FEVER 2025

The Eighth Fact Extraction and VERification Workshop

Proceedings of the Workshop

July 31, 2025

The FEVER organizers gratefully acknowledge the support from the following sponsors.

Supported by

©2025 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 978-1-959429-53-1

Introduction

With billions of individual pages on the web providing information on almost every conceivable topic, we should have the ability to collect facts that answer almost every conceivable question. However, only a small fraction of this information is contained in structured sources such as Wikidata; we are therefore limited by our ability to transform free-form text to structured knowledge. There is, however, another problem that has become the focus of a lot of recent research and media coverage: false information coming from unreliable sources.

To ensure accuracy, any content must be verified. However, the volume of information precludes human moderators from doing so. Hence, it is paramount to research automated means to verify accuracy and consistency of information published online and the downstream systems (such as Question Answering, Search and Digital Personal Assistants) which rely on it.

The eighth edition of the FEVER workshop collocated with ACL 2025 aims to continue promoting ongoing research in above area, following on from the first seven collocated with EMNLP 2018, EMNLP 2019, ACL 2020, EMNLP 2021, ACL 2022, EACL 2023, EMNLP 2024, and four shared tasks in 2018, 2019, 2021, and 2024. This year's workshop consists of 3 oral and 14 poster presentations of accepted papers (64% overall acceptance rate), 3 poster presentations from ACL Findings papers, and presentations from 5 invited speakers. FEVER 2025 also hosts the second AVeriTeC shared task on open-weights, reproducible and efficient fact verification systems of real-world claims. The annotation of the new shared task test set was conducted by a donation from Google. The workshop is held in hybrid mode with in-person and virtual poster sessions, live-streamed panel discussion, oral presentations, and invited talks.

The organisers would like to thank the authors of all submitted papers, the reviewers, the panelists, and the invited speakers for their efforts, and we are looking forward to next year's edition.

Best wishes,
The FEVER organisers

Organizing Committee

Workshop Organisers

Mubashara Akhtar, King's College London
Rami Aly, University of Cambridge
Rui Cao, University of Cambridge
Yulong Chen, University of Cambridge
Christos Christodoulopoulos, Amazon
Oana Cocarascu, King's College London
Zhenyun Deng, University of Cambridge
Zifeng Ding, University of Cambridge
Zhijiang Guo, Hong Kong University of Science and Technology
Arpit Mittal, Meta
Michael Schlichtkrull, Queen Mary University of London
James Thorne, KAIST AI
Andreas Vlachos, University of Cambridge
Chenxi Whitehouse, Meta

Program Committee

Program Committee

Mahmud Elahi Akhter, Queen Mary, University of London
Iqra Ali, Queen Mary, University of London
Anab Maulana Barik
Andrew Bell, New York University
Rui Cao, University of Cambridge and Singapore Management University
Jenny Chim, Queen Mary, University of London and Queen Mary University London
Zifeng Ding, University of Cambridge
Roksana Goworek, Queen Mary University of London
Kriti Goyal, Apple
Lovisa Hagström, Chalmers University of Technology
Spencer Hong, National University of Singapore
Maeve Hutchinson, City University
Saidah Zahrotul Jannah, Universitas Airlangga and Nara Institute of Science and Technology
Myunghoon Kang, Korea University
Pride Kavumba, SB Intuitions
David Kletz, IDSIA
Guneet Singh Kohli, Queen Mary, University of London and GreyOrange
Gaurav Kumar, Moveworks
Pietro Lesci, University of Cambridge
Irene Li, University of Tokyo
Christopher Malon, NEC Laboratories America
Kunwoo Park, Soongsil University
Heesoo Park, Dunamu
Parth Patwa, Amazon and University of California, Los Angeles
Dominik Stammbach, Princeton University
Herbert Ullrich, Czech Technical University
Juraj Vladika, Technische Universität München
Moy Yuan, University of Cambridge
Linda Zeng, The Harker School, San Jose and University of California, Santa Cruz
Yuji Zhang, University of Illinois at Urbana-Champaign
Dongzhuoran Zhou, University of Oslo
Yuxiang Zhou, Queen Mary, University of London

Invited Speakers

Christo Buschek, BuzzFeed
Pepa Atanasova, University of Copenhagen
Oana Maria Camburu, Imperial College London
Leyang Cui, Tencent
Michiel Bakker, Massachusetts Institute of Technology

Table of Contents

Automated Claim–Evidence Extraction for Political Discourse Analysis: A Large Language Model Approach to Rodong Sinmun Editorials
Gyuri Choi and Hansaem Kim 1

Language Model Re-rankers are Fooled by Lexical Similarities
Lovisa Hagström, Ercong Nie, Ruben Halifa, Helmut Schmid, Richard Johansson and Alexander Junge 18

Portuguese Automated Fact-checking: Information Retrieval with Claim extraction
Juliana Gomes, Eduardo Garcia and Arlindo Rodrigues Galvão Filho 34

Multilingual Symptom Detection on Social Media: Enhancing Health-related Fact-checking with LLMs
Saidah Zahrotul Jannah, Elyanah Aco, Shaowen Peng, Shoko Wakamiya and Eiji Aramaki 54

When Scale Meets Diversity: Evaluating Language Models on Fine-Grained Multilingual Claim Verification
Hanna Shcharbakova, Tatiana Anikina, Natalia Skachkova and Josef Van Genabith 69

Less Can be More: An Empirical Evaluation of Small and Large Language Models for Sentence-level Claim Detection
Andrew Bell 85

RAG based Question Answering of Korean Laws and Precedents
Kiho Seo and Takehito Utsuro 91

FACT5: A Novel Benchmark and Pipeline for Nuanced Fact-Checking of Complex Statements
Shayan Chowdhury, Sunny Fang and Smaranda Muresan 101

Correcting Hallucinations in News Summaries: Exploration of Self-Correcting LLM Methods with External Knowledge
Juraj Vladika, Ihsan Soydemir and Florian Matthes 118

The Law of Knowledge Overshadowing: Towards Understanding, Predicting, and Preventing LLM Hallucination
Yuji Zhang, Sha Li, Cheng Qian, Jiateng Liu, Pengfei Yu, Chi Han, Yi R. Fung, Kathleen McKeown, ChengXiang Zhai, Manling Li and Heng Ji 132

GQC: LLM-Based Grouped QA Consolidation for Open-Domain Fact Verification at AVeriTeC
Dongzhuoran Zhou, Roxana Pop, Yuqicheng Zhu and Evgeny Kharlamov 151

(Fact) Check Your Bias
Eivind Morris Bakke and Nora Winger Heggelund 162

EMULATE: A Multi-Agent Framework for Determining the Veracity of Atomic Claims by Emulating Human Actions
Spencer Hong, Meng Luo and Xinyi Wan 179

SemQA: Evaluating Evidence with Question Embeddings and Answer Entailment for Fact Verification
Kjetil Indrehus, Caroline Vannebo and Roxana Pop 184

The 2nd Automated Verification of Textual Claims (AVeriTeC) Shared Task: Open-weights, Reproducible and Efficient Systems
Mubashara Akhtar, Rami Aly, Yulong Chen, Zhenyun Deng, Michael Schlichtkrull, Chenxi Whitehouse and Andreas Vlachos 201

Team HUMANE at AVeriTeC 2025: HerO 2 for Efficient Fact Verification
Yejun Yoon, Jaeyoon Jung, Seunghyun Yoon and Kunwoo Park 224

Exploring Semantic Filtering Heuristics For Efficient Claim Verification
Max Upravitelev, Premtim Sahitaj, Arthur Hilbert, Veronika Solopova, Jing Yang, Nils Feldhus, Tatiana Anikina, Simon Ostermann and Vera Schmitt 229

OldJoe at AVeriTeC: In-context learning for fact-checking
Farah Ftouhi, Russel Dsouza, Lance Calvin Lim Gamboa, Asim Abbas, Mubashir Ali, Yue Feng, Mark G. Lee and Venelin Kovatchev 238

SANCTUARY: An Efficient Evidence-based Automated Fact Checking System
Arbaaz Dharmavaram and Saqib Hakak 247

Fathom: A Fast and Modular RAG Pipeline for Fact-Checking
Farrukh Bin Rashid and Saqib Hakak 258

Graph-of-Thoughts for Fact-Checking with Large Language Models
Sascha Rolinger and Jin Liu 266

AIC CTU@FEVER 8: On-premise fact checking through long context RAG
Herbert Ullrich and Jan Drchal 274

Program

Thursday, July 31, 2025

09:00 - 09:05 *Opening*

09:05 - 09:50 *Keynote Talk: Christo Buschek*

09:50 - 10:10 *Shared Task Overview*

The 2nd Automated Verification of Textual Claims (AVeriTeC) Shared Task: Open-weights, Reproducible and Efficient Systems
Mubashara Akhtar, Rami Aly, Yulong Chen, Zhenyun Deng, Michael Schlichtkrull, Chenxi Whitehouse and Andreas Vlachos

10:10 - 10:30 *Contributed Shared Task Talks*

AIC CTU@FEVER 8: On-premise fact checking through long context RAG
Herbert Ullrich and Jan Drchal

Exploring Semantic Filtering Heuristics For Efficient Claim Verification
Max Upravitelev, Premtim Sahitaj, Arthur Hilbert, Veronika Solopova, Jing Yang, Nils Feldhus, Tatiana Anikina, Simon Ostermann and Vera Schmitt

10:30 - 11:00 *Morning Break*

11:00 - 12:00 *Poster session*

Automated Claim–Evidence Extraction for Political Discourse Analysis: A Large Language Model Approach to Rodong Sinmun Editorials
Gyuri Choi and Hansaem Kim

Language Model Re-rankers are Fooled by Lexical Similarities
Lovisa Hagström, Ercong Nie, Ruben Halifa, Helmut Schmid, Richard Johansson and Alexander Junge

Portuguese Automated Fact-checking: Information Retrieval with Claim extraction
Juliana Gomes, Eduardo Garcia and Arlindo Rodrigues Galvão Filho

Multilingual Symptom Detection on Social Media: Enhancing Health-related Fact-checking with LLMs
Saidah Zahrotul Jannah, Elyanah Aco, Shaowen Peng, Shoko Wakamiya and Eiji Aramaki

When Scale Meets Diversity: Evaluating Language Models on Fine-Grained Multilingual Claim Verification
Hanna Shcharbakova, Tatiana Anikina, Natalia Skachkova and Josef Van Genabith

Less Can be More: An Empirical Evaluation of Small and Large Language Models for Sentence-level Claim Detection
Andrew Bell

RAG based Question Answering of Korean Laws and Precedents
Kiho Seo and Takehito Utsuro

FACT5: A Novel Benchmark and Pipeline for Nuanced Fact-Checking of Complex Statements
Shayan Chowdhury, Sunny Fang and Smaranda Muresan

Correcting Hallucinations in News Summaries: Exploration of Self-Correcting LLM Methods with External Knowledge
Juraj Vladika, Ihsan Soydemir and Florian Matthes

The Law of Knowledge Overshadowing: Towards Understanding, Predicting, and Preventing LLM Hallucination
Yuji Zhang, Sha Li, Cheng Qian, Jiateng Liu, Pengfei Yu, Chi Han, Yi R. Fung, Kathleen McKeown, ChengXiang Zhai, Manling Li and Heng Ji

GQC: LLM-Based Grouped QA Consolidation for Open-Domain Fact Verification at AVeriTeC
Dongzhuoran Zhou, Roxana Pop, Yuqicheng Zhu and Evgeny Kharlamov

(Fact) Check Your Bias
Eivind Morris Bakke and Nora Winger Heggelund

EMULATE: A Multi-Agent Framework for Determining the Veracity of Atomic Claims by Emulating Human Actions
Spencer Hong, Meng Luo and Xinyi Wan

SemQA: Evaluating Evidence with Question Embeddings and Answer Entailment for Fact Verification
Kjetil Indrehus, Caroline Vannebo and Roxana Pop

Team HUMANE at AVeriTeC 2025: HerO 2 for Efficient Fact Verification
Yejun Yoon, Jaeyoon Jung, Seunghyun Yoon and Kunwoo Park

Thursday, July 31, 2025 (continued)

Exploring Semantic Filtering Heuristics For Efficient Claim Verification
Max Upravitelev, Premtim Sahitaj, Arthur Hilbert, Veronika Solopova, Jing Yang, Nils Feldhus, Tatiana Anikina, Simon Ostermann and Vera Schmitt

OldJoe at AVeriTeC: In-context learning for fact-checking
Farah Ftouhi, Russel Dsouza, Lance Calvin Lim Gamboa, Asim Abbas, Mubashir Ali, Yue Feng, Mark G. Lee and Venelin Kovatchev

SANCTUARY: An Efficient Evidence-based Automated Fact Checking System
Arbaaz Dharmavaram and Saqib Hakak

Fathom: A Fast and Modular RAG Pipeline for Fact-Checking
Farrukh Bin Rashid and Saqib Hakak

Graph-of-Thoughts for Fact-Checking with Large Language Models
Sascha Rolinger and Jin Liu

AIC CTU@FEVER 8: On-premise fact checking through long context RAG
Herbert Ullrich and Jan Drchal

12:00 - 12:45 *Keynote Talk: Pepa Atanasova*

12:45 - 14:00 *Lunch Break*

14:00 - 14:45 *Keynote Talk: Oana Maria Camburu*

14:45 - 15:30 *Keynote Talk: Leyang Cui*

15:30 - 16:00 *Afternoon Break*

16:00 - 16:30 *Contributed Workshop Talks*

When Scale Meets Diversity: Evaluating Language Models on Fine-Grained Multilingual Claim Verification
Hanna Shcharbakova, Tatiana Anikina, Natalia Skachkova and Josef Van Genabith

FACT5: A Novel Benchmark and Pipeline for Nuanced Fact-Checking of Complex Statements
Shayan Chowdhury, Sunny Fang and Smaranda Muresan

The Law of Knowledge Overshadowing: Towards Understanding, Predicting, and Preventing LLM Hallucination
Yuji Zhang, Sha Li, Cheng Qian, Jiateng Liu, Pengfei Yu, Chi Han, Yi R. Fung, Kathleen McKeown, ChengXiang Zhai, Manling Li and Heng Ji

16:30 - 17:15 *Keynote Talk: Michiel Bakker*

17:15 - 17:30 *Closing Remarks*