evgeniiaraz.github.io
evgeniiarazum@gmail.com;
er563@cam.ac.uk

Evgeniia Razumovskaia

- Final year PhD student at the University of Cambridge researching multilingual NLP focusing on dialogue systems, factuality and data collection. Previously – research-based MSc at Carnegie Mellon University and BSc at the University of Edinburgh.
- Eight years of experience in NLP research, with almost two years in industry with internships at Google Research and the strongest startups in the area – Rasa and PolyAI.
- Eight first/co-first authored papers at NAACL, EMNLP, ACL, COLING, TACL and JAIR

EDUCATION

PhD in Computation, Cognition and Language, University of Cambridge   2020–now
Supervisor: ANNA KORHONEN. Co-supervisor: IVAN VULIĆ.
Research topic: *Multilingual dialogue systems in low data regimes*

> I am working on extending dialogue systems to new languages in low-resource scenarios.
> My work focuses on dialogue Natural Language Understanding and uses varied techniques, such as contrastive learning and weight alignment, to effectively use scarce annotated resources. In addition, I have conducted research in data collection for different tasks and languages.

MSc in Language Technologies, Carnegie Mellon University (GPA: 3.98; *sup.:* MAXINE ESKENAZI)   2018–2019
Research Topic: *Hybrid Dialogue Systems Combining End-to-End Systems with Rules*

> Projects: End-to-end dialog systems and hybrid systems with rules; human evaluation via crowdsourcing

BSc (Hons) Cognitive Science (Informatics), University of Edinburgh (1ˢᵗ degree; *sup.:* SIMON KING)   2014–2018
Thesis title: *Emotion Recognition from Speech*

RESEARCH VISITS AND INTERNSHIPS

Machine Learning Research Intern, Poly AI, London UK   Jun'23 — Oct'23

> *Poly AI is an innovative multilingual conversational AI startup with 200+ published papers and 6 registered patents.*
> Developed a new model for answering Frequently Asked Questions by finetuning Large Language Models, which led to improved factuality and 58% quality improvement in comparison to preceding model.

Research Intern, Google Research, London UK   May'22 — Sep'22

> Project: In-context few-shot multilingual story generation with planning using monolingually pretrained Large Language Models
> Hosts: JOSHUA MAYNEZ, SHASHI NARAYAN, MIRELLA LAPATA, ANNIE LOUIS

Machine Learning Research Intern, Rasa, Berlin, Germany   Jan'20 — Oct'20

> *Rasa is a leading open-source conversational AI framework.*
> Conducted research for end-to-end training for Conversational AI systems while accounting for project-specific rules; developed production level code deployed as an experimental feature in Rasa's main product, Rasa Open Source.

Graduate Research Assistant, Carnegie Mellon University, Pittsburgh, USA   Sep'18 — Dec'19

Summer Research Expedition, Johns Hopkins University, Baltimore, USA

> Project: Text classification based on keywords and topics extracted from the Wikipedia using Gensim
> Supervisor: Benjamin van Durme

Undergraduate Student Researcher, University of Edinburgh, Edinburgh, UK

> Project: Data Collection and Analysis of Usage of Scottish Words on Twitter during the Independence Referendum
> Supervisor: Sharon Goldwater

SELECTED PUBLICATIONS

Dial BeInfo for Faithfulness: Improving Factuality of Information-Seeking Dialogue via Behavioural Fine-Tuning
E. Razumovskaia, I. Vulić, P. Marković, T. Cichy, Q. Zheng, T.-H. Wen, and P. Budzianowski
Findings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)

SQATIN: Supervised Instruction Tuning Meets Question Answering for Improved Dialogue NLU
E. Razumovskaia, G. Glavaš, A. Korhonen, and I. Vulić
Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2024)

Little Red Riding Hood Goes Around the Globe: Crosslingual Story Planning and Generation with Large Language Models
E. Razumovskaia, J. Mainez, A. Louis, M. Lapata, and S. Narayan
Proceedings of the Language Resources and Evaluation Conference (LREC-COLING 2024)

Transfer-Free Data-Efficient Multilingual Slot Labeling
E. Razumovskaia, I. Vulić, and A. Korhonen
Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)

MULTI3NLU++: A Multilingual, Multi-Intent, Multi-Domain Dataset for Natural Language Understanding in Task-Oriented Dialogue
N. Moghe*, **E. Razumovskaia***, L.K. Guillou, I. Vulić, A. Korhonen, and A. Birch
Findings of 61st Meeting of the Association for Computational Linguistics (ACL 2023); 5th NLP4ConvAI Workshop

Cross-Lingual Dialogue Dataset Creation via Outline-Based Generation
O. Majewska*, **E. Razumovskaia***, E. M. Ponti, I. Vulić, and A. Korhonen
Transactions of Association for Computational Linguistics (TACL): 11, 139 – 156. (presented at EACL 2023)

Data Augmentation and Learned Layer Aggregation for Improved Multilingual Language Understanding in Dialogue
E. Razumovskaia, I. Vulić, and A. Korhonen
Findings of 60th Meeting of the Association for Computational Linguistics (ACL 2022)

Natural Language Processing for Multilingual Task-Oriented Dialogue
E. Razumovskaia, G. Glavaš, O. Majewska, E. M. Ponti and I. Vulić
Proceedings of 60th Meeting of the Association for Computational Linguistics (ACL 2022): *Tutorial Abstracts*

Crossing the Conversational Chasm: A Primer on Multilingual Task-Oriented Dialogue Systems
E. Razumovskaia, G. Glavaš, O. Majewska, E. M. Ponti , A. Korhonen, and I. Vulić
Journal of Artificial Intelligence Research (JAIR): 74 (2022), 1351-1402

Incorporating Rules into End-to-End Dialog Systems
E. Razumovskaia, and M. Eskenazi
3rd Conversational AI Workshop @ NeurIPS 2019

Pretraining Methods for Dialog Context Representation Learning
S. Mehri*, **E. Razumovskaia ***, T. Zhao and M. Eskenazi
Proceedings of 57th Meeting of the Association for Computational Linguistics (ACL 2019)

Beyond Turing: Intelligent Agents Centered on the User
M. Eskenazi, S. Mehri, **E. Razumovskaia**, and T. Zhao
arXiv preprint arXiv:1901.06613

AWARDS AND SCHOLARSHIPS

Nominated for Google PhD Scholarship and Apple AI/ML Scholars 2023 (one of 3 nominated by University of Cambridge for each Scholarship)	2022
Perry Travel Award, *Queens' College, University of Cambridge (to attend ACL'22)*	2022
Carnegie PhD Scholarship, *University of Edinburgh* (declined)	2020
Class Prize for Top Performance in BSc (Hons) Cognitive Science, *University of Edinburgh*	2018
Edinburgh Award, *University of Edinburgh*	2017

TEACHING

COURSES

Computational Linguistics, *Supervisor, University of Cambridge*	2021—2023
Deep Reinforcement Learning, *Teaching Assistant, Carnegie Mellon University*	2019
Introductory Applied Machine Learning, *Tutor, Lab demonstrator, University of Edinburgh*	2017—2018
Formal and Natural Language Processing, *Tutor, University of Edinburgh*	2017—2018
Reasoning and Agents, *Tutor, University of Edinburgh*	2016—2017

STUDENT SUPERVISOR

Undergraduate Student: Jadd Virji, *University of Cambridge*

TALKS

Building Dialogue Systems with Scarce Resources, *Invited Talk at **EdinburghNLP, University of Edinburgh***	2023
Natural Language Processing for Multilingual Task-Oriented Dialogue, ***Tutorial at ACL'22***	2022
End-to-end dialogue systems, *Talk at **Rasa Summit***	2021

SKILLS

Prog. Languages: Python, Bash, Matlab

ML Frameworks: PyTorch, Tensorflow, JAX, Keras

NLP Frameworks: HuggingFace toolkit, NLTK, Gensim

Languages: English, Russian German, Italian, Latin, Ancient Greek

SERVICE

Center of Human-Inspired Artificial Intelligence (CHIA): Student Representative, Events on Responsible AI
Reviewer: ACL ARR, EMNLP, ACL, NLP4ConvAI workshop