"What is your timeline for AGI?" The world is captivated by the vision of superintelligent AGI. However, in my research, I am driven by the belief that our priority should be improving the first and only superintelligence today: human intelligence. I explore two questions fundamental to this belief: (1) **How can we build AI systems that enhance human capabilities and benefit all people?** (2) **How can we leverage human insight and creativity to design more robust and trustworthy AI?**

1. BUILDING BETTER AI FOR HUMANS

My work is driven by the belief that AI should be accessible and beneficial to diverse global audiences, transcending language and cultural barriers. Through this principle, I have tackled challenges in multilingual readability, cultural bias, and personalization.

Multilingual Text Simplification and Readability. My first effort in AI to enhance human intelligence focused on making reading more accessible. As an undergraduate, I led the development of MultiSim, a multilingual text simplification benchmark spanning 12 languages and created in collaboration with researchers from 16 institutions [1]. Recognized with an **honorable mention for outstanding paper at ACL 2023**, MultiSim equips NLP researchers with essential tools for multilingual simplification. While working on this project, I identified a gap in assessing the readability of non-English text. This inspired the creation of ReadMe++, a multilingual readability dataset designed to enable automatic readability assessment across diverse languages [2]. These experiences honed my ability to develop impactful NLP resources for social good.

Cultural and Global Biases. Recognizing AI's global reach, I focused on addressing cultural biases in LLMs. Partnering with Tarek Naous and Dr. Wei Xu, I helped develop CAMeL (Cultural Appropriateness Measure Set for LMs), a dataset of culturally relevant Arabic prompts and entities, to measure and quantify cultural biases in AI systems [3]. Awarded Best Social Impact at ACL 2024, this work demonstrated how Western-centric data sources, such as Wikipedia, skew model outputs toward Western norms, often failing to reflect local cultural values. Building on this foundation, I explored the unintended biases in LLM alignment during my master's research with Dr. Diyi Yang. I identified dialect, language, and opinion as key dimensions where tuning for some preferences can inadvertently exclude others [4]. This work revealed the ethical implications of "whose preferences" models serve, showing how U.S.-centric opinions shape reward models.

Next Steps. As solutions, I see two key directions: personalization and community-driven design. Towards personalization I'm currently developing "In-context reward models" (ICRMs) with Dr. Diyi Yang to create lightweight, user-specific reward models, achieving SOTA results on user-stratified Chatbot Arena and PRISM datasets. On the other hand, personalization isn't always the final solution. Sometimes, single models will be used to serve communities. To address this, I am also leading a project with Dr. Diyi Yang and Dr. Josiah Hester to build a bias-auditing framework in collaboration with native Hawaiians, ensuring AI respects local values on sensitive topics. This project will have an immediate impact since Hawaii is integrating Google Gemini into public schools. These initiatives embody my commitment to human-centered AI and its potential to foster meaningful, inclusive technology. I am eager to continue this exploration of both personalization and community-driven design in my PhD.

2. BUILDING BETTER AI WITH HUMANS

Human creativity and insight are essential for building systems that are robust, adaptable, and aligned with real-world needs.

DSPy. I am a core contributor to DSPy, the StanfordNLP library for programming – not prompting LLMs. DSPy's design shifts AI development from complex, monolithic prompt engineering to a modular, human-forward approach, allowing developers to build robust systems through simple, declarative calls. This approach emphasizes high-level planning and system design, encouraging oversight rather than opaque, all-in-one models.

MIPRO. In collaboration with Omar Khattab and Dr. Chris Potts, I co-developed MIPRO (Multi-Instruction PRoposer & Optimizer), the first DSPy optimizer for enhancing both few-shot demonstrations and plain-text instructions [5]. MIPRO improves LLM performance by automating the selection of examples and optimizing instruction clarity, establishing state-of-the-art results across DSPy benchmarks. MIPRO is now widely adopted in industry and academia, including at companies like **Apple, Snowflake, and Databricks** and labs at **Stanford, MIT, Columbia, and UC Berkeley**. By automating prompt improvements while preserving human design principles, MIPRO has helped make modular LM Programs more accessible, efficient, and impactful.

Next Steps. I am excited to introduce DSPy CoPilot, a human-in-the-loop optimization tool that directly empowers domain experts to influence prompt refinements. Human expertise is often essential to ensure accuracy and relevance in complex domains such as medical or legal document processing. With DSPy CoPilot, experts can guide optimizations, fostering systems that reflect technical rigor and practical utility. This approach aligns with transparency and trustworthiness and helps bridge the gap between model optimization and user needs. In addition to optimization, I am researching methods to use minimal human feedback to define automatic evaluation metrics and design novel "LLM-as-a-judge" rubrics for sophisticated tasks. By integrating human feedback into automatic evaluations, I aim to make AI assessments more precise, scalable, and representative of real-world quality standards.

3. My Path Forward

Career Goals I aspire to be a professor of Computer Science and Natural Language Processing, focusing on mentorship, teaching, and research. Over the years, I have honed my teaching skills as a head TA for courses such as "Introduction to Artificial Intelligence" (Georgia Tech) and "Artificial Intelligence Principles & Techniques" (Stanford). I received the "Outstanding Undergraduate TA" award at Georgia Tech and was twice a "Top 5% CA" at Stanford. I aim to extend these skills to leading a research lab, mentoring students, and advancing the field of AI through innovative and socially impactful work.

Why Stanford? Two years ago, I wrote in my master's Statement of Purpose that I wanted to work with Dr. Diyi Yang or Dr. Chris Potts. Today, I am thrilled to say that I have published a first-author paper at a top conference with **both** of these inspiring researchers. Building on this foundation, I am eager to continue learning from Stanford's top talent in NLP. Dr. Yang's work in cultural representation and personalization, such as Culture Bank and DITTO, aligns with my vision of building inclusive, human-centered LLM technology. Similarly, Dr. Potts's focus on DSPy and RAG pipelines resonates with my interest in modular LM systems that integrate human insights. I am also inspired by Dr. Monica Lam's research, particularly projects like WikiChat and Co-STORM, which use human input and modular LMs to create personalized, high-quality content. Overall, I would be honored to continue my studies at Stanford, working collaboratively to push the boundaries of NLP in innovative and impactful ways.

REFERENCES

[1] **Michael J. Ryan**, Tarek Naous, Wei Xu, Revisiting non-English Text Simplification: A Unified Multilingual Benchmark, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (2023). https://aclanthology.org/2023.acl-long.269

[2] Tarek Naous, **Michael J. Ryan**, Anton Lavrouk, Mohit Chandra, Wei Xu, ReadMe++: Benchmarking Multilingual Language Models for Multi-Domain Readability Assessment, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (2024). https://aclanthology.org/2024.emnlp-main.682

[3] Tarek Naous, **Michael J. Ryan**, Alan Ritter, Wei Xu, Having Beer after Prayer? Measuring Cultural Bias in Large Language Models, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (2024). https://aclanthology.org/2024.acl-long.862

[4] **Michael J. Ryan**, William Held, Diyi Yang, Unintended Impacts of LLM Alignment on Global Representation, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (2024). https://aclanthology.org/2024.acl-long.853

[5] Krista Opsahl-Ong*, **Michael J Ryan***, Josh Purtell, David Broman, Christopher Potts, Matei Zaharia, Omar Khattab, Optimizing Instructions and Demonstrations for Multi-Stage Language Model Programs, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (2024). https://aclanthology.org/2024.emnlp-main.525

* denotes co-first authorship