

SHADEN SHAAR

Jan 18, 2026
Cornell University
Ithaca, United States

Phone: +1 (646) 286-3453
Email: sshaar31@gmail.com

Research Interests

My research focuses on advancing Natural Language Processing (NLP) techniques for complex textual analysis, specializing in long-form text generation and evaluation methods. My current work explores novel approaches to narrative summarization and document-level question answering systems, with an emphasis on maintaining semantic coherence across extended passages. Previously, I developed automated fact-checking systems and detection mechanisms for propaganda in multi-length documents, contributing to improved information integrity in digital content.

Education

Cornell University, Ithaca, United States 2021–Now
PhD, Computer Science
Minor: Applied Mathematics

Cornell University, Ithaca, United States 2021–2024
Master of Science, Computer Science

Carnegie Mellon University, Doha, Qatar 2015–2019
Bachelor of Science, Computer Science
Minor: Mathematics
University Honors

Work Experience

Applied Scientist Intern at Zillow Groups May 2025–Aug 2025
Natural Language Processing
Research Topics: Conversational AI

- Developed tools for conversational assistive AI tools using RL based approaches

Machine Learning Research Engineer Intern at ScaleAI, Inc. Jan 2025–May 2025
Natural Language Processing
Research Topics: Judicial Artificial Assistant

- Developed tools for judicial case summarization and event extraction
- Developed tools for case and article law retrieval

AI/ML Research Intern at Apple May 2022–Aug 2022
Natural Language Processing
Worked with: Dr. Alex Churchill
Research Topics: Multi-turn Dataset Generation

- Worked on zero-shot multi-turn data generation
- Developed a model to generate multi-turn conversation dataset with conditioning

Research Assistant at Qatar Computing Research Inst., HBKU Jul 2019–Aug 2021
Natural Language Processing
Worked with: Dr. Preslav Nakov, Prof. Giovanni Da San Martino
Research Topics: Automated Fact-Checking; Propaganda Detection

- Worked on automated fact-checking by using previously fact-checked claims
- Developed Fact-checking schema to fight misinformation on COVID-19
- Co-organized multiple tasks in CLEF2020–CheckThat! and CLEF2021–CheckThat! Lab focused on fact-checking
- Built demos for propaganda detection in media
- Co-organized propaganda detection in text/images shared task in SemEval-2021

Research Intern at The Robotics Inst. at Carnegie Mellon Uni. May 2018–Aug 2018

Computer Vision

Advisor: Dr. Christoph Mertz

Research Topic: Lane Detection for Autonomous Vehicles

- Worked on road image segmentation for lane detection and identification
- Explored deep learning and computer vision tools to identify & segment road markings

Part-Time Research Assistant at Carnegie Mellon Uni. May 2017–June 2018

Machine Learning

Advisor: Prof. Bhiksha Raj

Research Topic: Active Testing

- Developed two algorithms that compete with state-of-the-art techniques to estimate the accuracy of a classifier under limited resources
- Proved mathematically & showed empirically the efficiency of the proposed algorithms

Publications

Shaar, S., Thymes, B., Chaixanien, S., Cardie, C., , Hariharan, B. MovieRecapsQA: A Multimodal Open-Ended Video Question-Answering Benchmark *Conference on Computer Vision and Pattern Recognition (CVPR '2026)*

Shaar, S., Chen, W., Chatterjee, M., Wang, B., Zhao, W., Cardie, C. Are *Triggers* Needed for Document-Level Event Extraction? *Transactions of the Association for Computational Linguistics (TACL '2025)*

Chaixanien, S., Choi, E., **Shaar, S.**, Cardie, C. (2024). Pungene at dialam-2024: Identification of propositional and illocutionary relations *Workshop on Argument Mining at ACL 2024 (ACL '2024)*

Aydin, A. N., **Shaar, S.**, Cardie, C. (2024). Edward Said at Touché: Human Value Detection Using Transformers and Upsampling *Touché Lab at CLEF 2024 (CLEF '2024)*

Nakov, P., Alam, F., **Shaar, S.**, Martino, G.D., Zhang, Y. (2022). A second pandemic? Analysis of fake news about COVID-19 vaccines in Qatar *Findings of the Association for Computational Linguistics: NAACL 2022 (Findings of NAACL'2022)*

Shaar, S., Georgiev, N., Alam, F., Martino, G.D., Mohamed, A., Nakov, P. (2022). Assisting the Human Fact-Checkers: Detecting All Previously Fact-Checked Claims in a Document. *Findings of the Association for Computational Linguistics: EMNLP 2022 (Findings of EMNLP'2022)*

Shaar, S., Alam, F., Martino, G.D., Nakov, P., (2022). The Role of Context in Detecting Previously Fact-Checked Claims. *Findings of the Association for Computational Linguistics: NAACL 2022 (Findings of NAACL'2022)*

Hassan, S., **Shaar, S.**, Kareem, D. (2022). Cross-lingual Emotion Detection. *Proceedings of the Thirteenth Language Resources and Evaluation Conference. (LREC'2022)*

Dimitrov, D., Ali, Bishr., **Shaar, S.**, Alam, F., Silvestri, F., Firooz, H., Nakov, P., Martino, G.D. (2021). Detecting Propaganda Techniques in Memes. *Association for Computational Linguistics (ACL-IJCNLP'2021)*

Nakov, P., Alam, F., **Shaar, S.**, Martino, G.D., Zhang, Y. (2022). Findings of the NLP4IF-2021 shared tasks on fighting the COVID-19 infodemic and censorship detection (**ACL NLP4 Workshop 2021**)

Nakov, P., Martino, G.D., Elsayed T., Barrón-Cedeno, A., Míguez, R., **Shaar, S.**, Alam, F., Haouari, F., Hasanain, M., Babulkov, N., Nikolov, A., Shahi, G., Struß, J., Mandl, T.. (2021). The CLEF-2021 CheckThat! Lab on Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News. *European Conference On Information Retrieval. (ECIR'2021)*

Alam, F., **Shaar, S.**, Dalvi, F., Sajjad, H., Nikolov, A., Mubarak, H., Martino, G.D., Abdelali, A., Durrani, N., Darwish, K., Al-Homaid, A., Zaghoulani, W., Caselli, T., Danoe, G., Stolk, F., Bruntink, F., Nakov, P. (2021). Fighting the COVID-19 Infodemic: Modeling the Perspective of Journalists, Fact-Checkers, Social Media Platforms, Policy Makers, and the Society. *Findings of Empirical Methods in Natural Language Processing. (Findings of EMNLP'2021)*

Alam, F., Dalvi, F., **Shaar, S.**, Durrani, N., Mubarak, H., Nikolov, A., Martino, G.D., Abdelali, A., Sajjad, H., Darwish, K., Nakov, P. (2021). Fighting the COVID-19 Infodemic in Social Media: A Holistic Perspective and a Call to Arms. *International Conference On Web And Social Media. (ICWSM'2021)*

Nakov, P., Corney, D., Hasanain, M., Alam, F., Elsayed, T., Barrón-Cedeno, A., Papotti, P., **Shaar, S.**, Martino, G.D. (2021). Automated Fact-Checking for Assisting Human Fact-Checkers. *International Joint Conferences on Artificial Intelligence. (IJCAI'2021)*

Shaar, S., Martino, G.D., Babulkov, N., and Nakov, P. (2020). That is a Known Lie: Detecting Previously Fact-Checked Claims. *Association for Computational Linguistics. (ACL'2020)*

Martino, G.D., **Shaar, S.**, Zhang, Y., Yu, S., Barrón-Cedeño, A., and Nakov, P. (2020). Prta: A System to Support the Analysis of Propaganda Techniques in the News. *Association for Computational Linguistics. (ACL'2020)*, **Best Demo Award (Honorable Mention)**

Shaar, S., Nikolov, A., Babulkov, N., Alam, F., Barrón-Cedeno, A., Elsayed, T., Hasanain, M., Suwaileh, R., Haouari, F., Da San Martino, G. and Nakov, P. (2020). *Overview of CheckThat! 2020 English: Automatic identification and verification of claims in social media. Cappellato et al. (CLEF'2020)*

Barrón-Cedeño, A., Elsayed, T., Nakov, P., Martino, G.D., Hasanain, M., Suwaileh, R., Haouari, F., Babulkov, N., Hamdan, B., Nikolov, A., **Shaar, S.**, Ali, Z.S. (2020). Overview of CheckThat 2020: Automatic Identification and Verification of Claims in Social Media. *(CLEF'2020)* 11: 215–236

Shaar, S., Razak, S., Dalvi, F., and Moosavi, S. (2018). Group Identification in Crowded Environments Using Proximity Sensing. *IEEE Conference on Local Computer Networks. (LCN'2018)*

Hassan, S., **Shaar, S.**, Raj, B., and Razak, S. (2018). Interactive Evaluation of Classifiers Under Limited Resources, *IEEE International Conference on Machine Learning and Applications. (ICMLA'2018)*

Awards & Honors

- Awarded **University Fellowship** in 2021 for exceptional preparation and promise at Cornell University
- Paper awarded **Honorable Mention** at **ACL 2020**
- Graduated with **University Honors** in 2019 for outstanding grade point average at Carnegie Mellon Uni.

- Awarded the **Dean's List** for F15, S16, F16 and F17 at Carnegie Mellon Uni.
- Awarded a **50% Academic Merit Scholarship** in 2015 that is given to few promising students starting at Carnegie Mellon Uni.

Teaching Experience

Teaching Assistant at Cornell University

- CS4740 Natural Language Processing, Prof. Lillian Lee (F22)
- 15-251 Intro to Machine Learning (S22)

Teaching Assistant at Carnegie Mellon Uni.

- 11-785 Introduction to Deep Learning, Prof. Bhiksha Raj (F18 & S19)
- 15-251 Great Theoretical Ideas, Prof. Christos Kapoutsis (S18 & S19)
- 15-213 Introduction to Computer Systems, Prof. Khaled Harras (F17)
- 15-112 Fundamentals of Programming, Prof. Saquib Razak (F16 & F17)
- 21-127 Concepts of Mathematics, Prof. Zelealem Yilma (S16)

Technical Skills

- Programming Languages: Python, C, SML, Assembly
- Deep Learning Toolkits: HuggingFace, PyTorch, Tensorflow, Keras
- Machine Learning Toolkits: sklearn