

# BeyondFacts 2025: 5th International Workshop on Computational Methods for Online Discourse Analysis

Konstantin Todorov  
todorov@lirmm.fr

LIRMM, University of Montpellier, CNRS, Montpellier,  
France

Stefan Dietze  
stefan.dietze@gesis.org

GESIS, Cologne & Heinrich-Heine-University Düsseldorf,  
Germany

Pavlos Fafalios  
pfafalios@tuc.gr

Technical University of Crete and FORTH-ICS,  
Chania, Greece

Dimitar Dimitrov  
dimitar.dimitrov@gesis.org  
GESIS, Cologne, Germany

## Abstract

This workshop explores the intersection of computational and interdisciplinary approaches to analyzing online discourse, including claims, arguments, and opinions on controversial topics. With the rise of mis- and disinformation, bias, and echo chambers, NLP-based methods such as argument mining, stance detection, and fact verification have become essential. However, these tasks require robust conceptual foundations across fields like communication studies, computational linguistics, and computer science. BeyondFacts fosters collaboration among diverse research communities—including social sciences, political science, computational journalism, and computer science—to enhance machine-interpretation and analysis of societal debates using techniques from Web mining, AI, and NLP.

## CCS Concepts

• **Computing methodologies** → **Machine learning**; **Artificial intelligence**; **Natural language processing**; **Discourse, dialogue and pragmatics**; **Information extraction**;

## Keywords

Online discourse analysis, Web mining, Social media mining, Computational fact-checking, Mis- and dis-information spread and detection, Stance viewpoint discovery, intent detection, computational journalism, knowledge graphs, LLMs

## ACM Reference Format:

Konstantin Todorov, Pavlos Fafalios, Stefan Dietze, and Dimitar Dimitrov. 2025. BeyondFacts 2025: 5th International Workshop on Computational Methods for Online Discourse Analysis. In *Companion Proceedings of the ACM Web Conference 2025 (WWW Companion '25)*, April 28-May 2, 2025, Sydney, NSW, Australia. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3701716.3717540>

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
*WWW Companion '25*, Sydney, NSW, Australia  
© 2025 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-1331-6/2025/04  
<https://doi.org/10.1145/3701716.3717540>

## 1 Scope and Objectives of BeyondFacts

Expressing opinions and interacting with others on the Web has led to the production of an abundance of online discourse data, such as claims and viewpoints on controversial topics, their sources and contexts (events, entities). This data constitutes a valuable source of insights for studies into misinformation spread, bias reinforcement, echo chambers or political agenda setting. Computational methods, mostly from the field of NLP, have emerged that tackle a wide range of tasks in this context, including argument and opinion mining, claim detection, checkworthiness detection, stance detection or fact verification. However, computational models require robust definitions of classes and concepts under investigation. Thus, these computational tasks require a strong interdisciplinary and epistemological foundation, specifically with respect to the underlying definitions of key concepts such as claims, arguments, stances, check-worthiness or veracity. This requires a highly interdisciplinary approach combining expertise from fields such as communication studies, computational linguistics and computer science. As opposed to facts, claims are inherently more complex. Their interpretation strongly depends on the context and a variety of intentional or unintended meanings, where terminology and conceptual understandings strongly diverge across communities. From a computational perspective, in order to address this complexity, the synergy of multiple approaches, coming both from symbolic (knowledge representation) and statistical AI seem to be promising to tackle such challenges.

We define online discourse as any kind of narrative, debate or conversation that happens on the Web, including social networks or news media, involving **claims** and **stances on controversial topics**, their **sources** and **contexts** (such as related events or entities). Recently, a wide range of interdisciplinary research directions are being explored involving a variety of scientific disciplines. Such works either are focused on gaining new scientific insights, for instance, by investigating the spreading pattern of false claims on Twitter, or, they aim at computational methods, for instance, pipelines for detecting the stance of claim-relevant Web documents, approaches for classifying sources of news, such as Web pages, pay-level domains (PLDs), users or posts, or research into **fake news detection** and automatic **fact-checking** and mining of facts and claims from the Web for knowledge base augmentation.

One crucial requirement to facilitate the aforementioned research areas is the availability of reliable structured knowledge

about key notions such as claims, truth ratings, evidence, sources, arguments and their relations. On the one hand, initiatives such as the schema.org Claim Review vocabulary<sup>1</sup> aim at encouraging website providers to offer such data through embedded Web markup. On the other hand, knowledge graphs (KG) such as ClaimsKG<sup>2</sup> have been proposed aimed at consolidating Web-mined data about the aforementioned notions.

Initial efforts have been made to gather communities working in those areas, for instance through dedicated challenges (such as the Fake News Challenge<sup>3</sup> or sessions at major conferences, such as the Journalism, Misinformation and Fact Checking track at The Web Conf 2018.<sup>4</sup> The BeyondFacts Workshop brings together the various disciplines involved in or benefitting from (a) approaches for representing online discourse and involved notions, (b) methods for mining such notions (for instance, claims, stances, sources, etc) and their relations from the Web, or (c) inter-disciplinary research investigating online discourse.

Beyond research into information and knowledge extraction, data consolidation and modeling, the workshop targets communities focusing on the analysis of online discourse, relying on methods from **machine learning (ML)**, **natural language processing (NLP)** and **Web data mining (DM)**. These include communities on:

- discourse analysis
- social web mining
- argumentation mining
- computational fact-checking
- mis-, mal- and dis-information spread
- bias and controversy detection and analysis
- stance/viewpoint detection and representation
- opinion mining
- rumour, propaganda and hate-speech detection
- computational journalism

BeyondFacts provides a meeting point for these related but distinct communities that address similar or closely related questions from different perspectives and in different fields, using different models and definitions of the main notions of interest. Often these communities apply their research in particular domains, such as scientific publishing, medicine, journalism or social science. Therefore, the workshop is particularly interested in works that apply an interdisciplinary approach, such as works on **computational social sciences** or **computational journalism**.

Based on the high interest and participation (around 120 participants in all four past events<sup>5</sup>), we consider the first four editions of the workshop successful steps towards fostering a community on discourse analysis via computational methods and structured knowledge in the context of the Web. The **fifth anniversary edition** continues this effort, by opening up towards **novel communities** and strengthening the interdisciplinary nature of the forum. To do so, we have invited as members of the programme committee authors from the previous editions that come from fields such as

*social sciences* and *political sciences*, as well as confirmed experts in *computational journalism* and experts in analyzing *audiovisual content*, enabling to go beyond textual data (e.g. for the analysis of deep fakes). In addition, we have reached out to renowned experts from the *industry*. We have also attract scholars in ethics and law and have updated the current edition's topics of interest and relevant communities accordingly.

As a main difference in this year's edition, we (a) include of a poster session in addition to the presentations (for strengthening interactions among the workshop participants); (b) a round table discussing interactions between academia and industry.

## 2 Topics of Interest

The workshop focuses on the following main topics of interest:

- Large language models for online discourse
- Computational fact-checking, Computational journalism
- Narrative detection and extraction from the web
- Social, ethical and legal aspects of online discourse
- Bias and controversy detection and analysis
- Stance and viewpoint discovery
- Interpretability/explainability of computational methods
- Rumour, propaganda and hate-speech detection
- Intent discovery for claims
- Integration/aggregation/linking/enrichment of discourse data
- Multilingual analysis of online discourse data
- Knowledge graphs, data models and knowledge extraction techniques in the context of online discourse
- Semantic and exploratory search of online discourse data
- Argumentation and reasoning over online discourse
- Recommender systems for discourse data
- Quality, uncertainty, provenance, and trust of discourse data
- Dealing with online audiovisual content
- Benchmarks and training data for discourse analysis tasks
- Use-cases, applications and cross-community interfaces

A discussion on why this topic is relevant at this time is provided in the next section.

## 3 Workshop Format, Organisation and Programme Committee

*Format.* The workshop includes a mix of research papers, position papers, demo/system papers, resource papers, keynotes and invited talks. Note that we will accept position papers or papers with intermediate/incomplete results, the idea being to foster discussion as much as presenting finished results. The workshop will contain:

- Paper presentations (full, short, resource, position, posters, demos)
- Two one-hour poster sessions for all accepted contributions providing an opportunity for discussions and networking
- Two keynotes / invited talks from senior researchers on relevant topics (with one of them being highly interdisciplinary)
- Closing session and discussion on future directions and challenges, with a best paper award ceremony.

All contributions are eligible for the "Best BeyondFacts Paper" award, which will be awarded to the best contribution during the

<sup>1</sup><http://schema.org/ClaimReview>

<sup>2</sup><https://data.gesis.org/claimskg/site>

<sup>3</sup><http://www.fakenewschallenge.org/>

<sup>4</sup><https://archives.iw3c2.org/www2018/call-for-papers/misinformation-cfp/>

<sup>5</sup>The list of participants was closed to the workshop organisers in the first two hybrid events, therefore we do not have exact statistics.

closing session. In addition, the workshop organisation team is planning to invite selected workshop submissions to a dedicated journal special issue (to be specified).

The 2025 edition of BeyondFacts received **14 submissions**, out of which 11 were accepted (3 short and 8 full papers).

In addition, the workshop features two keynote talks.

*Organization.* The workshop is co-organized by:

**Konstantin Todorov**, University of Montpellier / LIRMM / CNRS, France, konstantin.todorov@lirmm.fr

**Stefan Dietze**, Heinrich-Heine-University Düsseldorf & GESIS, Cologne, Germany, stefan.dietze@gesis.org

**Pavlos Fafalios**, Technical University of Crete and FORTH-ICS, Chania, Greece, pfafalios@tuc.gr

**Dimitar Dimitrov**, GESIS, Cologne, Germany, dimitar.dimitrov@gesis.org

*Programme Committee.*

- Katarina Boland, Heinrich Heine University, Germany
- Sandra Bringay, Paul Valéry University of Montpellier, France
- Elena Cabrio, University Cote D'Azur, France
- John Culbert, GESIS, Germany
- Ronald Denaux, Amazon.com, Spain
- Gianluca Demartini, University of Queensland, Australia
- Vasilis Efthymiou, Harokopio University of Athens and FORTH-ICS, Greece
- Agata Gurzawska, Trilateral, Ireland
- José Manuel Gómez Pérez, Expert.AI, Spain
- Salim Hafid, LIRMM, Université Montpellier, France
- Kyle Hamilton, Technological University Dublin, Ireland
- Daniel Hardt, Copenhagen Business School, Denmark
- Andreea Iana, University of Mannheim, Germany
- Mali Jin, University of Sheffield, UK
- Mathieu Lafourcade, University of Montpellier, France
- Stephan Linzbach, GESIS, Germany
- Yannis Marketakis, University of Crete and FORTH-ICS, Greece
- Diana G. Maynard, Sheffield, UK
- Petr Motlicek, Idiap, Switzerland
- Michalis Mountantonakis, University of Crete and FORTH-ICS, Greece
- Preslav Nakov, MBZUAI, United Arab Emirates
- Wolfgang Otto, GESIS, Germany
- Panagiotis Papadakis, FORTH-ICS, Greece
- Katerina Papantoniou, FORTH-ICS, Greece
- Ensiyeh Raouffi, LIRMM, Université Montpellier, France
- Achim Rettinger, University of Trier, Germany
- Mathieu Roche, CIRAD, France
- François Scharffe, LIRMM, Université Montpellier, France
- Sebastian Schellhammer, GESIS, Germany
- Daniel Schwabe, Pontificia Universidade Católica, Brazil
- Kostas Stefanidis, Tampere University, Finland
- Andon Tchechmedjiev, Ecoles des Mines d'Alès, France
- Yannis Tzitzikas, University of Crete and FORTH-ICS, Greece
- Ran Yu, University of Bonn, Germany

*Subreviewers.*

- Xiaoou Wang, Inria, France
- Jonas Fegert, FZI Forschungszentrum Informatik, Germany

## 4 Past Editions of the Workshop

The **BeyondFacts 2020 workshop**<sup>6</sup> and the **BeyondFacts 2021 workshop**<sup>7</sup> took place as a virtual events (due to COVID-19 outbreak) jointly with the **29th and 30th Web Conferences (WWW 2020 and 2021)**, respectively. The third edition<sup>8</sup> and the fourth edition<sup>9</sup> took place as hybrid events jointly with the **31st and 32nd Web Conferences (WWW 2023 and 2024)**, respectively.

All four editions brought together diverse communities of researchers from different fields such as neural language models, argument mining and knowledge graphs or databases, but also social, political science and journalism scholars. The accepted and presented papers span a palette of topics such as claim detection, relation extraction for online discourse modeling, interpretable graph embeddings for misinformation detection, disinformation on social networks, explainability, stance detection, fact-checking in relation to argumentation schemes and false narratives, political and social scientific perspectives on propaganda chains and discourse mapping.

**In the first edition**, seven papers were accepted for publication, after a peer-review process (each submitted paper was reviewed by three PC members). The **proceedings** are available at: <http://ceur-ws.org/Vol-2877/>. The **best paper award** was given to the paper by Giovanni Luca Ciampaglia and John Licato: *Fact-checking, False Narratives, and Argumentation Schemes*. In addition, the first edition of the workshop hosted three excellent keynotes: **Preslav Nakov** talked about detecting 'Fake News' before it was even written, media literacy and flattening the curve of the COVID-19 infodemic; **Daniel Schwabe** proposed his take on trust and information disorders seen as disputes of narratives, while **Ioana Manolescu** gave an overview and lessons learned from the ANR ContentCheck project, focusing on content management approaches for assisting journalists in their day-to-day fact-checking efforts.

**In the second edition**, five papers were accepted for publication. The **proceedings** are available at: <https://dl.acm.org/doi/proceedings/10.1145/3487553>. The **best paper award** was given to the paper by Anab Maulana Barik, Wynne Hsu and Mong Li Lee: *Incorporating External Knowledge for Evidence-based Fact Verification*. The second edition of the workshop hosted three brilliant keynote talks: **Serena Villata** talked about argument-based explanatory dialogues, **Harith Alani** discussed computational tools and analysis of our interactions with false and corrective information while **José Manuel Gómez Pérez** gave an industry keynote on accurate and explainable misinformation detection.

**In the third edition**, three papers were accepted for publication. The **proceedings** are available at: <https://dl.acm.org/doi/proceedings/10.1145/3543873>. The **best paper award** was given to the paper by Youri Peskine, Raphaël Troncy and Paolo Papotti, "Analyzing COVID-Related Social Discourse on Twitter using Emotion, Sentiment, Political Bias, Stance, Veracity and Conspiracy Theories". In addition, the third edition of the workshop hosted two excellent keynotes: **Kai Shu** talked about combating disinformation on social media and its challenges, while **Colin Porlezza**, a journalism scholar, discussed the responsible future of AI in journalism.

<sup>6</sup><https://knod2021.wordpress.com/>

<sup>7</sup><https://knod2021.wordpress.com/>

<sup>8</sup><https://beyondfacts2023.wordpress.com/>

<sup>9</sup><https://beyondfacts2024.wordpress.com/>

**In the fourth edition**, seven papers were accepted for publication. The **proceedings** are available at: <https://dl.acm.org/doi/proceedings/10.1145/3589335>. The **best paper award** was given to the paper by Luca Luceri, Eric Boniardi and Emilio Ferrara “Leveraging Large Language Models to Detect Influence Campaigns on Social Media”. The workshop hosted keynotes by **Isabelle Augenstein** (“*Detection Distortions in Science Reporting*”), **Abraham Bernstein** (“*Escaping the Echo Chamber: The Quest for Normative News Recommender Systems*”), and **Tanmoy Chakraborty** (“*Insights from accessing and mitigating hate speech on the Web*”).

## 5 Organizers Short Bio

**Konstantin Todorov** is associate professor for Computer Science at the University of Montpellier since 2012. His research lies in the field of computational models for graphs and text, including Web data linking, knowledge graphs and social media with applications in the cultural heritage, social science, agro-bio and medical domains. He is scientific coordinator of the AISci French-German project and the ANR DACE-DL project. K. Todorov serves as a PC member of a number of international conferences (e.g. ESWC, ISWC, the WebConf, ECAI) and has (co-)organised: (i) the KnOD/BeyondFacts series of workshops (with the WebConf 2021 to 2025); (ii) the Check That! shared task at CLEF 2025; (iii) OAEI in 2016 and 2017, jointly with ISWC; (iv) the DOREMUS tutorial at ESWC 2016; (v) the Special Session “Uncertainty and imprecision on the Web of Data” at IPMU 2014; (vi) The International Workshop on Ontologies for Multimedia Interpretation and Retrieval at SAMT 2010. Konstantin Todorov serves as a senior PC member for ESWC since 2022.

**Pavlos Fafalios** is assistant professor at the Technical University of Crete (Chania, Greece) and affiliated researcher at FORTH-ICS. His research interests fall in the areas of information systems, data and knowledge management, and knowledge engineering, with a special focus on interdisciplinary research in these areas and the application of his research findings in different disciplines, including Cultural Heritage and the Humanities. He regularly participates to the programme committee of international conferences (including ISWC, ESWC and HT), and reviews papers for international journals (including ACM JOCCH and Journal of Web Semantics). He was PC chair of the 13th International Joint Conference on Knowledge Graphs (IJCKG 2024, organized jointly with CCKS 2024), co-organiser of the 1st, 2nd, 3rd and 4th KnOD/BeyondFacts workshops (collocated with The WebConf 2021, 2022, 2023 and 2024), and proceedings chair of the 15th International Conference on Web-based Learning (ICWL 2016).

**Stefan Dietze** is full professor for Data and Knowledge Engineering at the Institute for Computer Science at Heinrich-Heine-University Düsseldorf, Scientific Director of the department Knowledge Technologies for the Social Sciences at GESIS – Leibniz Institute for the Social Sciences. His research interests are at the intersection of information retrieval, natural language processing, and knowledge graphs, and in particular, the extraction, fusion and search of knowledge and data on the Web in various application domains. Stefan’s work has been published at major scientific venues, such as WWW/The Web Conf, SIGIR, or ISWC, where he also frequently serves as PC and OC member. Stefan has been

involved in the organisation of major conferences (ESWC, ACM WebSci, ISWC, or as General Co-Chair of CIKM2020), and he has been involved in the organisation of successful workshop series (for instance, KnOD/BeyondFacts, PROFILES, LILE and LDOW).

**Dimitar Dimitrov** is leading the team Information Extraction and Linking and a postdoctoral researcher at department Knowledge Technologies for the Social Sciences at GESIS – Leibniz Institute for the Social Sciences (Cologne, Germany). Dimitar was a visiting researcher at the Graz University of Technology, Graz, Austria, and at the eXascale Infolab at the University of Fribourg, Fribourg, Switzerland. His research focuses on discourse data analysis from social media platforms. He co-organized the Doctoral Consortium at ECIR 2020 and co-chaired the CIKM2020 AnalytiCup for which he also prepared the COVID-19 Retweet Prediction Challenge. He regularly serves as a programme committee member at computer science conferences, such as ISWC, ESWC, HT, and CIKM. Dimitar was also a co-organizer of the 3rd and 4th BeyondFacts workshops, collocated with The WebConf 2023 and 2024.