

# Mourad Heddaya

[mourad@uchicago.edu](mailto:mourad@uchicago.edu) | [mheddaya.com](http://mheddaya.com) | 425-753-5800

## Research Interests

---

### Language processing and LLM evaluation in domains with natural ambiguity.

- Automatically and efficiently extracting and generating useful information from large-scale, messy, and complex real-world data.
- Better understand capabilities and limitations of LLMs, particularly in challenging domains such as long-context summarization and understanding.
- Building compound systems, using LLM agents and feedback mechanisms, to achieve new capabilities.

## Education

---

### Ph.D. Student in Computer Science, 2021-,

*Expected Graduation June 2026.*

*University of Chicago, Chicago, IL.*

*Advisor: Chenhao Tan*

### B.S. in Informatics, 2015-2019,

*University of Washington, Seattle, WA.*

*Research Supervisor: Noah Smith & Mari Ostendorf*

## Selected Publications

---

[LLM training, evaluation, and analysis. Information extraction from real-world complex and noisy data.]

- [Causal Micro-Narratives](#). **Mourad Heddaya**, Q. Zeng, R. Voigt, A. Zentefis, Chenhao Tan. EMNLP 2024 Workshop on Narrative Understanding.
- [Language of Bargaining](#). **Mourad Heddaya**, S. Dworkin, R. Voigt, A. Zentefis, Chenhao Tan. ACL 2023 Main Conference.

[LLM Long-context summarization and evaluation.]

- [CASESUMM: A Large-Scale Dataset for Long-Context Summarization from U.S. Supreme Court Opinions](#). **Mourad Heddaya**, K. MacMillan, Hongyuan Mei, Chenhao Tan, A. Malani. NAACL 2025 Findings. Accepted with talk at [ALEA 2024](#).

## Internships

---

**Applied Scientist at Amazon AWS AI Labs**, Summer 2023. Earned return offer.

[Bedrock Team](#), JFK 14, New York City, NY.

*Mentor: Miguel Ballesteros*

- Proposed self-supervised alignment, an efficient method for aligning LLMs to human preferences for summarization and toxicity without RLHF (without RL and with less human feedback).

- Allow the model to score its own hypotheses (sampled sentences) and incorporate it as self-feedback in the SFT loop, providing more effective regularization for better alignment.
- Project outcome: delivered internal technical report, documented code, and presentation.

## Additional Research Experiences

---

**Research Engineer**, University of Washington, 2020-2021,

*Advisors: Noah Smith, Mari Ostendorf*

- Project outcome: [Unsupervised Learning of Hierarchical Conversation Structure](#)
- Industry collaboration designing and developing unsupervised & supervised information extraction systems to model noisy real-world conversational speech data.
- Based on learned topology from unsupervised HMM, identified distinct conversation paths corresponding to low & high customer service issue resolution, providing insight into successful vs unsuccessful interactions. Final methods and analyses delivered to industry partner.

## Invited Talks

---

**Max Planck Institute for Research on Collective Goods,**

*Research Group Engel, February 2025*

*Talk Topic: NLP In the Legal Domain (summarization, reasoning, etc). Talk to occur in early 2025.*

**University of Chicago,**

*Language Evolution Acquisition & Processing Workshop (LEAP), January 2023*

*Talk Title: Language of Bargaining*

## Service

---

**ARR Reviewer:** June, November 2024

## Teaching Assistantships

---

**University of Chicago,**

*CMSC 25400 – Machine Learning, Winter 2023*

*CMSC 25300 / 35300 – Mathematical Foundation of Machine Learning, Fall 2022*

*CMSC 35100 - Natural Language Processing, Winter 2022*