PEPA ATANASOVA

Department of Computer Science, University of Copenhagen

PERSONAL DETAILS

Pepa Atanasova
LinkedIn: https://www.linkedin.com/in/pepa-atanasova-65a2b417b/
WWW: apepa.github.io
Google Scholar: https://scholar.google.com/citations?user=CLOC3rEAAAAJ&hl=en

RESEARCH INTERESTS

Natural Language Processing • Machine Learning • AI Explainability • Explainability Faithfulness and Diagnostics • Mechanistic Interpretability • Knowledge Mechanisms • Factuality

EMPLOYMENT

2024 –	**Tenure-Track Assistant Professor** in Natural Language Processing Section at the Department of Computer Science, University of Copenhagen
2022 – 2024	**Postdoctoral Researcher** in Natural Language Processing Section at the Department of Computer Science, University of Copenhagen
01 – 04/2022	**Research Intern** at Meta AI Research, Mountain View, USA
05 – 09/2023	**Research Intern** at Google, New York, USA
2016 – 2019	**NLP Researcher** at Siteground, Sofia, Bulgaria
–	Maternity Leaves: 01 – 09/2023, 03 – 09/2025

ACADEMIC EDUCATION

2019 – 2022	Ph.D., Computer Science, **University of Copenhagen (*defense 8th November 2022*)**
2015 – 2017	M.S., Artificial Intelligence, **Sofia University "St. Kl. Ohridski", Bulgaria**
2011 – 2015	B.S., Computer Science, **Sofia University "St. Kl. Ohridski", Bulgaria**

RESEARCH PROJECTS

- 2023 - Present – *Explainable and Robust Automatic Fact Checking*, co-PI
  - ⋆ European Research Council (ERC) Starting Grant for Isabelle Augenstein
- 2022 - 2024 – *Understanding the Effects of Natural Language Processing-Based Trading Algorithms*, Postdoctoral Fellow
  - ⋆ Funded by Villum Synergy Initiator Grant
- 2019 - 2022 – *Accountable and Explainable Methods for Complex Reasoning over Text*, PhD student
  - ⋆ European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 801199

GRANTS AND SCHOLARSHIPS

- DDSA Large Event Grant, 2025. Co-organising a local pre-ACL seminar to bring leading Natural Language Processing researchers to Denmark to foster international collaborations.
- Cohere For AI Research Grant Program, 2024. Grant for computing resources for faithful explanations in LLMs. Cohere For AI Research Grants are designed to support academic partners who are conducting research with the goal of releasing a scientific artifact.
- Informatics Europe (IE) best dissertation award, 2023, sponsored by Springer. PhD thesis published as a book in a dedicated Springer series.
- European Laboratory for Learning and Intelligent Systems (ELLIS) best dissertation award, 2023, sponsored by Kühborth Stiftung GmbH.
- PhD Fellowship under European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 801199.
- Travel Grant for Participation in 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019.

TALKS

- *Facts Unveiled: Navigating Factuality in the Era of Generative Models.* Invited talk, Sheffield NLP Group, Nov. 2024.
- *Navigating the Right to Explanation in AI: Methods and Technical Challenges* Keynote talk, Romanian AI days 2024, Sep. 2024.
- *Facts Unveiled: Navigating Factuality in the Era of Generative Models.* Opening Keynote, iGeLU 2024, Sep. 2024.
- *Exploring the Explainability Landscape: Testing and Enhancing Explainability Techniques.* Invited talk, Chalmers University, Nov. 2023.
- *From Opacity to Clarity: Embracing Transparent and Accountable Fact Verification.* Invited conference talk, MISDOOM, Nov. 2023.
- *Faithfulness Tests for Natural Language Explanations.* Talk at Nordic AI Meet, Sep. 2023.
- *Explainable and Accountable Fact Checking.* Invited talk at The University of Massachusetts' NLP group, April 2023.
- *Methods for Accountable and Explainable Complex Reasoning Tasks.* Invited talk at the Responsible Data Science and AI Speaker Series at the University of Illinois at Urbana-Champaign, October 2022.
- *When Research Goes Wrong: Deepfakes!.* Invited for a panel as a part of the Legal Tech Research Talks at the University of Copenhagen's Faculty of Law, March 2022.
- *Explainable and Accountable Automatic Fact Checking.* Invited talk for the NLP group at Oxford, February 2022.
- *Explaining Automated Fact Checking Predictions and Current Vulnerabilities.* Invited talk at FAIR's AI and Society talk series, September 2021.

- *Check-worthiness of Claims in Political Debates.* Invited talk at Data Science Society Meetup, Sofia, September 2018.
- *Leveraging Expert Annotations for Fact-Checking.* Invited talk at DataBeers Copenhagen, May 2019.
- *Check-worthiness of Claims in Political Debates.* Invited talk at Information Retrieval Workshop invited speaker, RANLP, September 2017.
- *Finding the Right Articles – A Supervised Approach to Search.* Talk at PyData London, April 2017.

TEACHING

- Guest lectures at the IT-University of Copenhagen, graduate level, "Explainability and Explainability Evaluations". Master's course, 2023. Developed and delivered lecture materials for 30 students.
- Fairness and Transparency in Machine Learning. Master's course, University of Copenhagen, 2022-2023. Developed and delivered lab and lecture materials for 28 students, and participated in examinations.
- Introduction to Natural Language Processing. Master's course, University of Copenhagen, 2019-2022. Developed and delivered mainly lab materials for up to 50 students, supplemented by lecturing materials in the last year, and participated in examinations.
- Tutorial at the Advanced Language Processing Winter School (ALPS), "Explainability and Explainability Evaluations". Graduate level, 2021. Developed and delivered lab materials for 50 students.
- Information Retrieval. Master's course, Sofia University. Developed and delivered lab materials for 20 students.
- Natural Language Processing. Master's course, Sofia University, 2018-2019. Developed and delivered lab materials for 20 students.
- Data Mining. Master's course, Sofia University, 2016-2017. Developed and delivered lab materials for 20 students.
- Artificial Intelligence. Bachelor's course, Sofia University, 2016-2017. Developed and delivered lab materials for 20 students, and participated in examinations.
- Object-Oriented Programming. Bachelor's course, Sofia University, 2012-2013. Developed and delivered lab materials for 20 students.
- Introduction to Programming. Bachelor's course, Sofia University, 2012-2013. Developed and delivered lab materials for 20 students.
- Workshop "Introduction to Machine Learning" for industry practitioners. 2019 Developed and delivered lecture and lab materials for 20 practitioners.

ONGOING SUPERVISION

- Sekh Manuil Islam (2024 -) - PhD (co-supervised with Isabelle Augenstein)

- Haeun Yu (2023 -) - PhD (co-supervised with Isabelle Augenstein)
- Jingyi Sun (2023 -) - PhD (co-supervised with Isabelle Augenstein)
- *Bachelor's students*: Emma Fowler (2024), Nikolaj Højer (2021)
- *Master's students*: Yingming Wang (2024 - 2025), Wojciech Ostrowski (2020)

SELECTED ORGANISATION OF SCIENTIFIC MEETINGS

- **Conference/Workshop Co-Organiser**: Pre-ACL Workshop 2025 in Copenhagen; Repl4NLP 2024; EACL 2023 (website co-chair)
- **Shared Task Co-Organiser**: Fact Checking and Check-worthiness Detection at CLEF 2018 and CLEF 2019, Community Forum Fact Checking at SemEval 2019, OffenseEval at SemEval 2020
- **Area Chair**: ARR 2023- , ACL 2023

BIBLIOMETRICS

(source Google Scholar, May 31, 2025): Publications: 46; Citations: 2759; h-index: 22; i10-index: 26

SELECTED PUBLICATIONS

- **Pepa Atanasova**, Oana-Maria Camburu, Christina Lioma, Thomas Lukasiewicz, Jakob Grue Simonsen, and Isabelle Augenstein. *Faithfulness Tests for Natural Language Explanations.* Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), July 2023.
- Jingyi Sun, **Pepa Atanasova**, Isabelle Augenstein. *From Tokens to Span Interactions: a Multi-level Comparative Framework for Highlight-based Explanations.* In Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL) 2025, Long Papers.
- Haeun Yu, **Pepa Atanasova**, Isabelle Augenstein. *Revealing the Parametric Knowledge of Language Models: A Unified Framework for Attribution Methods.* Under review at ACL Rolling Review for the Annual Meeting of the Association for Computational Linguistics (ACL), Long Papers, 2024.
- **Pepa Atanasova**, Jakob Grue Simonsen, Christina Lioma, Isabelle Augenstein. *Diagnostics-Guided Explanation Generation.* In Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI), February 2022.
- **Pepa Atanasova**, Jakob Grue Simonsen, Christina Lioma, Isabelle Augenstein (2020). *A Diagnostic Study of Explainability Techniques for Text Classification.* Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), November 2020.
- Sara Vera Marjanovic, Haeun Yu, **Pepa Atanasova**, Maria Maistro, Christina Lioma, Isabelle Augenstein *DYNAMICQA: Tracing Internal Knowledge Conflicts in Language Models.* Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2024), November 2024.

PAPERS IN CONFERENCE PROCEEDINGS

- Lovisa Hagstrom, Sara Vera Marjanovic, Haeun Yu, Arnav Arora, Christina Lioma, Maria Maistro, **Pepa Atanasova**, Isabelle Augenstein. *A Reality Check on Context Utilisation for Retrieval-Augmented Generation.* In Proceedings of the Association for Computational Linguistics 2025, Long Papers.
- Jingyi Sun, **Pepa Atanasova**, Isabelle Augenstein. *From Tokens to Span Interactions: a Multi-level Comparative Framework for Highlight-based Explanations.* In Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL) 2025, Long Papers.
- Sara Vera Marjanovic, Haeun Yu, **Pepa Atanasova**, Maria Maistro, Christina Lioma, Isabelle Augenstein *DYNAMICQA: Tracing Internal Knowledge Conflicts in Language Models.* Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2024), November 2024.
- Haeun Yu, **Pepa Atanasova**, Isabelle Augenstein. *Revealing the Parametric Knowledge of Language Models: A Unified Framework for Attribution Methods.* Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), Long Papers, 2024.
- Sagnik Ray Choudhury, **Pepa Atanasova**, and Isabelle Augenstein. *Explaining Interactions Between Text Spans.* Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), December 2023.
- **Pepa Atanasova**, Oana-Maria Camburu, Christina Lioma, Thomas Lukasiewicz, Jakob Grue Simonsen, and Isabelle Augenstein. *Faithfulness Tests for Natural Language Explanations.* Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), July 2023.
- Momchil Hardalov, **Pepa Atanasova**, Todor Mihaylov, Galia Angelova, Kiril Simov, Petya Osenova, Veselin Stoyanov, Ivan Koychev, Preslav Nakov, and Dragomir Radev. *bgGLUE: A Bulgarian General Language Understanding Evaluation Benchmark.* Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL), July 2023.
- **Pepa Atanasova**, Jakob Grue Simonsen, Christina Lioma, Isabelle Augenstein. *Diagnostics-Guided Explanation Generation.* In Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI), February 2022.
- Wojciech Ostrowski, Arnav Arora, **Pepa Atanasova**, Isabelle Augenstein. *Multi-Hop Fact Checking of Political Claims.* Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI), August 2021.
- Sara Rosenthal, **Pepa Atanasova**, Georgi Karadzhov, Marcos Zampieri, Preslav Nakov. *SOLID: A Large-Scale Semi-Supervised Dataset for Offensive Language Identification.* Findings of the Association for Computational Linguistics (ACL-IJCNLP 2021), August 2021.
- **Pepa Atanasova**, Jakob Grue Simonsen, Christina Lioma, Isabelle Augenstein. *Generating Fact Checking Explanations.* Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), July 2020.
- **Pepa Atanasova**, Jakob Grue Simonsen, Christina Lioma, Isabelle Augenstein (2020). *A Diagnostic Study of Explainability Techniques for Text Classification.* Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), November 2020.

- **Pepa Atanasova**, Dustin Wright, Isabelle Augenstein. *Generating Label Cohesive and Well-Formed Adversarial Claims.* Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2020), November 2020.
- **Pepa Atanasova**, Jakob Grue Simonsen, Christina Lioma, Isabelle Augenstein. *Generating Fact Checking Explanations.* Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), July 2020.
- Slavena Vasileva, **Pepa Atanasova**, Lluis Márquez, Alberto Barrón-Cedeño, Preslav Nakov. *It Takes Nine to Smell a Rat: Neural Multi-Task Learning for Check-Worthiness Prediction.* Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP), September 2019.
- **Pepa Atanasova**, Preslav Nakov, Georgi Karadzhov, Mitra Mohtarami, Giovanni Da San Martino. *Overview of the CLEF-2019 CheckThat! Lab: Automatic Identification and Verification of Claims. Task 1: Check-Worthiness.* International Conference of the Cross-Language Evaluation Forum for European Languages, August 2019
- **Pepa Atanasova**, Georgi Karadzhov, Yasen Kiprov, Preslav Nakov, Fabrizio Sebastiani. *Evaluating Variable-Length Multiple-Option Lists in Chatbots and Mobile Search.* Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR). July 2019.
- Tamer Elsayed, Preslav Nakov, Alberto Barrón-Cedeño, Maram Hasanain, Reem Suwaileh, Giovanni Da San Martino, **Pepa Atanasova**. *CheckThat! at CLEF 2019: Automatic identification and verification of claims.* In European Conference on Information Retrieval (ECIR), April 2019.
- Preslav Nakov, Alberto Barrón-Cedeño, Tamer Elsayed, Reem Suwaileh, Lluis Márquez, Wajdi Zaghouani, **Pepa Atanasova**, Spas Kyuchukov, Giovanni Da San Martino. *Overview of the CLEF-2018 CheckThat! Lab on automatic identification and verification of political claims.* International conference of the Cross-Language Evaluation Forum for European languages (CLEF), August 2018.
- Israa Jaradat, **Pepa Atanasova**, Alberto Barrón-Cedeño, Lluis Márquez, Preslav Nakov. *ClaimRank: Detecting Check-Worthy Claims in Arabic and English.* Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations (NAACL), June 2018.
- Georgi Karadzhov, **Pepa Atanasova**, Preslav Nakov, Ivan Koychev. *We Built a Fake News / Click Bait Filter: What Happened Next Will Blow Your Mind!* In Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP), September 2017.
- **Pepa Atanasova**, Preslav Nakov, Lluis Márquez, AAlberto Barrón-Cedeño, and Ivan Koychev. *A Context-Aware Approach for Detecting Worth-Checking Claims in Political Debates.* In Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP), September 2017.
- Yasen Kiprov, **Pepa Atanasova**, Ivan Koychev. *Generating Labeled Datasets of Twitter Users.* In Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP), July 2017.

JOURNAL ARTICLES

- **Pepa Atanasova**, Jakob Grue Simonsen, Christina Lioma, Isabelle Augenstein. *Fact Checking with Insufficient Evidence.* Transactions of the Association for Computational Linguistics (TACL), Vol 10 (2022).
- Shailza Jolly, **Pepa Atanasova**, Isabelle Augenstein. *Generating Fluent Fact Checking Explanations with Unsupervised Post-Editing.* Information, Vol 13 (2022).
- Luna De Bruyne, **Pepa Atanasova**, Isabelle Augenstein. *Joint Emotion Label Space Modelling for Affect Lexica.* Computer Speech & Language, Volume 71, 2022.
- **Pepa Atanasova**, Preslav Nakov, Lluis Márquez, Alberto Barrón-Cedeño, Georgi Karadzhov, Tsvetomila Mihaylova, Mitra Mohtarami, James Glass. *Automatic Fact-Checking Using Context and Discourse Information.* J. Data and Information Quality 11, (JDIQ), 2019.

PAPERS IN WORKSHOP PROCEEDINGS

- Marcos Zampieri, Preslav Nakov, Sara Rosenthal, **Pepa Atanasova**, Georgi Karadzhov, Hamdy Mubarak, Leon Derczynski, Zeses Pitenis, Çağrı Çöltekin. *SemEval-2020 Task 12: Multilingual Offensive Language Identification in Social Media (OffensEval 2020).* Proceedings of the Fourteenth Workshop on Semantic Evaluation (SemEval), December 2020.
- Tsvetomila Mihaylova, Georgi Karadzhov, **Pepa Atanasova**, Ramy Baly, Mitra Mohtarami, Preslav Nakov. *SemEval-2019 Task 8: Fact Checking in Community Question Answering Forums.* Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval), June 2019.
- **Pepa Atanasova**, Preslav Nakov, Georgi Karadzhov, Mitra Mohtarami, Giovanni Da San Martino. *Overview of the CLEF-2019 CheckThat! Lab: Automatic Identification and Verification of Claims. Task 1: Check-Worthiness.* Sun SITE Central Europe Workshop (CEUR-WS), June 2018.
- Alberto Barrón-Cedeño, Tamer Elsayed, Reem Suwaileh, Lluis Márquez, **Pepa Atanasova**, Wajdi Zaghouani, Spas Kyuchukov, Giovanni Da San Martino, Preslav Nakov. *Overview of the CLEF-2019 CheckThat! Lab: Automatic Identification and Verification of Claims. Task 2: Factuality.* Sun SITE Central Europe Workshop (CEUR-WS), June 2018.
- Tsvetomila Mihaylova, **Pepa Atanasova**, Martin Boyanov, Ivana Yovcheva, Todor Mihaylov, Momchil Hardalov, Yasen Kiprov, Daniel Balchev, Ivan Koychev, Preslav Nakov, Ivelina Nikolova Galia Angelova. *SUper Team at SemEval-2016 Task 3: Building a Feature-Rich System for Community Question Answering.* In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval), June 2016.
- **Pepa Atanasova**, Martin Boyanov, Elena Deneva, Preslav Nakov, Yasen Kiprov, Ivan Koychev, and Georgi Georgiev. *PANcakes team: A Composite System of Genre-Agnostic Features for Author Profiling.* In CEUR Workshop Proceedings (CEUR-WS), June 2016.

BOOKS AND EDITED VOLUMES

- Chen Zhao, Marius Mosbach, **Pepa Atanasova**, Seraphina Goldfarb-Tarrent, Peter Hase, Arian Hosseini, Maha Elbayad, Sandro Pezzelle, Maximilian Mozes. *Proceedings of the 9th Workshop on Representation Learning for NLP (RepL4NLP-2024), August 2024.*
- **Pepa Atanasova**. *Accountable and Explainable Methods for Complex Reasoning over Text.* Last publishing stage of the PhD Thesis as a Springer book in a dedicated series, as part of the Informatics Europe best PhD Dissertation award.

- Venelin Kovatchev, Irina Temnikova, **Pepa Atanasova**, Yasen Kiprov, Ivelina Nikolova. *Proceedings of the Student Research Workshop Associated with RANLP 2017.* September 2017.

DISSERTATIONS

- **Pepa Atanasova**. *Accountable and Explainable Methods for Complex Reasoning over Text.* PhD Thesis, Department of Computer Science, University of Copenhagen, Denmark. September 2022.
- **Pepa Atanasova**. *A Context-Aware Approach for Detecting Worth-checking Claims in Political Debates.* Master's Thesis, Department of Mathematics and Informatics, Sofia University, Bulgaria. October 2017.