

Hosein Mohebbi

✉ h.mohebbi@tilburguniversity.edu | **in** hosein-mohebbi | **tw** hmohebbi75 | **gh** hmohebbi.github.io
☎ +31633928958 | Tilburg, The Netherlands

Education & Research

Tilburg University | Ph.D. candidate

2021 – 2025 (expected)

I'm part of the consortium project [InDeep: Interpreting Deep Learning Models for Text and Sound](#)

Supervisors: Afra Alishahi (Tilburg University), Willem Zuidema (University of Amsterdam), and Grzegorz Chrupała (Tilburg University)

University of Edinburgh | Visiting researcher

Jan – Apr 2024

Visited ILCC, School of Informatics, University of Edinburgh, worked with Ivan Titov

Iran University of Science and Technology | Master's in Artificial Intelligence and Robotics

Sep 2019 – Sep 2021

Ranked as top-four students

Master's Thesis: "Interpretability and Transferability of Linguistic Knowledge in Pre-trained Language Models"

Supervisor: Mohammad Taher Pilehvar

Ferdowsi University of Mashhad | Bachelor's in Computer Engineering

Sep 2014 – Sep 2019

Bachelor's Thesis: "Improving 2D Soccer Agent's Decision Making using Deep Reinforcement Learning"

Supervisor: Ahad Harati

Publications

Disentangling Textual and Acoustic Features of Neural Speech Representations

Under review

Hosein Mohebbi • Grzegorz Chrupała • Willem Zuidema • Afra Alishahi • Ivan Titov | [\[preprint\]](#) [\[video\]](#) [\[code\]](#)

What Do Self-supervised Speech Models Know about Dutch? Analyzing Advantages of Language-specific Pretraining

Interspeech 2025

Marianne de Heer Kloots • Hosein Mohebbi • Charlotte Pouw • Gaofei Shen • Willem Zuidema • Martijn Bentum

On the Reliability of Feature Attribution Methods for Speech Classification

Interspeech 2025

Gaofei Shen • Hosein Mohebbi • Arianna Bisazza • Afra Alishahi • Grzegorz Chrupała

How Language Models Prioritize Contextual Grammatical Cues?

BlackboxNLP 2024

Hamidreza Amirzadeh • Afra Alishahi • Hosein Mohebbi | [\[paper\]](#)

DecoderLens: Layerwise Interpretation of Encoder-Decoder Transformers

Findings of NAACL 2024

Anna Langedijk • Hosein Mohebbi • Gabriele Sarti • Willem Zuidema • Jaap Jumelet | [\[paper\]](#)

Homophone Disambiguation Reveals Patterns of Context Mixing in Speech Transformers

EMNLP 2023

Outstanding Paper Award

Hosein Mohebbi • Grzegorz Chrupała • Willem Zuidema • Afra Alishahi | [\[paper\]](#) [\[code\]](#)

Quantifying Context Mixing in Transformers

EACL 2023

Hosein Mohebbi • Willem Zuidema • Grzegorz Chrupała • Afra Alishahi | [\[paper\]](#) [\[code\]](#) [\[blog\]](#) [\[demo\]](#)

AdapLeR: Speeding up Inference by Adaptive Length Reduction

ACL 2022

Ali Modarressi • Hosein Mohebbi • Mohammad Taher Pilehvar | [\[paper\]](#) [\[blog\]](#)

Not All Models Localize Linguistic Knowledge in the Same Place: A Layer-wise Probing on BERToids' Representations

BlackboxNLP 2021

Mohsen Fayyaz • Ehsan Aghazadeh • Ali Modarressi • Hosein Mohebbi • Mohammad Taher Pilehvar | [\[paper\]](#) [\[blog\]](#)

Exploring the Role of BERT Token Representations to Explain Sentence Probing Results

EMNLP 2021

Hosein Mohebbi • Ali Modarressi • Mohammad Taher Pilehvar | [\[paper\]](#) [\[blog\]](#)

Public Activities

Workshop Organizer

- [BlackboxNLP](#) (co-located with EMNLP 2023, 2024, 2025)

Tutorial Instructor

- Tutorial on “Interpretability Techniques for Speech Models”, at Interspeech 2025 conference, Rotterdam | [\[material\]](#)
- Hands-on tutorial on “NLP & LLMs” for Dutch school students participating in Olympiad in Artificial Intelligence (IOAI'25)
- Tutorial on “Transformer-specific Interpretability”, at EACL 2024 conference, Malta | [\[material\]](#)
- Educational videos on LLMs Interpretability | [\[YouTube\]](#)

Area Chair

- “Interpretability and Analysis of Models for NLP” track for [ACL ARR 2025](#)
- “Speech Recognition, Text-to-Speech and Spoken Language Understanding” track for [ACL ARR 2025](#)

Session Chair

- Speech Processing oral session at ACL 2025, Vienna

Program Committee / Reviewer

- Conferences & Workshops: EMNLP (2022, 2023), ACL 2023, EACL 2023, ARR, Actionable Interpretability 2025

Awards & Honors

- Got an Outstanding Paper Award at EMNLP 2023 Dec 2023
- Ranked 4th in Sharif Data Days National Competition, Sharif University of Technology Mar 2020
- Ranked 5th in RoboCup Asia Pacific International Robotics Competition, 2D Soccer Simulation, Kish Island Dec 2018
- Ranked 1st in Portuguese International Robotics Open (DE ROBOTICA), 2D Soccer Simulation, University of Coimbra Apr 2017
- Ranked 2nd in IranOpen International Robotics Competition, 2D Soccer Simulation League Apr 2017

Teaching & Supervision Experience

Lecturer

- **Machine Learning**, co-lectured with Dr. Grzegorz Chrupala Tilburg University, Winter 2023, 2024
- **Methods for Responsible AI**, co-lectured with Dr. Seyed Mostafa Kia Tilburg University, Winter 2023, 2024

Students Supervised

- Mohsen Fayyaz (BSc), Ehsan Aghazadeh (BSc), Hamidreza Amirzadeh (MS). Topic: Interpretability.

Selected Invited Talks

- Invited talk at CardiffNLP, Cardiff University Feb 2024
- Invited talk at the Centre for Speech Technology Research (CSTR), University of Edinburgh Feb 2024
- Invited talk on at GroNLP, University of Groningen Jun 2023
- Invited talk at Cambridge/Cardiff Workshop in Natural Language Processing Jun 2021
- Gave a guest lecture on Interpretability to a graduate NLP course at Khatam University [\[slides\]](#) May 2021

Programming Skills

- PyTorch • Python • C • C++ • R • Bash Script • HuggingFace • Git

References

Afra Alishahi a.alishahi@tilburguniversity.edu | [\[homepage\]](#)

Professor of Computational Linguistics at the Department of Cognitive Science and Artificial Intelligence at Tilburg University

Willem Zuidema w.h.zuidema@uva.nl | [\[homepage\]](#)

Associate professor at the Institute for Logic, Language and Computation, University of Amsterdam

Grzegorz Chrupala grzegorz@chrupala.me | [\[homepage\]](#)

Associate Professor at the Department of Cognitive Science and Artificial Intelligence at Tilburg University

Mohammad Taher Pilehvar mp792@cam.ac.uk | [\[homepage\]](#)

Senior Lecturer in the School of Computer Science at Cardiff University & Affiliated Lecturer at the University of Cambridge

Ivan Titov ititov@inf.ed.ac.uk | [\[homepage\]](#)

Professor and Personal Chair in Natural Language Processing, at the University of Edinburgh & University of Amsterdam