

# Obed Junias

✉ obed.junias@colorado.edu | 🌐 obedjunias.com | 🌐 obed-junias | 📞 720-266-0046

## RESEARCH STATEMENT

---

I study the robustness of multi-step commonsense reasoning in LLMs, focusing on why fluent generation can mask brittle or spurious inference. My work uses explicit intermediate structure, such as reasoning trees and logical abstractions, to design evaluation frameworks that expose hidden reasoning failures beyond surface-level accuracy.

A central direction of my research is understanding how training-data distributions shape observed reasoning behavior. In particular, I plan to study how increasing reliance on model-generated data shifts the effective distribution over time, and how structured evaluations can detect and characterize these shifts.

## EDUCATION

---

**M.S. in Computer Science, University of Colorado Boulder** Aug 2024 – May 2026  
*Courses:* NLP, Deep Natural Language Understanding, Neurosymbolic NLP Methods (GPA: 4.0/4.0)

**B.E. in Computer Science, BMS College of Engineering** Aug 2017 – Aug 2021  
*Relevant Courses:* Algorithms, Databases, Operating Systems, Machine Learning

## PUBLICATIONS

---

- **Junias, Obed\***, et al. *Assessing Algorithmic Bias in Language-Based Depression Detection: A Comparison of DNN and LLM Approaches*. IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), 2025. (Peer-reviewed, Presented). <https://ieeexplore.ieee.org/abstract/document/11269509>
- **Junias, Obed**, and Maria Leonor Pacheco. *LOGICAL-COMMONSENSEQA: A Benchmark for Logical Commonsense Reasoning*. arXiv preprint arXiv:2601.16504, 2026. <https://arxiv.org/abs/2601.16504>

## RESEARCH EXPERIENCE

---

**Structured Commonsense Reasoning and Benchmarking** June 2025 – Present  
*Advisor:* *Maria Leonor Pacheco, BLAST Lab, Associated publication listed above*

- Designed **LOGICAL-COMMONSENSEQA**, a benchmark for logical commonsense reasoning, with structurally composed questions and answer options to isolate multi-fact and compositional inference in LLMs.
- Developed evaluation protocols to analyze failure modes of chain-of-thought and few-shot prompting, identifying cases where fluent reasoning masks incorrect logical inference.
- Designing entailment-like reasoning representations to study alignment between predicted answers and underlying reasoning structure.
- Conducting empirical analyses on how reasoning performance varies with task structure and inference depth, forming the basis of an ongoing research manuscript.

**Responsible LLM Evaluation in Mental Health Applications** Jan 2025 – Nov 2025  
*Advisor:* *Theodora Chaspari, HUBBS Lab, Associated publication listed above*

- Built large-scale evaluation pipelines to assess demographic bias and subgroup performance of LLMs (e.g., LLaMA, GPT-4) for mental health classification tasks.
- Studied how prompting strategies and fairness-aware losses affect reasoning consistency and robustness in low-resource and imbalanced data settings.
- Developed diagnostic metrics to characterize bias and reasoning instability beyond aggregate accuracy.
- Led the research design and analysis for a first-author, peer-reviewed publication presented at BHI 2025.

**Systems for Studying Multi-Agent LLM Behavior** Aug 2025 – Dec 2025  
*University of Colorado Boulder*

- Built a modular inference framework to study how memory constraints and execution order affect multi-agent LLM behavior under realistic deployment settings.
- Compared parameter-efficient adaptation (LoRA) with full-model specialization to analyze feasibility and performance trade-offs under constrained GPU memory.
- Used controlled scheduling policies to examine how system-level factors influence observed reasoning performance, framing systems behavior as a source of evaluation bias.

## AWARDS AND RECOGNITION

---

- **NSF–EMBS–Google Young Professional NextGen Scholar**, *IEEE EMBS BHI Conference*, 2025

## RELEVANT RESEARCH PROJECTS

---

### Lost in Plot: Contrastive Learning for Logical Semantic Retrieval

- Formulated a semantic retrieval task for “tip-of-the-tongue” queries, targeting underspecified and ambiguous natural-language descriptions requiring semantic rather than lexical matching.
- Designed a BERT-based dual-encoder trained with a multi-task contrastive learning objective to align free-form user descriptions with structured movie metadata.
- Constructed a 100K+ synthetic dataset to enable scalable training and controlled evaluation.
- Evaluated retrieval quality using Recall@K and MRR, outperforming a GPT-4 few-shot baseline.

### Medical Ethics Assessment of Large Language Models

- Designed evaluation protocols to assess ethical reasoning of LLMs in medical decision-making scenarios.
- Built a multiple-choice benchmarking pipeline to analyze how external knowledge affects reasoning consistency.
- Analyzed rationales and confidence signals to identify systematic failure modes in ethically sensitive settings.

## INDUSTRY EXPERIENCE

---

### Senior Member of Technical Staff at Oracle Corporation

Aug 2021 - Aug 2024

- Designed and implemented scalable automation and validation frameworks (**TestNG**, **BATS**) for end-to-end testing of **Oracle REST Data Services (ORDS)** and **SQLcl**, reducing manual validation effort by **95%** and improving correctness guarantees.
- Developed a Python-based migration framework to automate deployment and transfer of **Oracle APEX** applications, enabling reliable large-scale migrations and supporting 200+ internal users.
- Led technical validation and delivery of a new Oracle Cloud **DB Tools** feature, coordinating a 15-engineer effort and achieving production readiness within 2 months.

### Machine Learning Engineer Intern at Hewlett Packard Enterprise

Feb 2021 - Jul 2021

- Built OCR-based automation pipelines for processing unstructured enterprise documents, integrating learned document extraction into production workflows.
- Evaluated model performance across noisy and heterogeneous inputs, contributing to improvements in robustness and downstream automation reliability.

## SKILLS

---

<b>Languages</b>	Python, C++, Java, SQL
<b>Machine Learning</b>	PyTorch, HuggingFace Transformers, PEFT/LoRA, Contrastive Learning
<b>LLMs &amp; NLP</b>	Prompting, RAG, Evaluation Frameworks, Reasoning Benchmarks
<b>Systems</b>	Linux, Distributed Systems, GPU Computing, Kubernetes
<b>Data &amp; Tooling</b>	NumPy, Pandas, Redis, PostgreSQL
<b>Research Tools</b>	LaTeX, Git