

Yftah Ziser

ASSISTANT PROFESSOR · UNIVERSITY OF GRONINGEN · SENIOR RESEARCH SCIENTIST · NVIDIA RESEARCH

Bernoulli Institute, University of Groningen, Nijenborgh 9, 9747 AG Groningen, The Netherlands

✉ yftah89@gmail.com | 🏠 <https://yftah89.github.io/>

Education

Technion – Israel Institute of Technology Haifa, Israel
Ph.D. (Direct track) 2015 - 2019

- Supervised by Prof. Roi Reichart.
- Thesis: "Domain Adaptation for Natural Language Processing - a Neural Network Based Approach."
- My Ph.D. research goal is to adapt models from domains and languages rich in labeled training data to domains and languages that are poor in such data through representation learning.
- Outstanding Ph.D. Student Award, the DDS faculty.

Technion – Israel Institute of Technology Haifa, Israel
B.Sc. 2011 - 2015

- Bachelor of Science in Computer Science

Professional Experience

University of Groningen Groningen, the Netherlands
Assitant Professor 2025 - present

- A tenure-track assistant professor (research focused).

Nvidia Research Tel-Aviv, Israel (remotely)
Senior Researcher 2024 - present

- Working mainly on LLM alignment and low-resource NLP.
- Leading multiple research internships.

University of Edinburgh Edinburgh, United Kingdom
Postdoctoral Researcher 2021 - 2024

- Hosted by Dr. Shay Cohen.

Facbook Tel-Aviv, Israel
Researcher 2020 - 2021

- Designing state-of-the-art NLP solutions for e-commerce.

Amazon Alexa Haifa, Israel
Researcher 2019 - 2020

- Designing state-of-the-art question-answering systems using Amazon community data.
- Leading multiple research internships.

IBM Haifa, Israel
Part-time researcher 2017 - 2018

- Designing and building machine learning algorithms for natural language problems.
- Designing and teaching an internal "Introduction to Deep Learning" course for IBM colleagues.

Publications

Beyond Token Probes: Hallucination Detection via Activation Tensors with ACT-ViT, Guy Bar-Shalom, Fabrizio Frasca, Yaniv Galron, **Yftah Ziser** and Haggai Maron. To appear in NeurIPS 2025.

Policy Optimized Text-to-Image Pipeline Design, Uri Gadot, Rinon Gal, **Yftah Ziser**, Gal Chechik, Shie Mannor, To appear in NeurIPS 2025.

Iterative Multilingual Spectral Attribute Erasure, Shun Shao, **Yftah Ziser**, Zheng Zhao, Yifu QIU, Shay B Cohen, Anna Korhonen, To appear in EMNLP 2025.

A Simple Yet Effective Method for Non-Refusing Context Relevant Fine-grained Safety Steering in LLMs, Shaona Ghosh, Amrita Bhattacharjee, **Yftah Ziser**, Christopher Parisien, To appear in EMNLP 2025.

Knowing Before Saying: LLM Representations Encode Information About Chain-of-Thought Success Before Completion, Anum Afzal, Florian Matthes, Gal Chechik and **Yftah Ziser**, findings of ACL 2025.

TSPRank: Bridging Pairwise and Listwise Methods with a Bilinear Travelling Salesman Model, Waylon Li, **Yftah Ziser**, Yifei Xie, Shay B Cohen and Tiejun Ma, KDD 2025 (Research Track).

Spectral Editing of Activations for Large Language Model Alignment, Yifu Qiu, Zheng Zhao, **Yftah Ziser**, Anna Korhonen, Edoardo Ponti and Shay Cohen, NeurIPS 2024.

Layer by Layer: Uncovering Where Multi-Task Learning Happens in Instruction-Tuned Large Language Models, Zheng Zhao, **Yftah Ziser** and Shay Cohen, EMNLP 2024.

Are Large Language Models Temporally Grounded?, Yifu Qiu, Zheng Zhao, **Yftah Ziser**, Anna Korhonen, Edoardo Ponti and Shay Cohen, NAACL 2024.

Detecting and Mitigating Hallucinations in Multilingual Summarisation, Yifu Qiu, **Yftah Ziser**, Anna Korhonen, Edoardo Ponti and Shay Cohen, EMNLP 2023.

A Joint Matrix Factorization Analysis of Multilingual Representations, Zheng Zhao, **Yftah Ziser**, Bonnie Webber and Shay Cohen, findings of EMNLP 2023.

Rant or Rave: Variation over Time in the Language of Online Reviews, **Yftah Ziser**, Bonnie Webber and Shay Cohen, LRE 2023.

Erasure of Unaligned Attributes from Neural Representations, Shun Shao*, **Yftah Ziser***, and Shay Cohen, TACL 2023.

BERT Is Not The Count: Learning to Match Mathematical Statements with Proofs, Waylon Li, **Yftah Ziser**, Maximin Coavoux, and Shay Cohen, EACL 2023.

Gold Doesn't Always Glitter: Spectral Removal of Linear and Nonlinear Guarded Attribute Information, Shun Shao, **Yftah Ziser**, and Shay Cohen, EACL 2023.

Factorizing Content and Budget Decisions in Abstractive Summarization of Long Documents, Marcio Fonseca, **Yftah Ziser**, and Shay Cohen, EMNLP 2022.

Understanding Domain Learning in Language Models Through Subpopulation Analysis, Zheng Zhao, **Yftah Ziser**, and Shay Cohen, BlackboxNLP@EMNLP 2022.

Customized Pre-Training for Improved Domain Adaptation with Category Shift, Tony Lechtman, **Yftah Ziser**, and Roi Reichart, EMNLP 2021.

WikiSum: Coherent Summarization Dataset for Efficient Human-Evaluation, Nachshon Cohen*, Oren Kalinsky*, **Yftah Ziser***, and Alessandro Moschitti, ACL 2021.

Answering Product-Questions by Utilizing Questions from Other Contextually Similar Products, Ohad Rozen, David Carmel, Avihai Mejer, Vitali Mirkis, and **Yftah Ziser**, NAACL 2021.

Humor Detection in Product Question Answering Systems, **Yftah Ziser**, Elad Kravi, and David Carmel, SIGIR 2020.

Task Refinement Learning for Improved Accuracy and Stability of Unsupervised Domain Adaptation, **Yftah Ziser**, and Roi Reichart, ACL 2019.

Deep Pivot-Based Modeling for Cross-language Cross-domain Transfer with Minimal Guidance, **Yftah Ziser**, and Roi Reichart, EMNLP 2018.

Pivot Based Language Modeling for Improved Neural Domain Adaptation, **Yftah Ziser** and Roi Reichart, NAACL 2018.

Neural Structural Correspondence Learning for Domain Adaptation, **Yftah Ziser** and Roi Reichart, CoNLL 2017.

Teaching Experience

2019	Yearly Project in Data Science,	Project Mentor.
Fall 2018	Natural Language Processing for the Social Sciences,	Project Mentor.
Spring 2017	Learning Robust Representations for Natural Language Processing,	Project Mentor.
Fall 2017	Introduction to Natural Language Processing,	Teaching Assistant in Charge.
2015-2016	Introduction to Computer Science,	Teaching Assistant.

Mentoring

2022-present	Yifu Qiu,	Ph.D. student,	University of Edinburgh.
2021-2024	Marcio Fonseca,	Ph.D. student,	University of Edinburgh.
2021-present	Zheng Zhao,	Ph.D. student,	University of Edinburgh.
2021-2023	Shun Shao,	M.Sc. student,	University of Edinburgh.
2021-2023	Waylon Li,	M.Sc. student,	University of Edinburgh.
2020-2021	Entony Lekhtman,	M.Sc. student,	Technion.
Summer 2020	Ohad Rozen,	Ph.D. student,	Bar-Ilan University (Summer internship at Amazon).
Summer 2020	Eilon Sheetrit,	Ph.D. student,	Technion (Summer internship at Amazon).
Summer 2020	Ido Hakimi,	Ph.D. student,	Technion (Summer internship at Amazon).

Academic Service

Invited Talks

2025 *Knowing Before Saying: LLM Representations Encode Information About Chain-of-Thought Success Before Completion*. Nvidia Research, "System 2" seminar, Santa Clara, USA.

2023 *Navigating Distribution Shifts in Natural Language Processing*. IBM Research, Haifa, Israel.

2023 *Navigating Distribution Shifts in Natural Language Processing*. Gong Research Center, Tel-Aviv, Israel.

2022 *Guarded Information Removal for Better Debiasing*. Language Technology Lab group, University of Cambridge, Cambridge, United Kingdom.

2022 *Guarded Information Removal for Better Debiasing*. Department for Computational Linguistics, Heidelberg University, Heidelberg, Germany.

2020 *Domain Bias in Humor Detection*. Amazon Research Center, Tel-Aviv, Israel.

2018 *Cross-language Cross-domain Transfer with Minimal Guidance*. Israel Seminar on Computational Linguistics, Haifa, Israel.

2018 *Contextualized Representations for Domain Adaptation*. Computer Science department, Tel-Aviv University, Tel-Aviv, Israel.

2017 *Harnessing the Power of Neural Networks for Domain Adaptation*. Technion Computational Data Science seminar, Haifa, Israel.

Organization

2021 Co-Organizer of the *Domain Adaptation for NLP Workshop*. At EACL, virtual event.

2018 Committee member of the *Seminar of Computational Linguistics (ISCOL)*. Haifa, Israel.

Reviewing

2018-2021 Reviewing for main NLP venues, including NNACL, ACL, EMNLP, EACL and CoNLL.

2023-present Reviewing for TACL.

2023 Area Chair, Efficient/Low-resource Methods in NLP, EACL 2024.

2021-2024 Reviewing for ACL Rolling Review (ARR).

2024–present Action Editor for ACL Rolling Review (ARR).

Ph.D. Admissions

2021-present Evaluating candidates for the European Laboratory for Learning and Intelligent Systems (ELLIS) Ph.D. program.

Community Engagement

Yotzeem Leshinuy	Jerusalem, Israel
Lecturer	2020 - 2021

- Designing and instructing a friendly programming course for former members of the Haredi Jewish community, who are among the most disadvantaged groups in Israel.

Educating for Excellence	Acre , Israel
Tutor	2019 - 2020

- Teaching underprivileged children programming and math.