

Mubashara Akhtar

 mubasharaak |  mubashara-akhtar |  akhtarmubashara |  mubasharaakhtar.com
|  mubashara.akhtar@ai.ethz.ch |

EDUCATION

- PhD in Natural Language Processing, King's College London** Oct 2020 - Dec 2024
- Advisor: Prof. Elena Simperl, Dr. Oana Cocarascu
 - Topic: Multimodal Evidence Integration for Automated Fact-Checking
 - Thesis submitted: July 2024 (defended in Dec 2024)
- Visiting PhD student, University of Cambridge** Jan - June 2023
- Host: Prof. Andreas Vlachos (NLP research group)
- BSc & MSc in Computer Science, TU Wien** Oct 2011 - Jan 2020
- Minor in Economics
 - Grade Master's Degree: 1.6 (comparable to UK first)
 - Grade Bachelor Degree: 2.1 (comparable to UK 2.1)

RECENT WORK EXPERIENCE

- Postdoctoral Researcher & ETH AI Center Fellow, ETH Zurich (Zurich, Switzerland)** since Jan 2025
- Host: Prof. Mrinmaya Sachan, Prof. Elliott Ash, Prof. Menna El-Assady
 - Received the **ETH AI Center Fellowship** (3% acceptance rate)
 - Leading and supervising projects on benchmarking & evaluation, (multimodal) reasoning, and interpretability & VLMs.
- Student Researcher, Google Deepmind (Zurich, Switzerland)** July - Nov 2023 & Jan - April 2024
- Host: Julian Eisenschlos
 - Developed the first open-domain question answering dataset that requires retrieving and reasoning over text, tables, and semi-structured data (paper accepted at TACL).
- Head Teaching Assistant, King's College London (London, UK)** Sept 2021 - Dec 2022
- Course: Python programming for data scientists
 - Co-designed the coursework and lab assignments, supervised and marked programming projects.
- Data Scientist, Mayato AT (Darmstadt, Germany)** Jan 2019 - Sept 2020
- Developed a language model for the first German smart speaker project, **Deutsche Telekom AG**.
 - Technologies: Python (PyTorch, transformers, scikit-learn, nltk, etc.), MS LUIS, etc.
- Java Software Engineer, Ebcont GmbH (Vienna, Austria)** Oct 2017 - Dec 2018
- Developed a fraud detection tool for the Austrian Ministry of Finance.
 - Technologies: Java, OpenShift, Camunda, Spring Boot, Maven, Hibernate, etc.

SELECTED PUBLICATIONS

- When AI Benchmarks Plateau: A Systematic Study of Benchmark Saturation*
Mubashara Akhtar*, A. Reuel*, P. Soni, S. Ahuja, P. Sasanka Ammanamanchi, R. Rawal, V. Zouhar, S. Yadav, et al.
Accepted at **ICML**, 2026.
- Ev2R: Evaluating Evidence Retrieval in Automated Fact-Checking*
Mubashara Akhtar, Michael Schlichtkrull, Andreas Vlachos.
Transactions of the Association for Computational Linguistics (**TACL**), 2026; Presented at **EACL** 2026.
- TANQ: An open domain dataset of table answered questions.*
Mubashara Akhtar*, Chenxi Pang*, Andreea Marzoca, Yasemin Altun, Julian Martin Eisenschlos
Transactions of the Association for Computational Linguistics (**TACL**), 2025; Presented at **ACL** 2025.
- Croissant: A Metadata Format for ML-Ready Datasets. spotlight (top 3% of submissions)*
Mubashara Akhtar*, Omar Benjelloun*, Costanza Conforti*, et al.
Advances in Neural Information Processing Systems 36 (**NeurIPS** 2024).

Croissant: A Metadata Format for ML-Ready Datasets. (best paper award)

Mubashara Akhtar*, Omar Benjelloun*, Costanza Conforti*, et al.

Proceedings of the Eighth Workshop on Data Management for End-to-End Machine Learning, 2024.

ChartCheck: Explainable Fact-Checking over Real-World Chart Images.

Mubashara Akhtar, Nikesh Subedi, Vivek Gupta, Sahar Tahmasebi, Oana Cocarascu, Elena Simperl.

Findings of the Association for Computational Linguistics (**ACL**), 2024.

Exploring the Numerical Reasoning Capabilities of Language Models: A Comprehensive Analysis on Tabular Data.

Mubashara Akhtar*, Abhilash Shankarampeta*, Vivek Gupta, et al.

Findings of Annual Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2023.

SERVICE & SCIENTIFIC EXPERIENCE

Research chair, Evaluating Evaluations (EvalEval) Coalition	since Mar 2025
Organisation committee workshop on Evaluating Evaluations (EvalEval) , co-located with ACL '26	since Sept 2025
Co-Lead Responsible AI working group, Croissant ML Dataset Format , MLCommons	since Aug 2023
Organisation committee FEVER workshop , co-located with EACL '23 , EMNLP '24 , ACL '25	since Oct 2022
Organisation King's College London KG Lab's weekly seminar group	Dec 2020 - Dec 2022
Teaching Assistant, Oxford Machine Learning Summer School	July 2021
Organisation of a Women in ML (WiML) Workshop at NeurIPS 2020	Dec 2020
ACM Summer School on Recommender Systems	Aug 2017

GRANTS & AWARDS

- Awarded the [ETH Zurich AI Center Fellowship](#) (3% acceptance rate), 2025
- Best Short Paper Award, [DEEM workshop @SIGMOD](#), 2024
- Research Impact Award by Faculty of Natural, Mathematical & Engineering Sciences, [King's College London](#), 2023
- [King's College London NMES Faculty Studentship Award](#) (Total £79,9223), 2020-2024
- [European Forum Alpbach](#) scholarship for the years 2020/21 (5% acceptance rate), 2020
- US Embassy of Austria's Austria to Austin Entrepreneurship program, 2019
- [Ban Ki-moon Center for Global Citizens](#) mentorship (mentored by TU Wien vice-rector), 2019
- Research scholarship from TU Wien for master's thesis, 2017
- ACM RecSys summer school and conference grant from TU Wien, 2017
- Graduated with the highest possible grade (overall 1.0) in the Austrian 'Matura' (A-level equivalent)

MEDIA & OUTREACH (SELECTED)

Panel on AI & Data, AI House Davos , World Economic Forum	Davos, Switzerland (2026)
Talk on the Croissant Metadata Standard for ML Datasets, Circle U. Workshop	Vienna, Austria (2025)
Talk on Data-centric Machine Learning, ETH Zurich	Zurich, Switzerland (2025)
Interview on AI and misinformation with the Austrian public broadcaster, ORF	Vienna, Austria (2024)
Panel discussion on AI and misinformation at Österreichischer Journalistinnenkongress	Vienna, Austria (2024)
Talk on Multimodal Automated Fact-Checking (AFC) at CardiffNLP , Cardiff University	Cardiff, UK (2024)
Talk on Multimodal AFC at Int. Multimodal Communication Centre , University of Oxford	Oxford, UK (2024)
Talk on Chart Reasoning at iPragma group , Google DeepMind	Zurich, Switzerland (2023)
Panel discussion at the European Forum Alpbach 2023	Alpbach, Austria (2023)
Talk at Alan Turing Visualization Interest Group 5th Symposium	London, UK (2023)
Panel moderation A digital Future with Human Values , Austrian Embassy UK	London, UK (2023)
Interview for the podcast FrauenFunk , City Council of Vienna	Vienna, Austria (2021)

EXTRA-CURRICULAR

President, Club Alpbach London	Oct 2022 - Nov 2023
--	---------------------

PUBLICATIONS

Journal Articles

Ev2R: Evaluating Evidence Retrieval in Automated Fact-Checking

Mubashara Akhtar, M. Schlichtkrull, A. Vlachos.

TACL, 2026; Presented at **EACL 2026**.

TANQ: An Open Domain Dataset of Table Answered Questions

Mubashara Akhtar*, C. Pang*, A. Marzoca, Y. Altun, J.M. Eisenschlos.

TACL, 2025; Presented at **ACL 2025**.

Conference Papers¹

When AI Benchmarks Plateau: A Systematic Study of Benchmark Saturation

Mubashara Akhtar*, A. Reuel*, P. Soni, S. Ahuja, P. Sasanka Ammanamanchi, R. Rawal, V. Zouhar, S. Yadav, C. Whitehouse, et al.

Accepted at **ICML**, 2026.

Who Evaluates AI's Social Impacts? Mapping Coverage and Gaps in First and Third Party Evaluations

A. Reuel, A. Ghosh, J. Chim, A. Tran, Y. Long, J. Mickel, U. Gohar, S. Yadav, **M. Akhtar**, et al.

Accepted at **ICML**, 2026.

Efficient Test-Time Scaling of Multi-Step Reasoning by Probing Internal States of Large Language Models

J. Ni, E. Fadeeva, T. Wu, **M. Akhtar**, J. Zhang, E. Ash, M. Leippold, T. Baldwin, et al.

Accepted at **ACL**, 2026.

Compose and Fuse: Revisiting the Foundational Bottlenecks in Multimodal Reasoning

Y. Wang, Y. Hou, A. Javadov, **M. Akhtar**, M. Sachan.

ICLR, 2026.

Lexam: Benchmarking legal reasoning on 340 law exams

Y. Fan, J. Ni, J. Merane, Y. Tian, Y. Hermstrüwer, Y. Huang, **M. Akhtar**, et al.

ICLR, 2026.

Croissant: A Metadata Format for ML-Ready Datasets

Mubashara Akhtar, O. Benjelloun, C. Conforti, L. Foschini, J. Giner-Miguel, et al.

NeurIPS, 2024. **Spotlight** paper (top 3%).

ChartCheck: Explainable Fact-Checking over Real-World Chart Images

Mubashara Akhtar, N. Subedi, V. Gupta, S. Tahmasebi, O. Cocarascu, E. Simperl.

Findings of the Association for Computational Linguistics (**ACL**), 2024.

Exploring the Numerical Reasoning Capabilities of Language Models: A Comprehensive Analysis on Tabular Data

Mubashara Akhtar, A. Shankarampeta, V. Gupta, A. Patil, O. Cocarascu, E. Simperl.

Findings of the Association for Computational Linguistics: **EMNLP '23**, 2023.

Multimodal Automated Fact-Checking: A Survey

Mubashara Akhtar, M. Schlichtkrull, Z. Guo, O. Cocarascu, E. Simperl, A. Vlachos.

Findings of the Association for Computational Linguistics: **EMNLP '23**, 2023.

Reading and Reasoning over Chart Images for Evidence-Based Automated Fact-Checking

Mubashara Akhtar, O. Cocarascu, E. Simperl.

Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (**EACL**), 2023.

PubHealthTab: A Public Health Table-Based Dataset for Evidence-Based Fact Checking

Mubashara Akhtar, O. Cocarascu, E. Simperl.

Findings of the Association for Computational Linguistics: **NAACL '22**, 2022.

The Potential of Chatbots: Analysis of Chatbot Conversations

Mubashara Akhtar, J. Neidhardt, H. Werthner.

IEEE CBI, 2019.

¹In Artificial Intelligence research, peer-reviewed conferences are the primary publication venue. The conferences listed here follow rigorous review standards and typically have acceptance rates of around 20–25%.

Preprints

Benchmarking and Enhancing Text-to-Image Models for Generating Visual Representations in Early Arithmetic Education

J. Wang, B. Chen, H. Do, **M. Akhtar**, A. Wang, M. Sachan

arXiv:2605.31212, 2026.

Chimera: Diagnosing Shortcut Learning in Visual-Language Understanding

Z. Chi, Y. Hou, C. Pang, S. Cui, **M. Akhtar**, M. Sachan.

arXiv:2509.22437, 2025.

Ailuminate: Introducing v1.0 of the AI Risk and Reliability Benchmark from MLCommons

S. Ghosh, H. Frase, A. Williams, S. Luger, P. Röttger, F. Barez, S. McGregor, **M. Akhtar**, et al.

arXiv:2503.05731, 2025.

The Automated Verification of Textual Claims (AVeriTeC) Shared Task

M. Schlichtkrull, Y. Chen, C. Whitehouse, Z. Deng, **M. Akhtar**, R. Aly, et al.

arXiv:2410.23850, 2024.

A Standardized Machine-readable Dataset Documentation Format for Responsible AI

N. Jain*, **M. Akhtar***, J. Giner-Miguel*, R. Shinde*, J. Vanschoren, S. Vogler, et al.

arXiv:2407.16883, 2024.

Other / Technical Reports

Multimodal Evidence Integration for Automated Fact-Checking

Mubashara Akhtar.

Doctoral Thesis, 2024.

Croissant Format Specification

O. Benjelloun, E. Simperl, P. Marcenac, P. Ruysen, C. Conforti, M. Kuchnik, **M. Akhtar**, et al.

2024.

Workshop Papers

The 2nd Automated Verification of Textual Claims (AVeriTeC) Shared Task: Open-Weights, Reproducible and Efficient Systems

Mubashara Akhtar, R. Aly, Y. Chen, Z. Deng, M. Schlichtkrull, C. Whitehouse, et al.

Proceedings of the Eighth FEVER Workshop, 2025.

Proceedings of the Eighth Fact Extraction and VERification Workshop (FEVER)

Mubashara Akhtar, R. Aly, C. Christodoulopoulos, O. Cocarascu, Z. Guo, A. Mittal, et al.

FEVER Workshop, 2025.

Proceedings of the Seventh Fact Extraction and VERification Workshop (FEVER)

M. Schlichtkrull, Y. Chen, C. Whitehouse, Z. Deng, **M. Akhtar**, R. Aly, et al.

FEVER Workshop, 2024.