https://doi.org/10.18653/v1/2020.louhi-1.6).
- Manousogiannis, Emmanouil, Sepideh Mesbah, Alessandro Bozzon, Selene Baez, et al. (2019). “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*, pp. 114–116. DOI: [10.18653/v1/W19-3219](https://doi.org/10.18653/v1/W19-3219).
- Vossen, Piek, Selene Baez, Lenka Bajcetic, et al. (2019). “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, pp. 181–184. URL: <https://ceur-ws.org/Vol-2456/paper47.pdf>.
- Vossen, Piek, Lenka Bajcetić, et al. (2019). “Modelling context awareness for a situated semantic agent”. In: *International and Interdisciplinary Conference on Modeling and Using Context*. Springer, pp. 238–252. DOI: [https://doi.org/10.1007/978-3-030-34974-5\\_20](https://doi.org/10.1007/978-3-030-34974-5_20).
- Vossen, Piek, Selene Baez Santamaria, et al. (2019). “A communicative robot to learn about us and the world”. In: *2019 Annual International Conference on Computational Linguistics and Intellectual Technologies, Dialogue 2019*, pp. 728–743. URL: [http://www.dialog-21.ru/media/4872/\\_-dialog2019scopus\\_rev2.pdf](http://www.dialog-21.ru/media/4872/_-dialog2019scopus_rev2.pdf).

- Liang, Quan et al. (2018). “Individual travel behavior modeling of public transport passenger based on graph construction”. In: *Journal of Advanced Transportation* 2018. DOI: [10.1155/2018/3859830](https://doi.org/10.1155/2018/3859830).
- Vossen, Piek, Selene Baez, Lenka Bajčetić, et al. (2018). “Leolani: a reference machine with a theory of mind for social communication”. In: *International conference on text, speech, and dialogue*. Springer, pp. 15–25. DOI: [https://doi.org/10.1007/978-3-030-00794-2\\_2](https://doi.org/10.1007/978-3-030-00794-2_2).
- Baez, Selene (2015). “Predicting opponent team activity in a RoboCup environment”. In: *arXiv preprint arXiv:1503.01446*. DOI: [10.48550/ARXIV.1503.01446](https://doi.org/10.48550/ARXIV.1503.01446). URL: <https://arxiv.org/abs/1503.01446>.