

My research derives from a simple philosophy: progress emerges from action, right action stems from understanding, and understanding arises by distilling complex puzzles into simpler components. With this principle in mind, my work focuses on interpreting black-box systems across disciplines. This pursuit eventually led me to the field of Explainable AI, where I explore the interactions between intricate machine learning models and human interpretation. In my recent paper, *[SCENE: Evaluating Explainable AI Techniques Using Soft Counterfactuals](#)*, I introduced a novel, human-centered causal inference method to evaluate Explainable AI techniques in text classification. While dedicating 17-hour workdays during the academic break writing the SCENE paper, I realized the overwhelming fulfillment this work brought me. I knew then that pursuing a Ph.D. in Computer Science was my next step. After my doctorate, I aspire to continue exploring advanced Machine Learning systems and actionable Explainable AI as a university professor. Studying at ----- would allow me to contribute meaningfully to the field and work towards building a world of trustworthy AI systems.

To advance my interest in interpretability and acquire the technical skills needed for impactful research, I joined the M.S. in Applied Data Science program at the University of Chicago. For my master's thesis, advised by Professor Utku Pamuksuz, I conducted a comprehensive empirical assessment of the ten most cited Explainable AI techniques, such as LIME and SHAP, across various neural network architectures in text classification. I found that Saliency methods provide accurate insights while being computationally efficient, whereas Deconvolution ensures high fidelity to the model's behavior. The final results were presented to faculty members at the program's capstone presentation.

While researching evaluation metrics for my thesis, I recognized limitations in existing approaches. Motivated to fill these gaps, I proposed the SCENE method after defending my thesis. SCENE addresses two significant challenges in assessing Explainable AI techniques. First, traditional counterfactual-based approaches rely on flipping the prediction label, which is not always feasible. To overcome this, SCENE introduces the concept of "soft counterfactuals," where the outcome is measured by changes in model output probabilities, while also accounting for input distances. Second, existing erasure-based evaluation methods (e.g., Comprehensiveness) and perturbation-based evaluation methods (e.g., Infidelity) often face issues with plausibility and human comprehension, as altered inputs can fall outside the original data distribution. SCENE tackles this by leveraging language models for token-level substitutions, ensuring that the modified inputs are contextually appropriate and semantically meaningful.

The initial version of my paper attracted interest from the field. The Causal Inference Workshop at KDD 2024 invited me to present in Barcelona, but scheduling conflicts prevented my attendance. I then submitted the paper to conferences with archival tracks to receive rigorous peer-reviewed feedback. Reviewers praised the topic's timeliness, comprehensive literature review, and experimental depth. They also pointed out areas for improvement, such as the article's structure and overly complicated illustrations. I am currently revising the paper to address these concerns and aim to publish the improved version before starting the Ph.D. program in 2025.

The end goal of my SCENE project extends beyond simply proposing a novel metric; it aims to generate actionable recommendations that benefit human users in practical settings. While still in the early stages of development, SCENE has already been applied in conjunction with my current research. As a Research Assistant (RA) for Dr. Christopher A. Reed at the Ohio State University, I utilize modern NLP techniques and the SCENE method to identify the

significance of previously unseen documents, demonstrating how Explainable AI techniques can bring new clarity to traditional historical analysis. I also envision SCENE creating tangible benefits in more complex domains, such as healthcare and Reinforcement Learning. Realizing these ideas will require rigorous experimentation, but the potential progress are exciting prospects.

As a researcher, I embrace out-of-the-box thinking and draw insights from diverse disciplines. Before pinpointing my research interests in Explainable AI, I pursued a multidisciplinary academic journey through undergraduate majors in Finance, Economics, and History, served as a Statistics Teaching Assistant, and gained two years of full-time industry experience. These experiences shaped my understanding of the fundamental challenges in interpretability. In 2024, recognizing the critical role of human-centered perspectives in Explainable AI, I broadened my research focus to Behavioral Science, contributing as an RA in two labs studying human perceptions and decision-making.

At the UChicago Mindworks Lab, I co-manage studies and collect data for renowned scholars like Dr. Nicholas Epley and Dr. Sarah Sebo. Notably, in Dr. Sebo's study, *"Engaging with a Robot to Improve Your Overall Wellbeing,"* I collaborated with the UChicago Human-Robot Interaction Lab to study if participants' personality alignment with robots enhances the benefits. Similarly, at the Epley Lab, I coordinate *"The Multicultural Study,"* examining individuals from different cultures to offer nuanced insights into human cognition. These projects parallel the challenges of Explainable AI—bridging the gap between computational systems and human understanding—and have sharpened my skills in experiment design and human-computer interaction analysis.

I am extremely excited to resume my teaching journey as a doctoral student—a passion that began during my undergraduate studies at The Ohio State University. During this time, I worked as a TA for Statistics 2320 under Professor Bonnie Schroeder. This sophomore-level course, required for over 900 students each semester from diverse backgrounds, involved leading two weekly recitations of 65 students each, covering topics from hypothesis testing to ANOVA. Over four semesters, I honed my skills in conveying complex material to students with varying proficiency levels, utilizing R in active problem-solving sessions, and supporting students with customized review notes and visual aids. In recognition of my dedication, I received the Pace Setters Award, the Fisher College of Business's highest distinction. This reflects my commitment to making complex knowledge accessible—a mission I intend to continue as a PhD student and future professor. I firmly believe in breaking down barriers to knowledge so all students feel empowered to pursue STEM careers.

Beyond academia, while at China Automotive Systems, Inc., I designed a machine learning model that combined tabular and textual data to forecast production and identify potential issues. This full-time industry experience taught me the importance of adaptability—a key quality I aim to bring to my research and interactions within -----'s diverse community.

Outside of my professional life, I enjoy connecting with people through sports. I have played for the Columbus Crew's reserve soccer team, competed as a kickboxer, and contested as a professional e-sports player. These experiences have shaped my ability to build connections and excel as a team player. In the future, I would like to establish a student-led program where current PhD students offer grad school application assistance to underrepresented scholars. By fostering a sense of community and sharing my knowledge and experiences, I aim to support those who seek mentorship.

My personal journey has instilled in me the courage to persevere through profound challenges and defy seemingly impossible odds. In 2017, I was diagnosed with severe depression, facing a year of unbalanced diet, irregular sleep, and suicidal thoughts that made daily functioning a constant struggle. Discussing this period isn't easy, even now, but with the unwavering support of family and friends, I began the slow process of healing. Seeking professional help, I returned to my studies with renewed determination. Over nine semesters, I completed three majors while maintaining a 3.8 GPA. This experience marked a turning point, teaching me invaluable lessons about resilience, the importance of mental health, and the power of a supportive community. Overcoming this adversity didn't just restore me to where I was—it propelled me forward, shaping me into a stronger, more compassionate individual ready to face future challenges head-on. It also deepened my empathy for others facing similar struggles. Now, as an advocate for mental health awareness, I strive to create supportive environments wherever I go. These qualities will empower me to contribute positively to -----'s community and inspire others to confront their challenges with the same confidence and resilience.

The above experiences underscore my motivation to join -----'s Ph.D. program in Computer Science, where its unparalleled resources would allow me to receive regular research guidance from renowned scholars such as Dr. ----- and Dr. ----- . While my research background has been in NLP, my primary passion lies in developing innovative and actionable solutions, regardless of the data type. After taking a graduate-level Bayesian Methods class with Dr. Batu Gundoğlu, I have grown increasingly interested in Bayesian methods and Reinforcement Learning. I recently encountered Dr. -----'s work on Hidden-Parameter Markov Decision Processes and was fascinated by it. I am confident that I can contribute to ----- Lab's mission and help broaden the scope of research in actionable Explainable AI. It would also be an honor to work with Dr. ----- . His expertise in language models would offer valuable insights into complex interpretability challenges, should I choose to further my current research in NLP. Given these considerations, with its unmatched intellectual community, ----- is the premier destination for me to achieve my aspirations in interpretability and create a meaningful impact.