Omri Uzan

omriuz.github.io
uzan.omri@gmail.com

Research Interest

I am broadly interested in understanding Language Models, their potential capabilities, inherent limitations, and future social implications. I'm currently focused on exploring the limitations imposed on LLMs by their foundational word representation schemes, particularly in the context of languages with diverse linguistic properties or limited data availability.

Education

Department of Computer Science, Ben Gurion University — Israel

M.Sc. in Computer Science, Specialization in ML — 2023 - 2024 *(expected)*

- GPA (current): 3.74
- Advisor: Yuval Pinter
- Thesis: Enhancing Subword Tokenization Schemes and Their Evaluation Methods in Large Language Models

B.Sc. in Computer Science, Manga Cum Laude — 2020 - 2023

- GPA: 3.71
- Ranked first in class in the NLP and Applied ML courses, amongst both M.Sc. and B.Sc. students

Awards and Honors

- **Outstanding Paper Award as first author at ACL 2024**, awarded to 35 of 1915 accepted papers
- **Senior area chair Paper Award as first author at ACL 2024**, awarded to 21 of 1915 accepted papers
- **The Dean's award for outstanding honor students** 2024.06
- **Won 2nd place with my team at Samsung Next Gen AI Hackathon** 2023.08
- **The Head of Department award for honor students** 2023.06
- **'Dkalim' Honorary programs participant**, early research training during B.Sc.
- **'Ashalim' Honorary programs participant,** extra multidisciplinary studies during B.Sc.

Experience

Meta | Full Time, Tel Aviv, Israel — 2024.02 - present

- Software Engineer, User Routing Systems

Ben Gurion University | Beersheba, Israel — 2023.10 - present

- Research Assistant, MeLeL Lab (Yuval Pinter)

Ben Gurion University | Beersheba, Israel — 2023.10 - 2024.02

- Teaching Assistant, Introduction to Computer Science - 202-1-1011

Meta | Full Time, Tel Aviv, Israel — 2022.07 - 2022.10

- Software Engineer Intern, Facebook Operators Solutions

Nitzanim.tech (NGO) | Beersheba, Israel — 2021.08 - 2023.12

- Mentored a group of 15 underrepresented youths for two years through bi-weekly meetings, supporting their pathways into the tech industry

Publications

1. Khuyagbaatar Batsuren, Ekaterina Vylomova, Verna Dankers, Tsetsuukhei Delgerbaatar, **Omri Uzan**, Yuval Pinter, Gábor Bella. *Evaluating Subword Tokenization: Alien Subword Composition and OOV Generalization Challenge. Under Review*,
2. Craig W Schmidt, Varshini Reddy, Haoran Zhang, Alec Alameddine, **Omri Uzan**, Yuval Pinter, Chris Tanner. *Tokenization Is More Than Compression.* **EMNLP 2024**
3. **Omri Uzan**, Craig W.Schimdt, Chris Tanner, Yuval Pinter. *Greed is All You Need: An Evaluation of Tokenizer Inference Methods.* **ACL 2024, Outstanding Paper Award, Area Chair Paper Award, Presented as Oral talk.**

Technical Skills

Languages: *Python, SQL, PHP, JavaScript, Java, TypeScript, C++, HTML, CSS.*

Frameworks and Libraries: *Hugging Face, PyTorch, Scikit-Learn, NumPy, Pandas, Matplotlib, Seaborn, Tensorflow.*

Developer Tools: *Git, Linux, JetBrains IDE, VS Code, Jupyter Lab, Google Cloud Platform.*