Stanford



Miguel Feles

Graduate, Stanford Center for Professional Development

Bio

BIO

I graduated with honors in Data Science from Universidad Externado de Colombia and am currently taking graduate coursework at Stanford. My interests lie at the intersection of language, machine learning, and human decision-making, with a focus on building reliable AI systems that can learn from and interact with people.

During my undergraduate studies, I conducted research in graph theory and melanoma detection. My thesis explored knowledge distillation with knowledge graph-based verification, aiming to reduce the transfer of factual errors from teacher models. This work shaped my interest in the reliability, factuality, and evaluation of large language models.

I am particularly interested in how language serves as an interface between people and AI systems: how we communicate intentions, express preferences, and provide feedback. My current interests include learning from human preferences, human-AI interaction, and the behavior of AI agents when human preferences are uncertain, conflicting, or difficult to articulate.