Evgeniia Razumovskaia

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

I am a PhD candidate working on multilingual Natural Language Processing at the University of Cambridge. My project focuses on sample-efficient techniques to extend dialogue systems to new languages. I also worked on the topic as a research intern at Google AI Research, London. My other interests include factuality and controllability of dialogue systems; during my internship at PolyAI and my Masters research at Carnegie Mellon University, I worked on improving factuality of fully generative systems and combining them with their rule-based counterparts.

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 to improve multilingual capabilities of dialogue systems and other user-facing languages technologies, more generally.

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 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 Jan'16 — May'16

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

PUBLICATIONS

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; to appear)

MULTI3NLU++: A Multilingual, Multi-Intent, Multi-Domain Dataset for Natural Language Understanding in Task-Oriented Dialogue
N. Moghe, **E. Razumovskaia***, Guillou, L. K., Vulić, I., Korhonen, A., and Birch, A.
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)

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
arXiv preprint arXiv: 2212.10471. Work done while interning at Google Research.

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

Evaluation of Difficulty and Complexity of Tasks: Case Study of International On-line Competition "Beaver''
E. Yagunova, S. Pozdniakov, N. Ryzhova, **E. Razumovskaia**, and N. Korovkin
International Journal of Engineering Education: 32(3), 1141-1150

Tasks Classification and Age Differences in Task Perception. Case Study of International On-line Competition "Beaver"
E. Yagunova, S. Pozdniakov, N. Ryzhova, **E. Razumovskaia**, and N. Korovkin
Proceedings of International Conference on Informatics in Schools: Situation, Evolution and Perspectives (ISSEP 2015)

AWARDS AND SCHOLARSHIPS

Nominated for Google PhD Scholarship and Apple AI/ML Scholars 2023   2022
(one of 3 nominated by University of Cambridge for each Scholarship)

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   NLP Frameworks: HuggingFace toolkit, NLTK, Gensim

ML Frameworks: PyTorch, Tensorflow, JAX, Keras   Languages: English, Russian German, Italian, Latin, Ancient Greek

SERVICE

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