I am Sanchaita Hazra, a fifth-year ABD Ph.D. candidate in Economics at the University of Utah. I have completed the necessary coursework in Spring 2023 covering microeconomics, macroeconomics, behavioral and experimental economics, econometrics, and political economy. My dissertation is titled: **Three Essays on Lies: AI Advisors, Groups with Differing Incentives, and AI-Use Disclosures**. My dissertation committee includes: Haimanti Bhattacharya (UUtah), Subhasish Dugar (UUtah), Gabriel Lozada (UUtah), Daniel Martin (UCSB), and Chris Callison-Burch (UPenn).

In our modern economy, increasingly controlled by artificial intelligence (AI), my research investigates critical questions in **human–AI collaboration**. Broadly, my research is set in the context where generative AI both magnifies and offers new instruments for resolving information friction. Across three settings—**lie detection**, **labor markets**, and **scientific discovery**—my work interrogates how the alignment (or misalignment) of incentives between humans and AI shapes the production, verification, and economic value of information.

I am a quantitative behavioral scientist. Using the methodology of incentivized experiments from behavioral economics, I study **how AI systems influence human decision-making**. I look into how people calibrate their trust and strategically adapt to AI advice, and how incentive and disclosure rules affect individuals' performance and accountability in using AI tools in lie detection, information search, and scientific writing. My research has been recognized and funded by GCSC & The Wilkes Sustainability Fellowship, the J.J. Rasmussen Research Award, and the Haskell Research Award from the University of Utah.

I am also interested in how the appropriate use of **AI can align with the value of human labor**. This is shaped by our recent paper, "AI Safety Should Prioritize the Future of Work", which received *an outstanding paper award* (one of 8/12000 submissions) in the International Conference on Machine Learning (ICML) 2025. Much of my current work looks at how disclosure and audit mechanisms can build trust in AI-assisted decision-making in the hiring market.

Finally, I collaborate on several projects to **build and study AI systems** that assist with human reasoning in economic tasks. For my dissertation, I developed an AI framework to detect linguistic cues from deceptive text. I also collaborate with the Allen Institute for AI on Asta, a project to enable scientific discovery with AI, building and evaluating AI agents for hypothesis generation and data-driven discovery at scale. I pioneered a student-driven user study with an AI agent for data analysis in my Statistics course at the University of Utah that informed optimal UI design for AI agents, trust in AI, and the validity of AI-driven science.

My current work opens pathways to investigate the possibilities and pitfalls of integrating AI in information-rich environments. My future vision aims to illuminate the frictions of AI use in markets that value authenticity, while demonstrating how it can accelerate humanity's accumulation of knowledge.

1 AI and Misinformation

In 2021, Reddit took action against r/NoNewNormal and quarantined 54 subreddits spreading COVID-19 misinformation, anti-vax content, and targeted harassment, leading to millions restricting their accounts. 70% of Americans use social media to connect, engage with news content, share information, and attain entertainment, thus becoming more susceptible to falling into the 'misinformation trap'. Individuals are not good at detecting lies or misleading information, often performing no better than chance. My work on deception detection revolves around a high-stakes, quasi-naturalistic setting, *To Tell The Truth*, an American TV game show. In the game, a set of judges tries to decipher a real character from a set of three challengers, two of whom are impostors, giving rise to conversations filled with lies, half-truths, and true information, similar to conversations on social media.

The ugly effect: AI advice helps, but the lack of transparency hurts further [*Paper 1*, *Paper 2*]

As AI becomes increasingly integrated in our daily lives, we ask whether algorithmic detectors can identify language cues to discern the truth, which otherwise remain challenging for individuals. In our first explorations (*Paper 1*, published in **NAACL 2024**), we demonstrate an LLM-driven bottleneck framework that progressively scans deceptive conversations, analyzes each snippet by verifying utterances against objective truth, semantically understanding complex indicators of deception such as ambiguous responses, half-truths, and overconfidence, and can satisfactorily reason its prediction for detecting deception. Our AI system detects novel but accurate language cues only from linguistic cues in many cases where judges with incentives and audio-visual cues failed to detect deception. In *Paper 2* (**R&R in JBEE**), we use laboratory experiments to find that individuals perform better than random while detecting truth from text . We uncover that when relying on low-quality advisors without disclosures, participants' truth-detection rates fall below their own intrinsic abilities (*the ugly effect*), which recovers once the AI's true effectiveness is revealed. Overall, we find that truth detection from text remains challenging, even with the help of AI. While intuitive, disclosure isn't just about boosting trust; our findings highlight the need to protect users from systematically worse outcomes than if they had ignored AI altogether.

Groups at a disadvantage in detecting lies [*Paper*]

Social media groups (e.g., WhatsApp groups) often play a key role in detecting lies by facilitating fact-checking, collective scrutiny, and spreading or debunking misinformation. In a second study, I explore whether humans can rely on the wisdom of the crowd to help better detect lies. While true, group members often act with varying incentives ranging from genuine truth-seeking to reinforcing biases or commercial interests. I follow up to measure whether

the groups with heterogeneous incentives perform differently at detecting lies than homogeneous groups. Individual accuracy in both individual and group treatments remains close to chance, and I find there was no offsetting benefit from error variance between group members. Large social-media platforms increasingly experiment with wisdom-of-crowd ratings (e.g., Twitter/X "Community Notes"), assuming that when participants have even modest individual insight, majority opinion can match professional fact-checking performance. My experiment identifies the opposite, but equally important, regime: when the underlying content offers vanishing cues to truth—as is common with text-based lies—decision as a group adds no accuracy dividend.

Experience in deceptive environments helps in detecting lies [*Paper*]

Deception by nature is difficult to observe longitudinally in the field. We turn to the original game show, *To Tell The Truth*, and observe how judges perform in this quasi-naturalistic scenario. Originally, a person could appear as a judge in multiple sessions. As judges make many decisions in over 429 sessions, this allows us to observe the effect of experience. In this study, published in **Journal of Economic Criminology**, we find an apparent trend that increased experience reduces average detection error. We econometrically analyze two confounders: keeping a judge fixed, either challenger groups were easier in the latter sessions, or as the game show continued, only high-performing judges were repeatedly selected. Controlling for bias in the selection of judges, we find no evidence for challengers differing in terms of 'quality' across sessions. Using a joint estimation of performance and selection of the judges, evidence in favor of learning was strong with analytical standard errors but weak with bootstrapped standard errors.

Understanding trust in AI as an information source [*Paper*]

Beyond individual or group attempts at discerning truth, deception also interacts with the structure of communication channels themselves. My recent work with Marta Serra-Garcia (UCSD) highlights how biased or incomplete exposure to information can systematically distort beliefs, a mechanism central to the persistence of misinformation in digital environments. This insight complements my recent inquiry into text-based deception and my earlier work on how exposure to deceptive environments helps in lie detection by emphasizing that the challenge is not confined to intrinsic ability but also extends to identifying how information can amplify misleading signals. From a large-scale experiment and the World Values Survey (WVS), we examine the determinants of trust in LLM-generated answers to policy-relevant questions among over 2,900 participants across 11 countries. Trust in the LLM is significantly lower in high-income countries—especially among individuals with right-leaning political views and lower educational attainment—compared to low- and middle-income countries. Using large-scale data on trust from the WVS, we show that patterns of trust in the LLM differ from those in generalized trust but closely align with trust in traditional information sources. This underscores that improving deception detection requires not only better tools for evaluating linguistic cues but also careful attention to the architecture of information environments in which those cues are embedded.

2 AI for Future of Work and Science

LLMs are transforming labor markets by introducing automation capabilities that potentially impact traditional freelance work roles with a significant decrease in writing, coding (21%), and image creation jobs (17%). Equally, there has been an immense momentum in applying AI for scientific discovery. I raise critical questions about the moral and professional hazards of AI use in innovative work.

Future of Work [*Paper*]

There has been an emphasis on evaluating the risks of AI against bio-terrorism, automated warfare, and possible rogue AIs. While pressing, this narrow focus overlooks critical human-centric considerations that shape the long-term trajectory of a society. I argue *AI safety should prioritize the Future of Work*. As the lead author in this position paper, I was honored to receive the **outstanding paper** (one of 8 awardees out of 12000 submissions) award in **ICML** (2025) with co-authors Bodhisattwa Majumder (Ai2) and Tuhin Chakrabarty (Stony Brook). We identify several labor market risks stemming from AI use. Rushed deployments of models, vague claims on model transparency, and unchecked hallucinations contribute to piling technical debt and moral hazard. Unrestrained AI-driven automation accelerates skill disparity, thus disrupting job stability and otherwise smooth intertemporal consumption patterns, thereby creating wage polarization. The extractive, rent-seeking nature of AI and uneven democratization reinforces structural disparities while diminishing worker bargaining power and economic security. Rampant use of AI in education and research supplements self-doubt, increases homogenization of thoughts, and raises non-instrumental image issues. This is further exacerbated by the difficulty of detecting AI-generated content.

We recommend a strong pro-worker framework of global AI governance. Promoting worker interests in AI development aligns naturally with efforts to increase competition and reduce the dominance of big tech companies over the industry's direction. From an AI consumer's perspective, there needs to be accountability on whether something is solely generated by AI, produced entirely by humans, or created with AI assistance. Firms and institutions need to educate themselves on the state-of-the-art for AI detection. Finally, we argue in favor of a robust international copyright anatomy supported by implementing collective licensing that ensures fair compensation mechanisms for using data to train AI models. These findings in labor markets highlight a broader concern: when incentives and transparency are misaligned, AI tools can erode trust and welfare rather than enhance them, the same logic that further motivates me to inspect AI's increasing role in advancing science.

AI in scientific writing [*Paper*]

Scientific publishing is fundamental to the accumulation of knowledge, yet the craft of writing remains a persistent obstacle LLMs now promise to lower writing barriers, but their effects on the credibility and reception of scientific prose are not well understood. Popular tools, including Overleaf plug-ins, are seeing high adoption, while anecdotal cases of AI-generated papers and reviews have raised concerns about declining standards. This tension frames the central question: to what extent do LLMs affect the quality, credibility, and acceptance of scientific writing?

With Sachin Kumar (Ohio StateU), we designed a randomized controlled trial with a hypothetical conference setup where participants with relevant expertise are split into an author and reviewer pool. Our novel incentive structure encourages authors to edit the provided abstracts to an acceptable quality for a peer-reviewed submission. Our 2×2 between-subject design expands into two dimensions: the implicit source of the provided abstract and the disclosure of it. We find authors make most edits when editing human-written abstracts *without* source attribution. Upon disclosure of source information, we find an opposite trend: authors make significantly more edits when the abstracts are known to be AI-generated. Especially, careful stylistic edits in the presence of source information improve the chance of acceptance. We find that LLM-generated abstracts hold potential to reach comparable levels of acceptability to human-written ones with minimal revision, and that perceptions of AI authorship, rather than objective quality, drive much of the observed editing behavior. Our findings reverberate the significance of source disclosure in scientific writing, which also extends to many AI-heavy activities such as programming, design, and editing.

3 Future Vision

My research highlights how the use of generative AI models raises incentive and disclosure problems in labor and academic markets, as employers and evaluators increasingly face imperfect detection of AI assistance. At the same time, in scientific discovery, these same technologies open the possibility of scaling hypothesis generation and verification beyond the limits of traditional intuition-driven approaches. Here, I lay out my broad future vision at the cross-section of Economics and AI.

Data-driven Causal Discovery with Generative AI [*Position Paper*]

The deluge of data collected in the digital age has marked a transformative change in the process and pace of insight discovery. To facilitate future scientific progress to absorb new findings at scale, in a position paper with Allen Institute for Artificial Intelligence (Ai2), we propose a framework to develop automated systems that are capable of continuous ingestion of data, creative generation of hypotheses, and analytical reasoning at a massive scale (published in **ICML 2024**). Turning to the vast amount of observational datasets, both public (e.g., World Bank, National Longitudinal Survey) or private (e.g., clicks on an e-commerce website), social scientists often pose data-driven research questions, presenting a unique opportunity for automatically generating testable hypotheses and observationally verifying them from data. Current state-of-the-art systems for AI-driven data analysis can mine large-scale statistical insights (e.g., "Mass layoffs increase misinformation sharing among middle-income males"), but fail to assert causality in them due to the unobservability of latent factors or the lack of control in observational data. However, using LLMs, we can not only expose hidden factors (e.g., "Under financial stress, individuals disproportionately attend to information that offers either explanation or hope") but also design interventions to experimentally collect data. This further extends to autonomously building a theory or causal mechanism that explains a data-driven statistical insight, verifying through LLM-driven user simulation. With causality-first, we can now prioritize which hypotheses are test-worthy that are not only statistically true but also improve our understanding of the data-generating functions.

Signaling Authenticity vs AI

In all economic sectors, AI widens the highway of automation. But the effects of rampant automation hamper labor markets. As a consequence, the uncertainty about the extent of AI use (often to cheat) by candidates in job applications and interviews is on the rise. Many recruiters now propose clear guidance for hiring companies, urging applicants to clarify AI use during the application process, so candidates know where they stand. However, the implications of such policy prescriptions are yet to be seen. My proposal extends a classic principal-agent framework into AI-use disclosure regulation, where the principal (e.g., firm, instructor, journal editor) allows the agent (e.g., workers, students, reviewers) to *self-select a contract* for AI-use disclosure. Disclosure is rewarded highly if the submitted work is human-produced, but every work with undisclosed AI use is subject to scrutiny with AI detectors. Using incentivized experiments, I plan to capture how a firm's stated preferences for human authenticity may shape the job applicant's use of AI in hiring tasks, and how disclosure rules and audit risks affect the accuracy of self-reported AI use under different preference policies. Can self-selecting a contract type (mandatory disclosure/noisy audit) partition workers by type (high-skill/low-skill) when all work is audited at a negligible marginal cost? The underlying mechanism highlights important policy designs for firms that (i) reward authentic human effort when desired, yet (ii) do not unnecessarily deter relatively cheap but competent AI assistance when authenticity is immaterial.

In summary, sweeping from micro-to-macro designs, I identify behavioral interventions—AI accuracy disclosure, incentives for genuine human work, and algorithmic audit—that improve human-AI alignment, curb moral hazard, and enhance social welfare when humans and machines collaborate. Where natural information asymmetry exists in humans using AI, my findings present actionable insights to minimize the risks and maximize the potential of the technology of our lifetime.