

GEORGETA-CRISTINA GÂRBACEA

5735 S Ellis Ave, Chicago, Illinois, 60637

garbacea@uchicago.edu, <https://cristinagarbacea.com>

RESEARCH INTERESTS

My research interests are focused on machine learning and deep learning for natural language processing. I am particularly interested in the alignment, interpretability and control of large language models (LLMs) and their evaluation.

ACADEMIC POSITIONS

University of Chicago

Postdoctoral Scholar, Data Science Institute

Sept 2023 - present

Chicago, IL

EDUCATION

University of Michigan, Ann Arbor, USA

Sep 2016 - Aug 2023

Ph.D. in Computer Science & Engineering

Advisor: Prof. Qiaozhu Mei

PhD Thesis Committee:

Prof. Joyce Chai, Prof. Emily Mower Provost, Prof. Kevyn Collins-Thompson, Prof. Lu Wang

University of Amsterdam, The Netherlands

Sep 2013 - May 2016

M.Sc. in Artificial Intelligence, graduated “*cum laude*”

Advisors: Prof. Evangelos Kanoulas, Prof. Maarten de Rijke

Transilvania University of Brasov, Romania

Sep 2006 - July 2010

B.Eng. in Electrical Engineering and Computers

B.Sc. in Computer Science

National College of Computer Science “Grigore Moisil”, Romania

Sep 2002 - July 2006

ACADEMIC RESEARCH EXPERIENCE

Data Science Institute, University of Chicago

Sept 2023 - present

Postdoctoral Scholar

Chicago, IL

Advisors: Prof. Chenhao Tan, Prof. Victor Veitch

Conducting research on the alignment, interpretability and control of large language models (LLMs).

Foreseer Group, University of Michigan

2016 - Aug 2023

Graduate Student Research Assistant

Ann Arbor, MI

Advisor: Prof. Qiaozhu Mei

Conducting research on deep unsupervised generative models for natural language and their evaluation.

Information and Language Processing (ILPS), University of Amsterdam

2013 - 2016

Graduate Student Research Assistant

Ann Arbor, MI

Advisors: Prof. Evangelos Kanoulas, Prof. Maarten de Rijke

Conducting research on text summarization, sentiment analysis, entity extraction and linking.

Amsterdam Data Science Research Center, University of Amsterdam

2013 - 2016

Graduate Student Research Assistant

Amsterdam, NL

Advisors: Daan Odijk, Victor de Boer, Laura Hollink, Jacco van Ossenbruggen

We published a paper in TPD L 2015 on “Quantifying Historical Perspectives on World War II”.

INDUSTRY RESEARCH EXPERIENCE

Microsoft Research

Research Intern

Advisors: Tong Wang, Alessandro Sordoni

Mitigating the few-shot sensitivity of pre-trained language models.

*June 2022 - August 2022
(remote) Montreal, Canada*

Google Deepmind

Research Scientist Intern

Advisors: Thomas C Walters, Aäron van den Oord

Practical low-rate speech coding with VQ-VAE, WaveRNN and dynamic arithmetic coding.

*August 2020 - December 2020
(remote) London, UK*

Google Deepmind

Research Scientist Intern

Advisors: Thomas C Walters, Aäron van den Oord, Oriol Vinyals

We published a paper at ICASSP 2019 on speech compression using VQ-VAE. Patent approved.

*May 2018 - September 2018
London, UK*

Microsoft Research

Research Intern

Advisors: Paul Bennett, Fernando Diaz

I joined the CLUES research group on the project “*Augmented Memory Intelligence*”, and worked on temporal and contextual retrieval of information from personal information stores.

*May 2016 - August 2016
Redmond, USA*

Microsoft Research

“Bright Mind” Research Intern

Advisors: Fahad Dogar, Thomas Karagiannis, Antony Rowstron

I joined the Systems and Networking research group on the project “*Analyzing OneDrive Performance*”. I analyzed how OneDrive user requests travel through the data serving stack, and how different cache sizes and eviction algorithms affect the performance and latency of the cloud service.

*June 2014 - Sept 2014
Cambridge, UK*

PUBLICATIONS

- **PhD Thesis**

Neural Language Generation for Content Adaptation: Explainable, Efficient Low-Resource Text Simplification and Evaluation

Cristina Gârbacea, PhD dissertation, University of Michigan, 2023.

<https://deepblue.lib.umich.edu/handle/2027.42/178028>

- **Publications**

Accept or Resist? A Systematic Comparison of Online and Offline Alignment for Persuasion-Balanced Training

Justin Hwang*, Cristina Gârbacea*, Sohini Banerjee, Chenhao Tan (*submitted*)

Personalized Benchmarking: Evaluating LLMs by Individual Preferences

Cristina Gârbacea*, Heran Wang*, Chenhao Tan (*submitted*)

HyPerAlign: Interpretable Personalized LLM Alignment via Hypothesis Generation,

Cristina Gârbacea, Chenhao Tan <https://arxiv.org/pdf/2505.00038> (*submitted*)

Evaluating the Goal-Directedness of Large Language Models, Tom Everitt*, Cristina Gârbacea*, Alexis Bellot, Jonathan Richens, Henry Papadatos, Siméon Campos, Rohin Shah

<https://arxiv.org/pdf/2504.11844> (*submitted*)

RATE: Causal Explainability of Reward Models with Imperfect Counterfactuals

David Reber, Sean Richardson, Todd Nief, Cristina Garbacea, Victor Veitch, in Proceedings In Forty-second International Conference on Machine Learning (ICML), 2025

Why is Constrained Neural Language Generation Challenging?

Cristina Gârbaça and Qiaozhu Mei, published in Transactions of Machine Learning Research (TMLR), 2025.

BoNBoN Alignment for Large Language Models: on the Sweetness of Best-of-N Sampling, Lin Gui, **Cristina Gârbaça** and Victor Veitch, Advances in Neural Information Processing Systems (NeurIPS), 2024: 2851-2885.

What's your Use Case? A Taxonomy of Causal Evaluations of Post-hoc Interpretability (NeurIPS 2023 Workshop on Causal Representation Learning)

David Reber, **Cristina Gârbaça** and Victor Veitch

Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models, Srivastava et al, published in Transactions of Machine Learning Research (TMLR) 2023, <https://arxiv.org/pdf/2206.04615.pdf>

Adapting Pre-trained Language Models to Low-Resource Text Simplification: The Path Matters, **Cristina Gârbaça** and Qiaozhu Mei

In Proceedings of 1st Conference on Lifelong Learning Agents (CoLLAs), Montreal, Canada, 2022, <https://proceedings.mlr.press/v199/garbacea22a/garbacea22a.pdf>

Neural Language Generation: Formulation, Methods, and Evaluation

Cristina Gârbaça and Qiaozhu Mei, <https://arxiv.org/abs/2007.15780.pdf>

GEMv2: Multilingual NLG Benchmarking in a Single Line of Code,

Sebastian Gehrmann et al, <https://arxiv.org/pdf/2206.11249.pdf>

In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing

Explainable Prediction of Text Complexity: The Missing Preliminaries for Text Simplification, **Cristina Gârbaça**, Mengtian Guo, Samuel Carton and Qiaozhu Mei

In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL) and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 1086-1097)., <https://aclanthology.org/2021.acl-long.88/>

The GEM Benchmark: Natural Language Generation, its Evaluation and Metrics

Sebastian Gehrmann et al, <https://arxiv.org/pdf/2102.01672.pdf>

In Proceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021) (pp. 96-120). Association for Computational Linguistics.

UMSIForeseer at SemEval-2020 Task 11: Propaganda detection by fine-tuning BERT with resampling and ensemble learning

Yunzhe Jiang, **Cristina Gârbaça**, Qiaozhu Mei

In Proceedings of the 14th Workshop on Semantic Evaluation at COLING 2020 (pp. 1841-1846)

Judge the Judges: A Large-Scale Evaluation Study of Neural Language Models for Online Review Generation

Cristina Gârbaça, Samuel Carton, Shiyang Yan and Qiaozhu Mei

In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 3959-3972) <https://aclanthology.org/D19-1409/>

Low Bit-Rate Speech Coding with VQ-VAE and a Wavenet Decoder

Cristina Gârbaça, Aäron van den Oord, Yazhe Li, Felicia SC Lim, Alejandro Luebs, Oriol Vinyals and Thomas C Walters, *In International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 735-739, 2019.

<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8683277>

A Systematic Analysis of Sentence Update Detection for Temporal Summarization

Cristina Gârbasea and Evangelos Kanoulas

In *Proceedings of the 39th Conference on Information Retrieval (ECIR 2017)*, Aberdeen, UK.

The University of Amsterdam (ILPS.UvA) at TREC 2015 Temporal Summarization

Cristina Gârbasea and Evangelos Kanoulas

In *Proceedings of the 24th Text REtrieval Conference (TREC 2015)*, NIST, Maryland, USA.

Supporting Exploration of Historical Perspectives Across Collections

D. Odijk, **C. Gârbasea**, T. Schoegje, L. Hollink, V. de Boer, K. Ribbens, and J. van Ossenbruggen

In *Research and Advanced Technology for Digital Libraries (TPDL 2015)*, pp. 238-251, Poland

Combining Multiple Signals for Semanticizing Tweets: U. Amsterdam at #Microposts2015

Cristina Gârbasea, Daan Odijk, David Graus, Isaac Sijaranamual, and Maarten de Rijke

In *Proceedings of the 5th Workshop on Making Sense of Microposts, WWW 2015*, Florence, Italy

Feature Selection and Data Sampling Methods for Learning Reputation Dimensions

Cristina Gârbasea, Manos Tsagkias, and Maarten de Rijke

In *Proceedings of Conference and Labs of the Evaluation Forum (CLEF 2014)*, Sheffield, UK

Detecting the Reputation Polarity of Microblog Posts

Cristina Gârbasea, Manos Tsagkias, and Maarten de Rijke

In *European Conference on Artificial Intelligence (ECAI 2014)*, Prague, Czech Republic

PATENTS

Speech Coding Using Latent Content and Speaker Embedding Vectors

Cristina Gârbasea, Aäron Gerard Antonius Van Den Oord, Yazhe Li, Sze Chie Lim, Alejandro Luebs, Oriol Vinyals, Thomas Chadwick Walters

PRESS COVERAGE

Turing Test 2.0, Physics World, May 2021 edition

<https://physicsworld.com/a/the-turing-test-2-0/>

INVITED TALKS

2022: Microsoft Research Efficient Large-Scale AI Workshop

<https://www.microsoft.com/en-us/research/event/efficient-large-scale-ai-workshop/>

AWARDS AND HONOURS

2021: D. E. Shaw Zenith Fellowship, New York, USA

2021: Oxford Machine Learning Summer School (OxML) Scholarship, University of Oxford, UK

2021: Probabilistic AI Summer School (ProbAI) Scholarship, NTNU Trondheim, Norway

2021: Women in Machine Learning (WiML) Award for attending virtual ICLR 2021

2020: Women in Machine Learning (WiML) Award for attending virtual NeurIPS 2020

2020: Student Award for attending virtual EMNLP 2020

2020: University of Michigan College of Engineering Award for Fall 2020 tuition fee

2020: University of Michigan Award for attending virtual Grace Hopper Conference (vGHC) 2020

2020: Student Award for attending virtual ACL 2020

2020: Student Award & Women in Machine Learning (WiML) Award for attending virtual ICML 2020

2019: NSF Women in Machine Learning (WiML) Travel Award at NeurIPS 2019, Vancouver, Canada

2019: Student Travel Award for attending NeurIPS 2019, Vancouver, Canada

2019: Student Travel Award for attending EMNLP 2019, Hong Kong, China

2019: Student Travel Award for attending ICML 2019, Long Beach, California

2019: INRIA Travel Grant for attending PAISS Summer School on Artificial Intelligence, Paris, France
2019: AMII Travel Grant for attending DLRL Summer School at University of Alberta, Canada
2019: Unity Scholarship for attending GamesAI Summer School (5 scholarships awarded overall)
2018: Self-Organizing Conference in Machine Learning (SOCML), Google Toronto, Canada
2018: Vector Institute Travel Grant for DLRL Summer School at University of Toronto, Canada
2018: CRA-W Grad Cohort Scholar, San Francisco, California, USA
2017: Anita Borg Scholar at Grace Hopper Conference (GHC) 2017, Orlando, Florida, USA
2017: Google Student Research Summit, Machine Intelligence track, San Bruno, California, USA
2017: Student Travel Award for attending KDD 2017, Halifax, Nova Scotia, Canada
2016: Rackham Graduate Fellowship, University of Michigan Ann Arbor, USA
2015: Google EMEA "Anita Borg" Memorial Scholarship Finalist
2015: ELIAS Scholarship for attending ESSIR 2015, Thessaloniki, Greece
2014: ACM Travel Grant for attending ECAI 2014, Prague, Czech Republic
2014: Microsoft "Bright Mind" Research Intern, Microsoft Research Cambridge, UK

TEACHING EXPERIENCE

2024: Data Science Clinic Instructor, University of Chicago
2022: Graduate Student Instructor for U. Michigan Master of Applied Data Science (MADS)
2021: Graduate Student Instructor for U. Michigan EECS 492 (Introduction to Artificial Intelligence)
2019: Lecturer at AI4ALL, University of Michigan
2019: Lecturer at Explore Engineering, University of Michigan

Students mentored:

Marie Wang, undergraduate student, University of Chicago
Justin Hwang, undergraduate student, University of Chicago
Sohini Banerjee, undergraduate student, University of Chicago
Karen Zhou, PhD student, University of Chicago
David Reber, PhD student, University of Chicago
Lin Gui, PhD student, University of Chicago
Mengtian Guo, UM CSE undergraduate, now PhD student at University of North Carolina Chapel Hill
Yunzhe Jiang, UM SI master student
Taoan Lu, UM CSE undergraduate

Teaching Training:

2021: Relate Workshop - Making Science Communication a Conversation, University of Michigan
2019, 2020, 2021: Next Prof Science Workshop, University of Michigan
2006-2009: Teacher Qualification Certificate, Transilvania University of Brasov

SUMMER SCHOOLS

2025 Oxford Machine Learning Summer School (OxML), University of Oxford, UK
2024 Oxford Machine Learning Summer School (OxML), University of Oxford, UK
2023 Oxford Machine Learning Summer School (OxML), University of Oxford, UK
2022 Deep Learning and Reinforcement Learning (DLRL), MILA Montreal, Canada
2022 Oxford Healthcare Machine Learning Summer School (OxML), University of Oxford, UK
2022 Oxford Finance Machine Learning Summer School (OxML), University of Oxford, UK
2021 Oxford Machine Learning Summer School (OxML), University of Oxford, UK
2021 Deep Learning and Reinforcement Learning (DLRL), MILA Montreal, Canada
2021 Quantum Machine Learning Summer School (Qiskit), IBM Quantum
2021 Deep Learning and Computational Neuroscience, Neuromatch Academy
2021 Eastern European Machine Learning Summer School (EEML), Budapest, Hungary
2021 Nordic Probabilistic AI Summer School (ProbAI), NTNU Trondheim, Norway
2021 Mexican NLP Summer School at NAACL 2021, Mexico City, Mexico

2021 Mediterranean Machine Learning Summer School (M2L), Milano, Italy
2020 Deep Learning and Reinforcement Learning (DLRL), MILA Montreal, Canada
2020 Brain, Minds and Machines (BMM) Summer Course, CBMM, MIT CSAIL, USA
2020 Regularization Methods for Machine Learning (RegML), MIT and University of Genova
2020 Machine Learning Summer School (MLSS), Telkom University, Indonesia
2019 PAISS Summer School on Artificial Intelligence, INRIA Paris, France
2019 Deep Learning and Reinforcement Learning (DLRL), University of Alberta and AMII, Canada
2019 Artificial Intelligence and Games (GameAI), New York University, New York, USA
2018 Deep Learning and Reinforcement Learning, University of Toronto and Vector Institute
2015 European Summer School in Information Retrieval, Thessaloniki, Greece
2011 Functional Programming, Utrecht University, The Netherlands

WORKSHOPS

2021 Stanford Compression Workshop, University of Stanford
2021 Deep Learning and Combinatorial Optimization, IPAM, University of California Los Angeles
2020 Deep Reinforcement Learning, Simons Institute, University of California Berkeley
2020 Theory of Reinforcement Learning, Simons Institute, University of California Berkeley

PROFESSIONAL EXPERIENCE

2010 - 2013: Software Development Engineer, US Outsourcing, Brasov, Romania
2009: Software Development Engineer Intern, Dynamic Ventures, Brasov, Romania
2008: Software Development Engineer Intern, iQuest Technologies, Brasov, Romania

TECHNICAL STRENGTHS

Computer Languages	Python, C#, Java, C/C++, MATLAB
Frameworks	Tensorflow, PyTorch, Keras, Scikit-Learn, Matplotlib
Databases	MSSQL, MySQL, MSAccess, ElasticSearch
Tools	Visual Studio, Git, Latex
Platforms	Linux, Mac OS X, Windows

PROFESSIONAL SERVICE

Reviewer: NeurIPS 2025, NeurIPS 2024, ACL 2024, COLM 2024, ICLR 2024, EMNLP 2023, ACL 2023, NeurIPS 2023, IJCAI 2023, KDD 2023, JMLR 2022, NeurIPS2022, ICLR 2022, ACL 2022, ACL 2021, AAAI 2021, INLG 2021, COLING 2020, WiML NeurIPS 2019

Volunteer: AISTATS 2021, NAACL 2021, ICLR 2020, ICML 2020, SIGIR 2018, KDD 2017

Program Committee:

3rd ACL Workshop on Advances in Language and Vision Research (ALVR) at ACL 2023
Program Committee Member at the WebConf (WWW) 2023
2nd Worskhop on Generation, Evaluation and Metrics (GEM) at EMNLP 2022
1st Worskhop on Generation, Evaluation and Metrics (GEM) at ACL 2021
CtrlGen: Disentanglement and Controllable Generation for Vision and Language at NeurIPS 2021
2nd Workshop on Advances in Language and Vision Research at NAACL 2021
14th International Conference on Natural Language Generation (INLG) 2021

Community Service:

2021-present: ACM Transactions of the Web (TWeb) Information Director
2021-present: ACL Rolling Review Action Editor
2021-2023: Michigan Institute for Data Science (MIDAS) Student Leadership Board
2018: Webmaster for 41st International ACM SIGIR Conference on Research and Development in

Information Retrieval (SIGIR 2018), <https://sigir.org/sigir2018/>
2016-2018: Graduate Rackham International (GRIN) Student Leadership Board

LANGUAGES SPOKEN

English (advanced), French (Advanced), Italian (intermediate), Spanish (intermediate),
Dutch (intermediate), Romanian (native)