Nishant Balepur

Email: nbalepur@umd.edu, nishantbalepur@gmail.com
Website: nbalepur.github.io

RESEARCH SUMMARY

I am a computer science Ph.D. <u>candidate</u> at the University of Maryland, advised by Professors Jordan Boyd-Graber and Rachel Rudinger. I mostly work in Natural Language Processing, with a sprinkle of Human-Computer Interaction and Machine Learning. I am broadly interested in building text generation and question answering systems that help users achieve their goals, with a focus on improving factuality, reasoning, preference learning, alignment, and evaluations.

EDUCATION

- **University of Maryland, College Park (UMD)** — College Park, MD
  Ph.D. Computer Science; GPA: 4.00/4.00 — Aug 2023 - May 2027 (Expected)
  Advisors: Jordan Boyd-Graber, Rachel Rudinger
  Thesis (Proposed): Teaching AI to Answer Questions with Reasoning that Actually Helps You
  Committee: Jordan Boyd-Graber, Rachel Rudinger, Shi Feng, Fumeng Yang, David Weintrop

- **New York University** — New York City, NY
  Visiting Student — Aug 2025 - May 2026
  Host: Eunsol Choi

- **University of Illinois at Urbana-Champaign (UIUC)** — Urbana, IL
  B.S. Computer Science; B.S. Statistics (Dual Degree); GPA: 4.00/4.00 — Aug 2019 - May 2023
  Collaborators: Kevin Chen-Chuan Chang, Jiawei Han, Hari Sundaram, Diyi Yang

CONFERENCE PAPERS

* denotes equal contribution, ✝ denotes mentored student

1. Which of These Best Describes Multiple Choice Evaluations? A) Forced B) Flawed C) Fixable D) All of the Above
   ACL 2025, Oral and Best Paper Nomination, Best Paper Award at MASC-SLL 2025
   Nishant Balepur, Rachel Rudinger, Jordan Boyd-Graber
   Selected: We critique multiple-choice evaluations with LLMs, and propose solutions from education to remedy them

2. Whose Boat Does it Float? Improving Personalization in Preference Tuning via Inferred User Personas
   ACL 2025
   Nishant Balepur, Vishakh Padmakumar, Fumeng Yang, Shi Feng, Rachel Rudinger, Jordan Boyd-Graber

3. Reverse Question Answering: Can an LLM Write a Question so Hard (or Bad) that it Can't Answer?
   NAACL 2025, Oral
   Nishant Balepur, Feng Gu, Abhilasha Ravichander, Shi Feng, Jordan Boyd-Graber, Rachel Rudinger

4. MoDS: Moderating a Mixture of Document Speakers to Summarize Debatable Queries in Document Collections
   NAACL 2025, Oral
   Nishant Balepur, Alexa Siu, Nedim Lipka, Franck Dernoncourt, Tong Sun, Jordan Boyd-Graber, Puneet Mathur

5. A Smart Mnemonic Sounds like Glue Tonic: Mixing LLMs with Student Feedback to Make Mnemonic Learning Stick
   EMNLP 2024
   Nishant Balepur, Matthew Shu, Alexander Hoyle, ..., Shi Feng, Seraphina Goldfarb-Tarrant, Jordan Boyd-Graber
   Selected: We use DPO to generate mnemonics aligned with what users prefer and aid learning

6. KARL: Knowledge-Aware Retrieval and Representations aid Retention and Learning in Students
   EMNLP 2024
   Matthew Shu*✝, **Nishant Balepur***, Shi Feng*, Jordan Boyd-Graber

7. Plausibly Problematic Questions in Multiple-Choice Benchmarks for Commonsense Reasoning
   EMNLP 2024 (Findings)
   Shramay Palta, **Nishant Balepur**, Peter Rankel, Sarah Wiegreffe, Marine Carpuat, Rachel Rudinger

8. Artifacts or Abduction: How Do LLMs Answer Multiple-Choice Questions Without the Question?
   ACL 2024, Best Paper Award at MASC-SLL 2024
   Nishant Balepur, Abhilasha Ravichander, Rachel Rudinger
   Selected: We discover LLMs score highly in MCQA benchmarks without the question, and analyze how

9. It's Not Easy Being Wrong: Large Language Models Struggle with Process of Elimination Reasoning
   ACL 2024 (Findings)
   Nishant Balepur, Shramay Palta, Rachel Rudinger

10. Expository Text Generation: Imitate, Retrieve, Paraphrase
    EMNLP 2023, Added to UIUC's AI Reading List for Qualifying Exams
    Nishant Balepur, Jie Huang, Kevin Chen-Chuan Chang

11. Text Fact Transfer
    EMNLP 2023
    Nishant Balepur, Jie Huang, Kevin Chen-Chuan Chang

12. DynaMiTE: Discovering Explosive Topic Evolutions with User Guidance
    ACL 2023 (Findings)
    Nishant Balepur*, Shivam Agarwal*, Karthik Ramanan, Susik Yoon, Diyi Yang, Jiawei Han

Workshop Papers

* denotes equal contribution, ✝ denotes mentored student

1. Language Models Generate Multiple-Choice Questions with Artifacts
   MASC-SLL 2025
   Atrey Desai✝, **Nishant Balepur**, Rachel Rudinger

2. Mental Model for Machine Translation in Human-MT Scenario
   MASC-SLL 2025
   Hyojung Han, **Nishant Balepur**, Jordan Boyd-Graber, Marine Carpuat

3. Is Your Large Language Model Knowledgeable or a Choices-Only Cheater?
   ACL 2024 (KnowLLM Workshop)
   Nishant Balepur, Rachel Rudinger

Preprints and Reports

* denotes equal contribution, ✝ denotes mentored student

1. Can They Dixit? Yes they Can! Dixit as a Playground for Multimodal Language Model Capabilities
   Under Review (EMNLP Wordplay Workshop)
   Nishant Balepur*, Dang Nguyen*, Dayeon Ki*

2. A Good Plan is Hard to Find: Aligning Models with Preferences is Misaligned with What Helps Users
   Under Review (EMNLP)
   Nishant Balepur, Matthew Shu✝, ..., Shi Feng, Fumeng Yang, Rachel Rudinger, Jordan Boyd-Graber

3. Mitigating Gender-Occupation Stereotypes in Llama through Consistency-Guided Finetuning
   Under Review (EMNLP)
   Haozhe An, **Nishant Balepur**, Christabel Acquaye, Rachel Rudinger

4. A Survey on LLM Jailbreaking Attacks and Defenses
   Under Review (AAAI)
   Sander Schulhoff*, **Nishant Balepur***, Arjun Akkiraju*, Michael Ilie*, ..., Rob Voigt, Denis Peskoff

5. The Prompt Report: A Systematic Survey of Prompting Techniques
   Preprint, Huggingface #1 Paper of the Day
   Sander Schulhoff*, Michael Ilie*, **Nishant Balepur**, ..., Shyamal Anadkat, Alexander Hoyle, Phillip Resnik

6. Mastering the ABCDs of Complex Questions: Answer-Based Claim Decomposition for Self-Evaluating LLMs
   Preprint
   Nishant Balepur, Jie Huang, Samraj Moorjani, Kevin Chen-Chuan Chang, Hari Sundaram

Industry Research Experience

- **Allen Institute for AI (Ai2)**   Seattle, WA
  Research Scientist Intern: Semantic Scholar (Personalized QA)   May 2025 - Aug 2025
- **Adobe**   San Jose, CA
  Research Scientist Intern: Document Intelligence (Multi-LLM QA)   May 2024 - Aug 2024
  Outcome: Patent Submission, Paper at NAACL 2025
- **Meta**   Menlo Park, CA
  Software Engineering Intern: Facebook Creators Well-being (Comment Safety)   May 2022 - Aug 2022
  Outcome: 15% drop in negative interactions among Facebook's 1 billion users

Startup Experience

- **LearnPrompting**   Remote
  Contractor: Co-leading research team on LLM jailbreak competitions   Nov 2024 - Present
  Designing harmfulness evaluations for HackAPrompt, lead author on two papers
- **XpertNetwork**   New York City
  Consultant: Technical advice for startup using LLMs for expert network outreach   May 2025 - Present

Fellowships and Grants

- **NSF Graduate Research Fellowship Program (GRFP)** — April 2023 - April 2028
  Wrote proposal on NLP for information accessibility—$159,000 over 3 Years of Ph.D.
- **Cohere for AI Research Grant (2x)** — April 2024
  Wrote proposals on LLMs Education and HackAPrompt 2.0—full access to Cohere models
- **Dean's Fellowship** — April 2023 - April 2025
  Awarded the Dean's Fellowship from UMD for outstanding academic achievement

Invited Talks

- **Which of These Best Describes Multiple-Choice Evaluation with LLMs?** — Vienna, Austria
  ACL 2025 — Aug 2025
- **Can an LLM Write a Question so Hard (or Bad) that it Can't Answer?** — Albuquerque, NM
  NAACL 2025 — May 2025
- **MoDS: Summarizing Debatable Queries in Document Collections** — Albuquerque, NM
  NAACL 2025 — May 2025
- **Which of These Best Describes Multiple-Choice Evaluation with LLMs?** — State College, PA
  Penn State University — April 2025
- **Which of These Best Describes Multiple-Choice Evaluation with LLMs?** — Remote
  Google Translate Reading Group — March 2025
- **Aligning LLMs to Answer Questions with Reasoning That Actually Helps You** — London, U.K.
  Imperial College, London — March 2025
- **How Do LLMs Answer Multiple-Choice Questions Without the Question?** — Baltimore, MD
  Johns Hopkins University — May 2024

Awards

- **MASC-SLL 2025 Best Paper Award** — April 2025
  Best paper award for our MCQA Position Paper by PSU
- **MASC-SLL 2024 Best Paper Award** — May 2024
  Best paper award for "Artifacts or Abduction" by JHU
- **UIUC Computer Science Graduation with Highest Honors** — May 2023
  Recommended by the UIUC computer science department to graduate with highest honors
- **C.W. Gear Outstanding Undergraduate Student** — May 2022
  Awarded to two seniors that have demonstrated excellence in research and service

Professional Service

- **Conference Reviewer** — UMD — 2022-2024
  Reviewer for: *ACL/ARR 2023-Present, IEEE TASLP 2024, TrustNLP 2024, SRW 2025
  Great Reviewer Nomination: Apr, June (#6), Aug (#4), Oct (#2), Dec 2024, Feb 2025 ARR
  Outstanding Reviewer: EMNLP 2024
- **Conference Volunteer** — UMD — 2023, 2025
  ACL 2023, NAACL 2025
- **Visiting Student Day Volunteer** — UMD — Mar 2024, 2025
  Volunteer and ambassador for UMD's visiting student day
- **Winter Storm LLM Workshop** — UMD — Jan 2023
  Led a 5-day workshop on LLMs for non-CS graduate students
- **SIGNLL** — UIUC — Aug 2020 - May 2021
  President of Special Interest Group for Natural Language Learning

Student Mentees

- **Michael Xie** (M.S. UMD), 2025-Present, LLMs for Education

- **Atrey Desai** (B.S. UMD), 2024-Present, LLM Evaluation
  One paper at MASC-SLL 2025 Workshop

- **Matthew Shu** (B.S. Yale → M.S. Yale), 2023-2025, LLMs for Education
  Two papers at EMNLP 2024 (long, main); one paper under review at EMNLP 2025 (long)

- **Jerry He** (HS Student → B.S. GTech), 2024-2025, LLMs for Education
  Submitted Project on LLMs+Education for Marriott's Ridge High School Gifted and Talented Research Program

SKILLS

- **Languages:** Python, JavaScript/HTML/CSS R, C++, Java, OCaml
- **Libraries:** Huggingface, Datasets, TRL, Pytorch, nltk, Spacy, BeautifulSoup