Underreporting of AI Use: The Role of Social Desirability Bias

Yier Ling
University of Chicago
Department of Economics
yier.ling@uchicago.edu

Alex Kale
University of Chicago
Department of Computer Science
Data Science Institute
kalea@uchicago.edu

Alex Imas
University of Chicago
Booth School of Business
alex.imas@chicagobooth.edu

Abstract

Rapid integration of artificial intelligence (AI) into work and educational settings challenges organizations to gauge and respond to adoption rates. However, most measures of AI adoption come from self-reported surveys, producing estimates of AI use that disagree by up to 40 percentage points within the same setting. We investigate whether social desirability bias—the tendency to answer surveys in ways that would be viewed favorably by an outside party—can explain this discrepancy. Surveying 338 university students, we assess potential social desirability bias using a method from psychology, indirect questioning: students report both their own AI use and that of their peers. We find a significant gap, with approximately 60% of students reporting that they use AI compared to 90% of their peers. Through qualitative analysis of student explanations for this gap, we conclude that social desirability bias is a key driver of mis-measurement, causing underestimates of AI adoption in educational settings.

Keywords: artificial intelligence, social desirability bias, self-reports

1 Introduction

Understanding the scale of AI adoption is crucial in order to develop and implement policies and tools promoting beneficial adoption and oversight of AI usage. For example, adoption of AI in organizations without system-wide support can lead to conflicts between interdependent AI-influenced decisions, diminishing operational cohesion (Agrawal et al., 2024). In educational settings, generative AI can significantly enhance efficiency of learning and student engagement or result in unhelpful overreliance depending on how AI use is managed (Salih et al., 2025). However, it remains unclear whether current measurements of AI usage capture true levels of AI adoption. A majority of studies on AI adoption rely on self-reported surveys yielding highly diverging results, sometimes varying by as much as 40 percentage points within the same time frame and setting. For example, in the case of education, a Wiley survey conducted in July 2024 reports that 45% of students had used AI in their classes in the past year (John Wiley & Sons, 2024), whereas the Digital Education Council Global AI Student Survey 2024 reports that 86% of students indicated regular use of AI (Digital Education Council, 2024). *In this study, we explore the hypothesis that this discrepancy may be driven by bias in self-reports, and we demonstrate a corrective method for improving measurement of AI usage.*

We posit that bias in self-reports of AI usage may be driven in large part by ***social desirability bias***, the tendency to answer surveys in a way that is perceived as positive or acceptable by others instead of truthfully (Krumpal, 2013). Misreporting on sensitive topics due to social desirability bias is common in surveys, often driven by a desire to avoid embarrassment in the presence of an interviewer or to prevent potential repercussions from third parties (Tourangeau and Yan, 2007). This bias can manifest as an overreporting of socially favorable attitudes and behaviors, such as compliance with medical advice or adherence to safety measures, and an underreporting of less desirable ones (Miller, 2020). Prior work has shown that social desirability bias may be particularly relevant in educational settings, where students are hesitant to state opinions that may be viewed as undesirable by others (Bowman and Hill, 2011; Nauta, 2007). Such bias can be measured through survey strategies such as ***indirect questioning***, where individuals are asked about others rather than themselves regarding certain behaviors to elicit more truthful responses on sensitive topics (Fisher, 1993). Indirect questioning, as commonly used in social psychology research, reduces bias in reporting because individuals feel more comfortable attributing behaviors perceived as undesirable within their social group to others than themselves. Thus, the method reduces pressure on survey respondents to present themselves in a socially desirable way. As a consequence, the gap between self-reports under direct questioning and peer-reports under indirect questioning should reflect misreporting induced by social desirability bias.

Without accurate measurement of AI usage—and social desirability bias itself—researchers may lack insight into the social dynamics and norms around AI usage in the educational institutions and organizations they serve. One social norm increasingly documented across professions such as research and writing is “AI shaming”, criticism or ostracization driven by beliefs that AI-assisted content is deceitful and lacks creativity (Giray, 2024; Reif et al., 2025). Social image concerns can discourage actual AI use in workplace settings, even when such concerns have little to do with AI’s instrumental impact on work performance (Almog, 2025). In the context of education, the use of AI tools is also linked to concerns about ethics, academic honesty, plagiarism, and trustworthiness (Kostopolus, 2025; Cotton et al., 2024; Rodrigues et al., 2025). For students, declaring AI use may be seen as an admission of not being able to complete the work independently, potentially leading to a perception of lower academic ability (Gonsalves, 2024). In other social settings such as firms where there is high pressure to adopt AI, social desirability bias could result in a pattern of “AI washing", overreporting one’s own use to appear more savvy or satisfy expectations (Barrios et al., 2024; Al Haddi, 2024; Simonian, 2025; Miller et al., 2024). Effective policy interventions and AI integration into workflows and software requires awareness and understanding of these social dynamics and norms—e.g, AI policy and tools should be responsive to whether a culture of shame reflects legitimate concerns versus irrational prejudice, or whether a culture of hype reflects market pressures versus real benefits. We demonstrate how a combination of indirect questioning and qualitative analysis can provide this critical perspective for researchers, policy makers, and tool builders.

We contribute two studies that explore social desirability bias around AI use among undergraduates. First, we run a large representative survey of students at a medium-sized Midwestern university, where students are asked detailed questions about their own AI use (direct questioning) and the same questions about the AI use of peers within their social circle (indirect questioning). We find a dramatic gap in AI usage: whereas approximately 60% report using AI tools themselves, this number increases to nearly 90% when asked about the use of others. The gap between own and other AI use appears in nearly every dimension of AI use in the survey. In the second study, we collect data from a separate sample of college students on Prolific to explore how they would explain possible reasons for the gap between own and other AI use. Through qualitative analysis of free-text responses, we find that the vast majority of students attribute the gap to people underreporting their own use relative to others. When choosing among plausible explanations for the gap in a forced-choice question, by far the most selected response is that *Students are embarrassed to admit they use LLMs, but are okay saying that their friends do.* Our results demonstrate how indirect questioning can contribute to more accurate knowledge of the social norms against AI adoption reflected in these responses. We conclude that these social norms seem to be strongly established in organizational cultures such as education. We argue that failure to account for how these cultural norms bias measures of AI adoption can spur under- or overreliance in ways that undermine the transformative potential of AI.

2 Background

Artificial intelligence (AI) is no longer a futuristic concept but an increasingly pervasive technology impacting numerous facets of daily life. Firms across multiple industries are reported to have an AI adoption rate between 20% to 40%, growing rapidly over time (Crane et al., 2025). In a 2024 in-depth survey of self-reported AI use, almost 40 percent of 18–64 year olds use AI in some capacity, with 9 percent reporting using it every day at work(Bick et al., 2025). Such AI adoption has impacted numerous industries and labor markets (Tyson and Zysman, 2022). Focusing on the public sector, Tveita and Hustad (2025) review literature portraying the benefits of AI adoption in terms of efficiency and cost-saving, as well as challenges in terms of trust and ethics. Their review finds that through automation of repetitive tasks, AI can allow employees to focus on enhancing public services. AI adoption is also shown to motivate employees by increasing the sense of task importance (Henk and Nilssen, 2021). On the other hand, AI-led outcomes can be deemed less reliable and trustworthy due to the lack of transparency in the process (Gualdi and Cordella, 2021; Asatiani et al., 2021). Other concerns over AI adoption are ethical in nature, where the privacy of training data and prompts, the gap between theoretical guidance and practical implementations in terms of reliability and safety, and bias in the algorithms or their training data could lead to potential harm (Wei and Zhou, 2023; Siqueira De Cerqueira et al., 2021; Asatiani et al., 2021).

2.1 AI in Education

Many have specifically highlighted the potentially transformative role of AI in education. The integration of AI tools in education is becoming more pronounced, with a significant proportion of both teachers and students acknowledging the essential role of AI tools in achieving success in college and future careers (Li and Towne, 2025). Some prior research highlights the potential benefits such as enhanced learning experiences, performance, and teaching methodologies (Thomson et al., 2024; Dai et al., 2023; Vartiainen and Tedre, 2023; Mazzoli et al., 2023; Cardona et al., 2023; Salih et al., 2025; Al-Nsour, 2024). For example, AI's ability to generate various forms of media could aid learning by providing perspectives hardly achievable for traditional teaching tools, across disciplines such as medicine and engineering (Mazzoli et al., 2023; Lan and Azimi, 2025). AI can also serve as a personalized tutor, complementing and sometimes even surpassing human tutors in providing customized feedback (Bassner et al., 2024; Cardona et al., 2023; Salih et al., 2025; Dai et al., 2023). Specifically, Dai et al. (2023) finds that AI can provide more consistent, readable, and process-focused feedback to student assignments compared to instructors at a graduate level. Qu et al. (2025) similarly shows that AI tools are able to enhance understanding and application of knowledge by reducing cognitve load and providing prompt feedback. Such advantages of AI usage have been documented in many disciplines, such as ESL, programming, dentistry, and pharmacy (Lyu et al., 2024; Mahapatra, 2024; Ododo et al., 2024; Svendsen et al., 2024; Kavadella et al., 2024). Besides serving as a personalized tutor, AI tools can also be integrated with other technology, such as educational gamification, to adaptively and automatically generate materials and assessments for students (Velazquez-Garcia et al., 2025). Morever, AI can also improve students' experience outside of learning: for instance, AI has been documented to support logistic and repetitive tasks for students, such as scheduling and planning, and sometimes even fulfill a social support role when needed (Heyer et al., 2025; George et al., 2025).

Nonetheless, others note the threats of AI use in education. Reliance on AI can act as an impediment to students' creativity, critical thinking, independence, and higher-level abstraction (Darvishi et al., 2024; Yan et al., 2024; Ododo et al., 2024; Jošt et al., 2024; Urban et al., 2024), as students may be accustomed to "outsourcing" thinking. Apart from student academic ability, AI in education can also adversely impact students' ethics and conduct, as well as social ability. Plagiarism and normalization of mediocrity may be amplified in the presence of widespread AI content (Deng and Joshi, 2024). Students may also become over-reliant on AI and avoid crucial learning processes, such as the direct query of facts and solving problems themselves (Huang and Chang, 2025). Relying on AI as a computer-based tool can foster anxiety in students, for example, in interpersonal settings (George et al., 2025), and also impede students from developing emotional and empathetic abilities (Huang and Chang, 2025). Another concern around AI use in education is the reliability of AI-generated content. Accuracy of AI-generated eductional content is often in doubt (Schefer-Wenzl et al., 2025), and such AI-generated content can also reflect bias embedded in their training data (Deng and Joshi, 2024). Velazquez-Garcia et al. (2025) further discusses concerns around ineffective personalization of AI, raising doubts about personalization as an advantage of AI in education. This underscores the need to accurately assess the extent to which people are adopting and utilizing these technologies.

Given the promise and perils of AI adoption, effective AI adoption in education requires institutional support, which prior work suggests are currently inadequate (Videla et al., 2025). Numerous papers indicate the importance of establishing appropriate guidelines and systems around AI use, including the establishment of norms (Agrawal et al., 2024; Salih et al., 2025; Jin et al., 2025). Kim and Wu (2024) showcase inconsistencies in existing AI policies across institutions and call for efforts to build a systematic and standardized framework for AI integration analyses and guidelines. Institutions should dedicate resources to adaptively refine AI policies and build AI literacy among faculty and students (Jin et al., 2025), taking into account both teacher and student perspectives (Heyer et al., 2025; Yin et al., 2025). In practice, a blended model combining human educators and AI tools is proposed by many reseachers. AI can take on tasks such as customizing learning plans, explaining concepts, and lower-level assistance, while human educators pivot towards emphasizing aspects such as creativity and critical thinking (Salih et al., 2025; Feng et al., 2025; George et al., 2025). AI tools in education should also be carefully chosen according to student's learning objectives and the model's ability (Cardona et al., 2023). More concretely, George et al. (2025) proposes that educators could guide students' AI use towards its most productive purposes, discouraging AI usage specifically for outsourcing homework and exams while allowing AI for logistic and lower-level tasks. Educators are also advised to adjust the format, focus, and evaluation of assigments, taking how students use AI into consideration (Rajabi et al., 2023; Feng et al., 2025). Specifically, Petrovska et al. (2024) proposes tasks encouraging students to critically evaluate AI outputs, potentially building AI literacy in students without fostering AI reliance. *This motivates the need to accurately measure and understand the current AI adoption, as an foundation for any analysis.*

2.2 Measuring AI Adoption

Current measurements and the future potential of AI adoption remains unclear. The central concern of our work is how widely varying measures of adoption may be influenced by social pressure around AI use. "AI shaming", involving beliefs that AI-assisted content is deceitful and lacks creativity, has increasingly been documented across several occupations such as researchers and academic writers (Giray, 2024). Research has shown that social image concerns discourage actual AI use in workplace settings, even when such concerns are not instrumental to work performance (Almog, 2025). Evidence from large-scale experiments also corroborate that individuals employing AI at work both predict and receive negative judgments from others (Reif et al., 2025). In the context of education, the use of AI tools is also linked to concerns about ethics, academic honesty, plagiarism, and trustworthiness (Kostopolus, 2025; Cotton et al., 2024; Rodrigues et al., 2025; Heyer et al., 2025). Consequently, declaring AI use may also be seen as an admission of not being able to complete coursework independently, potentially leading to a perception of lower academic ability (Gonsalves, 2024). These reactions to AI adoption can cultivate norms to hide or avoid using AI, rendering self-reported AI adoption statistics suspect and dampening the potential of AI integration. We seek to validate this hypothesis and demonstrate a way to measure these social dynamics.

3 Methods

We ran two surveys to investigate whether social desirability bias impacts students' self-reported AI usage. The first survey aims to capture AI usage among a sample of students representative of the college campus, using both direct and indirect questioning to see how social desirability bias may impact on reported results. The second survey aims to explore the mechanism behind the observed reporting gap in the first survey, validating social desirability bias as a cause and providing some alternative perspectives from a second sample of undergraduate students.

3.1 Representative student survey on AI Use

3.1.1 Participants

We recruited undergraduate students at medium-sized Midwestern university in collaboration with its behavioral research center, by reaching out to students on campus via multiple methods (e.g., flyers in campus buildings, volunteers contacting students on campus, posting survey links on local marketplace websites). Student participants were given a $5 gift card for completing the survey, and a university email address was required for the gift card. After survey data was collected, participants who indicated that they were not undergrads (i.e., either reporting that they are graduate students or choosing "Other" in terms of their year in college without providing more details) or that they are studying majors not offered by the university were removed. The final remaining sample size was 338.

Among the 338 participants, around 76.3% of them reported at least one STEM major; 77.5% of the sample has at least one STEM major or minor. Participants were about evenly distributed across years in college, ranging from 20.7% in the third year to 28.1% in the first year. Among participants who reported their gender identity, slightly over half of the sample identify as female, while around 4.2% of the participants identify as Non-binary or other gender. Among participants reporting their racial identity, Asian students make up the largest racial group in this sample (36.1%), followed by White students (33.5%), Other and Multi-racial students (20.7%) and African American students (8.8%), with less than 1% of Native American or Pacific Islander students. A non-trivial minority of the sample was composed of international students (26.6%). Among participants reporting their family income, 43.5% of participants report an annual household income over $150,000, followed by $25,000 to $50,000 (12.5%) and $75,000 to $100,000 (12.2%). Distributions of these demographic variables are also presented in Figure 9. This sample is broadly representative of undergraduate population at the university where the survey is conducted.

3.1.2 Survey Design

This survey consists of two major components examining the participant's own use of AI (in the form of LLMs) and their beliefs about others' usage of LLMs. After the welcome page, all participants were asked if they ever used an

LLM. If they answered “yes”, they are then directed to answer a series of additional questions about LLMs: how reliant they are on LLMs; which LLM models they use; whether they pay for LLM models; main reasons for using LLMs; how many days a week they use LLMs; what other AI tools they use (if any); and whether they use the built-in features of LLMs. These questions constitute the self-reported own use part of the survey. Each participant also answered the same set of questions about average peers in their social group (not any specific person). Notably, we asked two questions about participants’ peers’ AI reliance: (i) a single-choice question, where participants choose one of five levels of reliance as options, identical to how they evaluated their own reliance; and (ii) a distributional question, where participants report what percentage of their peers fall into the same five levels of reliance. The distributional question was impossible to ask for own-reliance since they are assessing the use of only one person rather than a distribution of people.

Finally, participants complete a set of demographic questions asking about: their majors (and minors if applicable); year in college; their study devices; whether they are international students; whether English is their first language; their ideal prospects after college; how much they work in major-related or major-unrelated fields part-time during a week (if any; in two separate questions); gender; family income; and race.

3.1.3 Analysis Method

We ran logistic regression analyzing whether or not students report AI use conditional on whether the report is about Own vs Other AI use, participant gender, whether or not they are in a STEM-major, and the interactions between these three factors. We also performed bootstrapping, where we resample with replacement and repeat the estimation, to capture estimation uncertainty.

We aggregate the ordinal responses about AI reliance or frequency to a binary outcome by assigning 1 to AI use if the reliance level is not “None at all" or the frequency level is not “0-1 days". We perform this aggregation focusing on the lowest level (“None at all"; “0-1 days") because: (i) it straightforwardly highlights the stark difference between self- and peer- reported AI usage; (ii) a binary divide is less prone to different interpretations of reliance versus frequency in participants; and (iii) we found a logistic model was better able to fit the empirical distribution of our data than a Bayesian ordinal regression analysis (see Appendix B).

3.2 Exploring the Mechanism

3.2.1 Participants

For the follow-up study exploring the mechanism behind the gap between own and other AI use, we recruited a separate group of participants from *Prolific*. We restricted the sample to those who reported being an undergraduate student—all others were screened out. The final sample was 96 participants, which was sufficient to reach codebook saturation when performing qualitative analysis on student explanations for the gap.

3.2.2 Survey Design

In this survey, we first described to participants the findings from the initial student survey, showing both the percentage of students’ own self-reported reliance and frequency of AI use as well as their answers about peer use. All participants were then asked three questions about this gap. The first question was a free-text entry directly asking for their explanation of the gap. The second question asked whether they believed this gap was due to the overreporting of others’ use or the underreporting the students’ own use. The last question asked participants to choose the most likely explanation: students being embarrassed to admit LLM use; students inflate others’ use; students are unaware of others’ use; and none of the above.

3.2.3 Analysis Method

Our primary analysis for the second survey involved qualitative coding of participants' free text responses to categorize their explanations for the gap between own versus other AI use. We used lightweight thematic analysis to group participants' free text responses into emergent categories. The first author made an initial pass of open coding to describe participants' proposed explanations. Then, all authors convened to discuss these open descriptions and agree upon inductive codes proposed by the first author. The first author applied inductive codes to categorize each participant's free text response, iterating a few times and consulting other authors as needed to resolve ambiguous cases. We report on the final set of inductive codes describing participant explanations and their prevalence in our sample.

For the other two single-choice questions, we provide simple summary statistics reflecting the fractions of participants choosing each option.

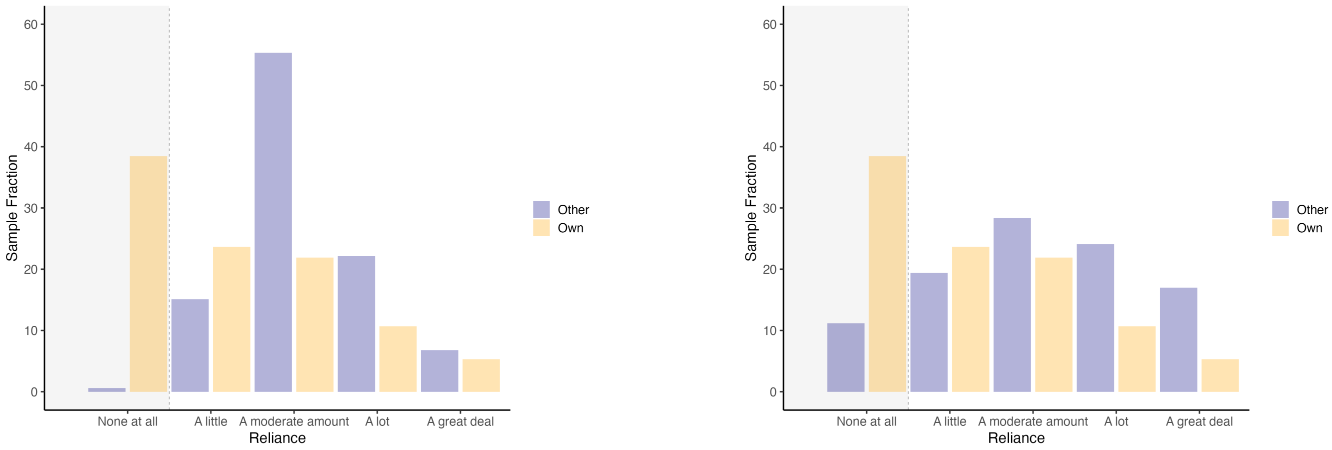
4 Results

4.1 Representative Student Survey on AI Use

We aim to evaluate the difference between self- versus peer- reported AI usage in college students. All students in the sample answered a series of questions about their own and others' reliance on AI, as well as questions on demographics and major. We focused on AI in the form of Large Language Models (LLMs), which are by far the most prevalent form of AI in the setting we studied. Questions thus referred to the use of LLMs, and also included queries on the specific models used (e.g., ChatGPT 4o, Claude, etc.). The full set of questions is listed in the Methods section. Our main questions of interest are (i) Do you/peer ever use LLMs (none at all, a little,...,a great deal; *reliance*) and (ii) the frequency of LLM use per week (0–1 days, 2–3 days,..., 6–7 days; *frequency*).

Recall from section 3.1.2 we ask participants about their own and their peers' behaviors around AI. Specifically, we ask participants to evaluate their own reliance using a single-choice question, and their peers' reliance using both a single-choice and a distributional question. In the single-choice questions, participants are asked to choose between themselves/their peers relying on AI “None at all", “A little", “A moderate amount", “A lot", or “A great deal"; in the distributional question, they are asked what percent of others falls into each of the categories listed in the single-choice questions. These questions correspond to (i) in the above paragraph, which we denote as *reliance*. We separately ask about AI use frequency using two single-choice questions. Participants are asked whether they/their peers use AI “0-1 days", “2-3 days", “4-5 days" or “6-7 days" a week. These questions correspond to (ii) in the above paragraph, which we denote as *frequency*. We focus on these two measures as indicators of reported AI usage.

Figures 1 and 2 present descriptive statistics on the reporting gap in our sample regarding both measures of AI usage. Figure 1a shows the difference in levels of AI reliance for Own and Other when participants were asked to choose one reliance level for both themselves and their peers (single-choice), while Figure 1b shows the difference when participants gave distributions of their peers' AI reliance across all levels (distributional), where the average distribution across participants is presented. Shaded regions suggest no AI use, highlighting the reporting gap of interest in contrast with non-shaded regions of the x-axis. We observe a stark difference between self- and peer-reported outcomes across all measures—substantially more participants report that they do not use AI than what they report for their peers. The gap ranges from 27.3% to 41.1% across measures. We also provide sample heatmaps of the Own and Other reports in Appendix A.1.



(a) Reliance - Single Choice

(b) Reliance - Distribution

Figure 1: Reliance reports for both self-reports and peer-reports. 1a presents single-choice reports for others' AI use, and 1b presents the average distributional reports for others' AI use across participants.

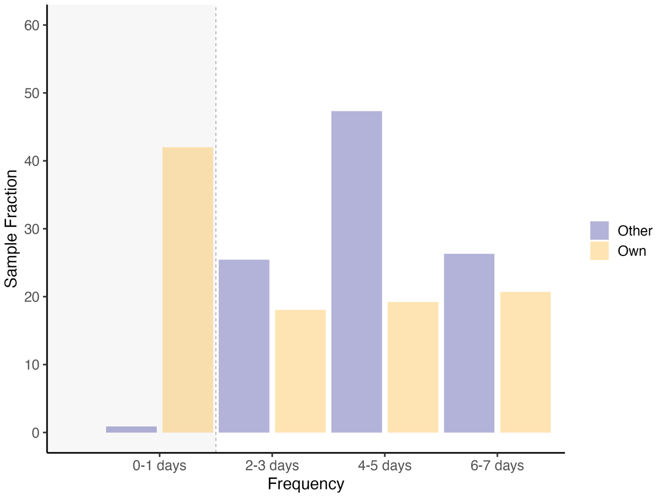


Figure 2: Frequency reports for both self-reports and peer-reports.

4.1.1 Heterogeneity Analysis with LogisticRegression

Section 3.1.3 presents our quantitative analysis method for the representative survey results using logistic regressions. Table 1 shows the similarity between the two measures of AI usage after aggregation of ordinal responses into binary categories of use and no use. In both regressions, Own is the only regressor significantly impacting the probability of reported AI usage (0.1% level). This impact suggests that participants are significantly less likely to report AI usage for themselves than peers, aligning with a SDB-driven reporting gap. Figures 3, 4, and 5 present the distributions of model predictions compared to sample fractions using AI, while Figures 6, 7, and 8 shows the same results using AI frequency. Distributions of model predictions are obtained from 5,000 rounds of bootstrapping iteration, where in each iteration we randomly resample participants from the survey with replacement and repeat the estimation and prediction procedures. Error bars denote the 95% intervals of the distribution across the 5,000 iterations. Reporting gaps are preserved in all figures: we always observe higher predicted probabilities of AI use in others, regardless of participant demographics.

Table 1: Logistic regression coefficients for two models fit to (1) reliance responses and (2) frequency responses, respectively. This table demonstrates that results are robust to the choice of measure.

Dependent Variables: Model:	Reliance (1)	Frequency (2)
Own	-3.720***	-3.720***
	(1.059)	(1.059)
Male	14.76	13.76
	(1,262.3)	(765.6)
STEM	1.184	1.184
	(1.432)	(1.432)
Own × Male	-14.47	-13.47
	(1,262.3)	(765.6)
Own × STEM	-0.7165	-0.9157
	(1.473)	(1.472)
Male × STEM	-1.184	-14.21
	(1,430.5)	(765.6)
Own × Male × STEM	1.067	14.15
	(1,430.5)	(765.6)
Constant	3.807***	3.807***
	(1.017)	(1.017)
Fit statistics		
Observations	632	632
Squared Correlation	0.22697	0.24266
Pseudo R^2	0.29028	0.29091
BIC	487.61	508.99

IID standard-errors in parentheses
*Signif. Codes: ***: 0.001*

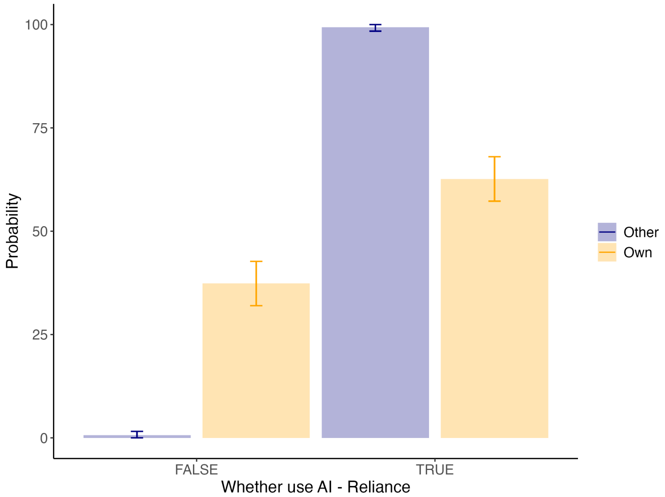


Figure 3: Reliance - Logistic Results. Error bars denote 95% interval across 5,000 bootstrap iterations.

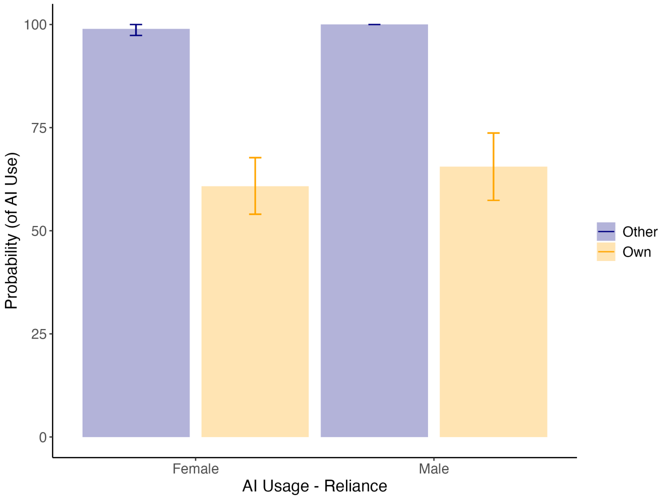


Figure 4: Reliance - Logistic Results by gender. Error bars denote 95% interval across 5,000 bootstrap iterations.

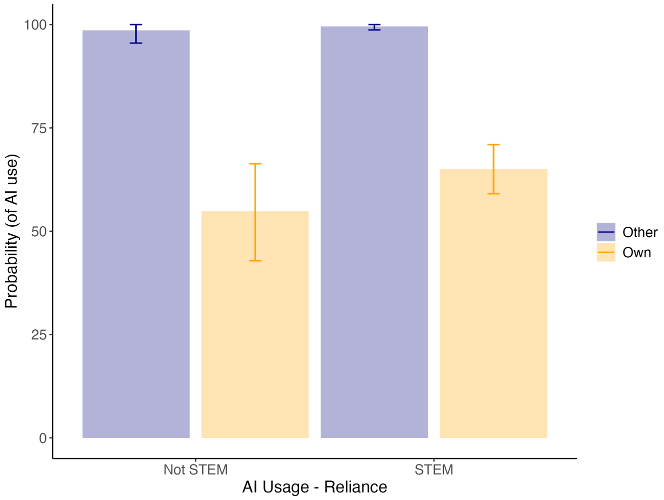


Figure 5: Reliance - Logistic Results by STEM Major. Error bars denote 95% interval across 5,000 bootstrap iterations.

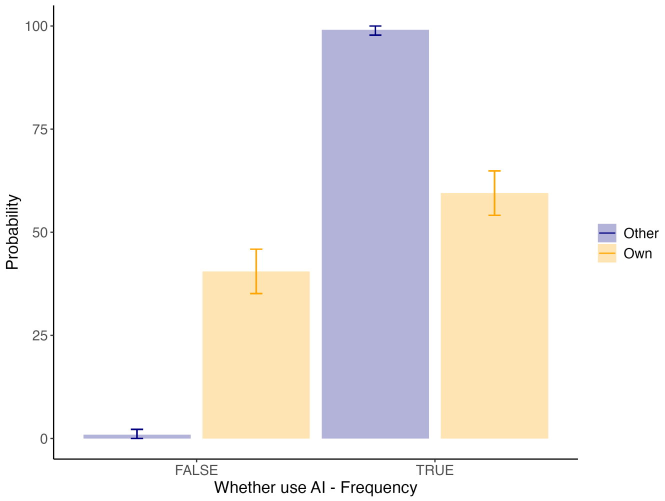


Figure 6: Frequency - Logistic Result. Error bars denote 95% interval across 5,000 bootstrap iterations.

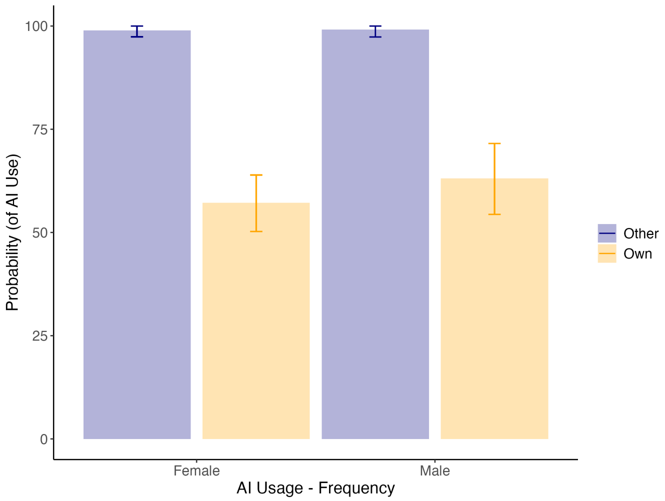


Figure 7: Frequency - Logistic Results by gender. Error bars denote 95% interval across 5,000 bootstrap iterations.

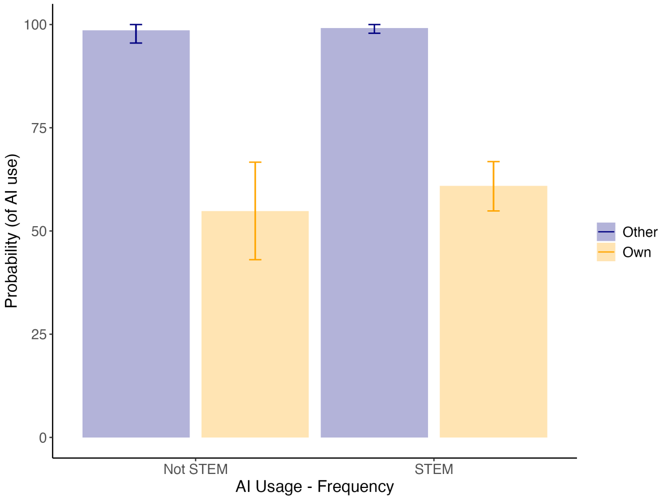


Figure 8: Frequency - Logistic Result by STEM Major. Error bars denote 95% interval across 5,000 bootstrap iterations.

4.2 Exploring the Mechanism

The results from the representative student survey paint a clear picture of a large gap between self-reports of own and others' AI use. We now proceed to explore the potential reasons for this gap. In a follow-up survey of students, we presented information about this gap and elicited reasons for it both through direct questions and free response. As mentioned in section 3.2.3, we analyse the single-choice questions using straightforward summary statistics, while providing a qualitative analysis for the free text responses.

4.2.1 Single-choice Questions

Table 2 lists the specific questions asked and the percent of people who select each answer. The first set of direct questions asks whether the own vs. other AI use gap is due to students underreporting their own use or overreporting the use of others. The vast majority (79%) report that the gap is due to underreporting of own use. When asked to provide a reason for the gap, by far the largest category of answers is consistent with social desirability bias (70%); the next largest category ("Students don't know how much their friends are using LLMs") received less than one third as many responses.

Table 2: Proportions of participants reporting reasons and mechanisms behind the own/other AI use gap in single-choice questions.

Question	Percent	Response
Reason	70.8	Students are embarrassed to admit they use LLMs, but are okay saying that their friends do
	20.8	Students have no idea how much their friends are using LLMs
	6.3	Students are inflating how much others are using LLMs
	2.1	None of the above
Under- vs. Over-reporting	79.2	Students under-report how much they use LLMs
	11.5	Students over-report how much others use LLMs
	9.4	Neither

4.2.2 Free response

We manually review responses from Prolific on students' beliefs about why a gap between own and other AI use exists. Among 96 respondents, 71 provide valid responses; 25 are removed for providing nonsensical responses (e.g., random integers), or for implying that they misunderstood or were confused by the question (e.g., explaining a reporting gap in the reverse direction).

Social Desirability Bias SDB is the most mentioned rationale behind the reporting gap, suggested by 43 respondents. 7 respondents explicitly point to SDB (P3, P43, P46, P56, P60, P65, P71), supplemented with more detailed explanations. For example, P46 says *"The disparity is most likely caused by social desirability bias"*, which they defined as *"students underreport their own LLM usage to look more academically honest"*. This captures the essential mechanism behind SDB: to be academically honest is socially desirable in education, hence students "hide" their actual AI usage which is oftentimes deemed dishonest. P3 concurs this point by suggesting they underreport own usage *"due to fear of being judged as being dishonest or lazy"*. Additionally, P71 emphasizes the social norms aspect of SDB: *"Students might not report using LLM if they think that academic organizations look down on it."* Students avoid reporting their own socially undesirable behaviors that external forces disapprove of.

Other responses reflect the same rationale without explicitly mentioning SDB. 18 respondents point to the idea that students underreport because they do not wish to admit using AI. Some respondents specifically suggest that the same SDB is much less likely to apply to peer reports: for example, *"[social norms around AI usage] has made individuals to not want to disclose fully their use of AI due to possible stigmatization. The same individuals won't have a problem reporting how their friends use the AI."* (P62), and *"College students themselves are not as willing to admit to their LLM usage, but are fine with discussing their friends' usage."* (P40). Both arguments emphasize another aspect of SDB—students care more about social perceptions and desirable behaviors about themselves than their peers, as they typically bear the burden of how their peers are perceived to a lesser extent, especially regarding individual conduct in education.

A varied range of sources of social desirability associated with reporting AI usage are provided by the respondents as well. Academic integrity concerns, such as casting AI usage as cheating or dishonest, is most frequently mentioned (15/71). Apart from appearing dishonest or less credible as previously discussed, some respondents explicitly mention that using AI can sometimes be deemed a violation of academic integrity: for instance, P17 wrote *"For most schools, getting caught using AI for homework can get students expelled from their program at school."*, along with 4 other respondents explicitly mentioning that AI usage could potentially be seen as cheating. These suggest the wide-spread and substantial aversion to students' AI usage—that it is not only socially undesirable, but also misconduct to be punished. More broadly, respondents also expressed how AI usage can be socially undesirable in institutions: *"[There are] concerns about academic integrity or stigma around relying on AI for assignments"* (P28), and *"They think that academic organizations look down on it"* (P71). Meanwhile, respondents also provided other reasons behind SDB for underreporting own AI usage: a sense of embarrassment and shame (5/71); appearing lazy (4/71); appearing dishonest or unethical (8/71); stigma surrounding AI adoption (7/71); and appearing less academically capable (6/71).

Overestimating Others' Usage The second most commonly mentioned rationale behind the reporting gap is overestimation of others' usage (22/71). Among these participants, multiple respondents suggest that this overestimation could come from biased perception: students are "primed" by AI reliance in certain groups and over-generalize it to all other students (9/71). For example, *"I think only a small portion of students actually rely on LLMs to do coursework, while most students do not. That small portion leads some students to assume most are using it."* (P14), and *"Students might assume their friends use LLMs more frequently because AI generated content is widespread, and discussions about AI tools are common."* (P28). These arguments implicitly point to a manifestation of the availability bias, which is explicitly mentioned by P65: *"The availability heuristic also plays a role; hearing about classmates using LLMs makes it seem widespread."* ***Availability bias***, as previously established by Tversky and Kahneman (1973), suggest that students can have upward-biased beliefs of how likely their peers use AI if they can more easily observe and recall other students using AI. P7 also suggests that students are more prone to notice others' usage than their own, potentially exacerbating this bias. Another respondent similarly suggests this availability bias can worsen with network effects: by observing *"certain friend groups have more of use of LLM than others"* (P30), students in more AI-reliant networks can more easily recall their peers using AI and perceive AI adoption in other students to align with their networks. Consequently, students with this upward-biased beliefs would form assumptions and speculations overestimating other students' AI usage, as mentioned by 5/71 participants.

This availability bias does not necessarily need to emerge from a student observing actual peers using AI. 4 respondents also mention how debates or discussions on AI usage, not necessarily in students' own social circles, could contribute to such overestimation. P16 says, *"It is likely due to how much people talk about AI nowadays."* P44 attributes overestimation of others' use to *"social media portraying AI use, lack of understanding of how much they use it"*, and P46 mentions *"Students may mistake their awareness of LLM debates or availability for actual usage, distorting their views of others' conduct."* All three comments indicate that discussions around AI usage alone, not necessarily actual use observed or experienced by students themselves, can also "prime" them with availability bias. Another respondent (P28) further mentions how wide-spread AI content, not necessarily any discussion about AI use patterns, could similarly aid in such availability-based overestimation: *"Students might assume their friends use LLMs more frequently because AI generated content is widespread, and discussions about AI tools are common."*

Apart from availability bias, a few additional rationales behind overestimation are provided by the respondents. P33 suggest students believe other students rely on AI more to protect their self-esteem (P33)—unlike SDB involving social perceptions, this statement suggests an internal force from the students to overestimate others' AI usage, such that their self-image is not harmed by using AI since "others use it more than me". Another respondent suggests a possible SDB in the opposite direction as a cause for overestimation: *"Some students tend to say they use LLMs to sound intellectual and trendy but they're lying."* (P12). This suggests students can be misinformed directly by their peers who pretend to be "tech-savvy" by claiming to be AI-reliant. This could reflect an opposite SDB in the pretending peers—they exaggerate what they perceive as socially desirable, using AI, to improve social perceptions about themselves, appearing trendy.

Information Asymmetry The third most mentioned reason is information asymmetry, where students are unaware of other students' actual AI adoption (19/71). Responses in this category include examples such as, *"They simply don't know much about what their friends do on the daily basis or how much their friends use LLMs to be exact,"* (P2), and *"Their friends may not disclose to college students how much they use LLMs,"* (P6), pointing to the lack of accurate information about other students' AI usage. P64 suggests this as an outcome of *"people do not talk about how much they use LLMs"*. A few respondents (7/71) point out how this lack of information could facilitate overestimation of AI adoption—e.g., *"Perception gaps cause students to misjudge their peers' behavior based on assumptions rather than direct knowledge,"* (P43) and *"The gap could be because we don't see each other use LLMs, we have to speculate, and we are more likely to speculate radically than moderately."* (P39). Combined with the aforementioned availability bias, students are more likely to assume and speculate that other students use AI more than the factual benchmark.

Underestimating Own Usage Another rationale behind the reporting gap is that students may be underestimating their true AI usage (5/71). Note that we differentiate underestimating from underreporting, where the latter suggests the students are aware of their usage but choose to report lower usage. For example, P27 says *"Many of us underestimate our use of AI"*, and P49 says *"Students may underestimate their own use of LLMs"*. These statements are also consistent with P7's abovementioned statement: students are more prone to notice otthers' AI usage than their own.

Self-esteem 3 respondents (P6, P33, P51) also indicate that students may underreport their AI usage to protect their self-esteem, a motivation similar to SDB but different in caring about students' self-image. This can also be seen as another side of the self-image motive for overestimating others' AI usage—by admitting to lesser AI usage, students create a self-image that are more AI-independent and capable. P6 suggests *"They may feel ashamed or subpar compared to their peers if they tell others they use LLMs for their work."* P33 says the students are *"protecting their self-esteem by disguising their true ai usage,"* and P51 indicates that students may underreport due to *"probably a sense of denial and underlying embarrassment...it suggests that their own work is not up-to-par with the professional levels...and that they as students may not be as well adept, intelligent, and skilled/talented."* These statements point out feelings of inferiority may drive students to underreport their own AI usage, a factor that does not involve social perception per se but self-evaluation, which we categorize as distinct from SDB.

Privacy Concerns 2 respondents (P17, P25) mentions privacy concerns as a reason behind the underreporting of own AI usage. P17 suggests the reporting gap is due to the *"prospect of getting caught"*, and P25 says *"the gap could be explained by a combination of personal privacy, social perceptions, and varied usage patterns."* Both suggest students underreport their own AI usage to avoid adverse outcomes by keeping their true AI usage private. This could suggest that students generalize stigma and even punishments around AI usage and carry extra caution to contexts outside student evaluation, corroborating the previously discussed academic integrity concerns.

Self-reporting Bias 1 respondent (P5) suggests that the underreporting can be an outcome of students exhibiting a self-favoring bias when reporting. Specifically, P5 says *"We like to judge others for their use and our perceptions of others is jaded when they tell us they use an outside source to help them on assignments, as we view it as a crutch. Yet we use the same crutch, most of us are self-reflective."* This suggests that students use different evaluating standards for themselves versus other students, and the standards are more in favor of themselves. As a consequence, the same level of factual AI usage may appear different to students when they evaluate themselves. We distinguish this from underestimating one's own AI usage, in that this reasoning suggests that students are aware of how much they use AI, but choose to judge their own usage more leniently or favorably. We also distinguish this bias from SDB, as social factors may not be needed for students to judge themselves more favorably.

Truthful Reporting 2 respondents (P20, P42) believe the reporting gap accurately reflects a gap in AI usage between survey respondents and their peers, and the students were truthfully reporting. For example, P20 says, *"The people taking the surveys don't use AI as much as the people that they know do."* While this remains a possibility, we do not agree that it is truly what drives the reporting gap. Since our survey ask representative sample of students on campus (without targeting specific students) and vaguely ask about their peers without specifying any individuals, one participant could be a "peer" to another participant. In other words, this survey design should impose that the students surveyed and their peers come from the same "population" distribution in terms of AI usage. If the reporting were truthful, it would suggest the surveyed students are substantially different from the rest of the students in AI use, and they are not considered as "peers" to the other surveyed students, which we deem implausible as an explanation.

5 Discussion

The findings of this study reveal a substantial disparity between students' self-reported AI usage and their perception of peer AI usage. We provide evidence that this gap is primarily driven by social desirability bias, where norms against AI use lead people to underreport their own usage relative to others. The significant discrepancy observed in this study has important implications for the accuracy of current statistics on AI usage. Especially in settings such as education and in private firms—where social desirability bias is likely to distort self-report measures—it may be difficult for organizations to develop strategic policies around AI adoption that are responsive to actual use. We discuss how the indirect questioning method demonstrated in this study might be useful for elucidating and responding to the social norms about AI use within different organizational cultures.

Our findings suggest that measurement error in self-reports of AI usage is so large that many current statistics on AI adoption are likely not credible. Specifically, we find up to a 40% difference in AI adoption when college students are asked about their own AI use versus that of their peers. However, the magnitude and direction of this measurement error are likely to vary across settings. For example, unlike in educational settings where students seem to underreport AI usage by a large margin, workers in private firms work may be incentivized to overreport AI usage due to labor market pressure to adopt AI or due to social pressure among peers. In organizational cultures where social desirability bias plays a role in discourse about AI, researchers cannot rely on naive self-report to track adoption rates. Other measures of AI use, such as software companies reporting the number of lines of code written with AI assistance in the last quarter Novet and Vinian (2025), may be distorted by similar forms of social desirability bias.

Student perspectives on social pressure around AI usage reveal a high degree of stigma around AI adoption in educational settings. Many survey participants point to concerns about social perception of their ability, academic honesty, and feelings of shame as primary reasons why a college student might underreport their own AI use. A smaller subset of participants attribute the gap in estimates own versus other AI use to overestimation of rates of AI adoption among peers due to availability bias Tversky and Kahneman (1973), stemming from economic pressures and hype around AI. Although our study design cannot fully distinguish underestimation of own use versus overestimation of other use, we conclude that the gap is driven by a sort of moral panic about AI adoption. For students, who might see productivity gains or improved job market outcomes from learning to use AI—and for universities, whose reputations are in part staked on these outcomes—this stigma poses a barrier to appropriate reliance on AI and effective adoption.

The indirect questioning methodology demonstrated in this study provides a corrective against the mis-measurement of AI adoption through self-report methods. We show how a mix of direct and indirect questioning can be used to uncover the magnitude and direction of social desirability bias around AI use in organizational settings. Although this alone does not reveal the true rate of AI adoption, it can reveal social dynamics within an organization that might be disruptive to strategic planning about AI use. We show how supplementing this method with lightweight qualitative analysis can elucidate the nature of social desirability bias within a firm—in our case, we interpret the qualitative evidence to suggest that students likely underreport their own AI use. We discuss how this approach can be useful for planning organizational interventions (e.g., workplace trainings).

5.1 Mitigating the Risks of Mis-Measurement for AI Policy

The mis-measurement of AI adoption create a major obstacle to institutional planning and response. For example, in universities, our findings suggest a need for policies to address stigma around AI use. The depth of this problem would be unknown without a way to measure social desirability bias. We consider how our approach to uncovering social desirability bias might be useful for planning organizational policies around AI in three settings.

In an educational setting, we find strong evidence of stigma against AI use, such that students report the use of their peers is much higher than their own use. Recognizing this gap, university leaders must ask when this stigma is warranted due to real ethical concerns, and conversely, when it presents a dogmatic barrier to students' upward mobility. Universities should aim to develop strategic initiatives and training programs that encourage faculty and teaching assistants to distinguish academic honesty violations from appropriate reliance on AI. For example, training programs could focus on helping teaching staff identify positive examples of responsible AI use, provide students with opportunities to practice such use cases, and otherwise avoid demonizing AI use that does not actually constitute an ethical lapse. At the same time, policies prohibiting inappropriate reliance should focus on addressing problematic behavior rather than committing fundamental attribution error by casting students themselves as lazy or unscrupulous.

Although our findings do not speak to non-educational settings, it stands to reason that social desirability bias will manifest in the opposite pattern of survey responses in firms where AI use is celebrated, such that survey respondents would report their own use of AI is greater than their peers' use. This could happen if a company or institution provides incentives for AI adoption or if employees in an industry believe that they need to appear savvy with AI in order to maintain a competitive edge in the labor market. Again, measuring these social dynamics through a mix of direct and indirect questioning presents such organizations with an opportunity to formulate a strategic response depending on the perceived reasons for social desirability bias.

First, consider a scenario where AI use is celebrated despite it conferring little to no instrumental value to productivity or worker satisfaction. This hypothetical represents the case where pressure to adopt AI results mostly from hype. In such cases, social desirability bias reflects a culture of overreporting driven by status anxiety within an organization or industry. An effective policy intervention in this case would aim to encourage employees to use AI only where doing so is prudent and would signal that non-adoption will not be viewed unfavorably by management. For example, a company could train managers to avoid costly misadventures in AI use (e.g., replicating existing work processes with a lower degree of reliability) and could eliminate internal incentives (e.g., raises, promotions) for AI adoption.

Conversely, consider a second scenario where AI use is rightly celebrated within a firm. This hypothetical represents the case where AI boosts employee performance and job satisfaction in ways that benefit the firm as a whole. In such cases, social desirability bias would reflect the striving of employees to demonstrate their commitment to or excitement about positive changes within the firm. Here, there is no need to mitigate social desirability bias, as it reflects well-calibrated expectations about the instrumental value of AI use. Instead of policies of risk mitigation, organizations in this position would benefit from openly celebrating and promoting AI use internally. For example, companies could offer raises or promotions for effective uses of AI. However, companies in this position should still exercise caution that promoting AI use as a performance target does not lead to irresponsible or costly uses as a way of gaming incentives.

5.2 Limitations

Our study recruits a sample of college students from a mid-sized Midwestern University. Although we recruit a representative sample at this institution, the socioeconomic conditions at this university may not be representative of all educational contexts. For example, our sample skews toward students from more affluent families relative to the student body at a typical state or community college. This may affect social desirability bias in ways that our study was not designed to address. Similarly, as described above, the social desirability bias measured in our study will not replicate in all organizations and contexts—e.g., the direction of the observed gap between own versus other AI use may reverse in firms where AI use is celebrated rather than stigmatized. Future work should replicate our survey methodology across a wide range of institutions to paint a clearer picture of heterogeneity of social desirability bias across contexts.

6 Conclusion

We contribute two surveys on AI usage among college students. By adopting an indirect questioning method we observe a large gap between students' self-reported AI use and their estimates of peer use. Qualitative analysis of student explanations for this gap suggest that social desirability bias, induced by stigma around AI adoption, causes students to underreport their own AI usage by a substantial margin. Given that mis-measurement of AI adoption may disrupt strategic efforts to respond to AI adoption in organizations, we offer recommendations on how institutions and firms can make use of indirect questioning methodology in planning and implementing trainings and policies around AI.

References

Agrawal, A., J. S. Gans, and A. Goldfarb (2024). Artificial intelligence adoption and system-wide change. *Journal of Economics & Management Strategy 33*(2), 327–337.

Al Haddi, H. (2024). Ai washing: The cultural traps that lead to exaggeration and how ceos can stop them. *California Management Review Insights*.

Al-Nsour, R. (2024, October). Ai tools in matlab course education: Instructor point of view. *Journal of Computing Sciences in Colleges 40*(2), 95–104.

Almog, D. (2025). AI recommendations and non-instrumental image concerns. *Available at SSRN 5226740*.

Asatiani, A., P. Malo, P. R. Nagbøl, E. Penttinen, T. Rinta-Kahila, and A. Salovaara (2021). Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the association for information systems 22*(2), 325–352.

Barrios, J. M., J. L. Campbell, R. G. Johnson, and Y. C. Liu (2024, August). Signals or smoke? the determinants and informativeness of corporate artificial intelligence (ai) disclosures. Available at SSRN.

Bassner, P., E. Frankford, and S. Krusche (2024). Iris: An ai-driven virtual tutor for computer science education. In *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1*, ITiCSE '24, New York, NY, USA. Association for Computing Machinery.

Bick, A., A. Blandin, and D. J. Deming (2025). The rapid adoption of generative AI.

Bowman, N. A. and P. L. Hill (2011). Measuring how college affects students: Social desirability and other potential biases in college student self-reported gains. *New Directions for Institutional Research 2011*(150), 73–85.

Cardona, M. A., R. J. Rodríguez, and K. Ishmael (2023). Artificial intelligence and future of teaching and learning: Insights and recommendations. Technical report, U.S. Department of Education.

Cotton, D. R., P. A. Cotton, and J. R. Shipway (2024). Chatting and cheating: Ensuring academic integrity in the era of chatgpt. *Innovations in education and teaching international 61*(2), 228–239.

Crane, L., M. Green, and P. Soto (2025). Measuring AI uptake in the workplace.

Dai, W., J. Lin, H. Jin, T. Li, Y.-S. Tsai, D. Gašević, and G. Chen (2023). Can large language models provide feedback to students? a case study on chatgpt. In *2023 IEEE international conference on advanced learning technologies (ICALT)*, pp. 323–325. IEEE.

Darvishi, A., H. Khosravi, S. Sadiq, D. Gašević, and G. Siemens (2024). Impact of AI assistance on student agency. *Computers & Education 210*, 104967.

Deng, X. N. and K. Joshi (2024, August). Promoting ethical use of generative AI in education. *SIGMIS Database 55*(3), 6–11.

Digital Education Council (2024). What students want: Key results from dec global AI student survey 2024.

Feng, T. H., A. Luxton-Reilly, B. C. Wünsche, and P. Denny (2025). From automation to cognition: Redefining the roles of educators and generative AI in computing education. In *Proceedings of the 27th Australasian Computing Education Conference*, ACE '25, New York, NY, USA, pp. 164–171. Association for Computing Machinery.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of consumer research 20*(2), 303–315.

George, A., V. C. Storey, and S. Hong (2025, February). Unraveling the impact of chatgpt as a knowledge anchor in business education. *ACM Trans. Manage. Inf. Syst. 16*(1).

Giray, L. (2024). AI shaming: the silent stigma among academic writers and researchers. *Annals of Biomedical Engineering 52*(9), 2319–2324.

Gonsalves, C. (2024). Addressing student non-compliance in AI use declarations: implications for academic integrity and assessment in higher education. *Assessment & Evaluation in Higher Education*, 1–15.

Gualdi, F. and A. Cordella (2021). Artificial intelligence and decision-making: The question of accountability. In *54th Annual Hawaii International Conference on System Sciences*, pp. 2297–2306.

Henk, A. and F. Nilssen (2021). Can AI become a state servant? a case study of an intelligent chatbot implementation in a scandinavian public service. In *54th Annual Hawaii International Conference on System Sciences*, pp. 5515–5524.

Heyer, C., E. M. Nilsson, and J. Pedersen (2025). Ai-assisted learning in hci education: Opportunities and dilemmas from a student perspective. In *Proceedings of the 7th Annual Symposium on HCI Education*, EduCHI '25, New York, NY, USA. Association for Computing Machinery.

Huang, H.-W. and J. C.-Y. Chang (2025). Human-ai interactions in teacher education: Examining social presence and friendship. In *Proceedings of the 2024 International Conference on Artificial Intelligence and Teacher Education*, ICAITE '24, New York, NY, USA, pp. 64–69. Association for Computing Machinery.

Jin, Y., L. Yan, V. Echeverria, D. Gašević, and R. Martinez-Maldonado (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence 8*, 100348.

John Wiley & Sons, I. (2024). Ai has hurt academic integrity in college courses but can also enhance learning, say instructors, students.

Jošt, G., V. Taneski, and S. Karakatič (2024). The impact of large language models on programming education and student learning outcomes. *Applied Sciences 14*(10), 4115.

Kavadella, A., M. A. D. Da Silva, E. G. Kaklamanos, V. Stamatopoulos, K. Giannakopoulos, et al. (2024). Evaluation of chatgpt's real-life implementation in undergraduate dental education: mixed methods study. *JMIR Medical Education 10*(1), e51344.

Kim, D. and J. Wu (2024). Artificial intelligence in higher education: Examining the AI policy landscape at u.s. institutions. In *Proceedings of the 25th Annual Conference on Information Technology Education*, SIGITE '24, New York, NY, USA, pp. 68–73. Association for Computing Machinery.

Kostopolus, E. (2025). Student use of generative AI as a composing process supplement: Concerns for intellectual property and academic honesty. *Computers and Composition 75*, 102894.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & quantity 47*(4), 2025–2047.

Lan, R. and E. Azimi (2025). Ai-driven microelectronics education using digital twins and extended reality. In *Proceedings of the Great Lakes Symposium on VLSI 2025*, GLSVLSI '25, New York, NY, USA, pp. 492–497. Association for Computing Machinery.

Li, D. H. and J. Towne (2025, Jan). How AI and human teachers can collaborate to transform education.

Lyu, W., Y. Wang, T. Chung, Y. Sun, and Y. Zhang (2024). Evaluating the effectiveness of llms in introductory computer science education: A semester-long field study. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale*, pp. 63–74.

Mahapatra, S. (2024). Impact of chatgpt on esl students' academic writing skills: a mixed methods intervention study. *Smart Learning Environments 11*(1), 9.

Mazzoli, C. A., F. Semeraro, and L. Gamberini (2023). Enhancing cardiac arrest education: exploring the potential use of midjourney. *Resuscitation 189*.

Miller, K., M. Brown, and C. Calvino (2024). Ai washing erodes consumer and investor trust, raises legal risk. *Bloomberg Law*. Analysis of AI washing regulatory enforcement and business implications.

Miller, P. R. (2020). Tipsheet – sensitive questions.

Nauta, M. M. (2007). Assessing college students' satisfaction with their academic majors. *Journal of career assessment 15*(4), 446–462.

Novet, J. and J. Vinian (2025). Satya nadella says as much as 30% of microsoft code is written by ai.

Ododo, E. P., N. P. Essien, and A. E. Bassey (2024). Effect of generative artificial intelligence (ai)-based tool utilization and students' programming self-efficacy and computational thinking skills in java programming course in nigeria universities. *International Journal of Contemporary Africa Research Network 2*(1).

Petrovska, O., L. Clift, F. Moller, and R. Pearsall (2024). Incorporating generative AI into software development education. In *Proceedings of the 8th Conference on Computing Education Practice*, CEP '24, New York, NY, USA, pp. 37–40. Association for Computing Machinery.

Qu, X., J. Sherwood, P. Liu, and N. Aleisa (2025). Generative AI tools in higher education: A meta-analysis of cognitive impact. CHI EA '25, New York, NY, USA. Association for Computing Machinery.

Rajabi, P., P. Taghipour, D. Cukierman, and T. Doleck (2023). Exploring chatgpt's impact on post-secondary education: A qualitative study. In *Proceedings of the 25th Western Canadian Conference on Computing Education*, WCCCE '23, New York, NY, USA. Association for Computing Machinery.

Reif, J. A., R. P. Larrick, and J. B. Soll (2025). Evidence of a social evaluation penalty for using AI. *Proceedings of the National Academy of Sciences*.

Rodrigues, M., R. Silva, A. P. Borges, M. Franco, and C. Oliveira (2025). Artificial intelligence: Threat or asset to academic integrity? a bibliometric analysis. *Kybernetes 54*(5), 2939–2970.

Salih, S., O. Husain, M. Hamdan, S. Abdelsalam, H. Elshafie, and A. Motwakel (2025). Transforming education with ai: A systematic review of chatgpt's role in learning, academic practices, and institutional adoption. *Results in Engineering 25*, 103837.

Schefer-Wenzl, S., C. Vogl, S. Peiris, and I. Miladinovic (2025). Exploring the adoption of generative AI tools in computer science education: A student survey. In *Proceedings of the 2024 16th International Conference on Education Technology and Computers*, ICETC '24, New York, NY, USA, pp. 173–178. Association for Computing Machinery.

Simonian, J. (2025). Ai washing: Signs, symptoms, and suggested solutions for investment stakeholders. Technical report, CFA Institute Research Policy Council. Research report addressing ethical concerns and risks of AI washing in finance.

Siqueira De Cerqueira, J. A., L. Dos Santos Althoff, P. Santos De Almeida, and E. Dias Canedo (2021). Ethical perspectives in ai: A two-folded exploratory study from literature and active development projects.

Svendsen, K., M. Askar, D. Umer, and K. H. Halvorsen (2024). Short-term learning effect of chatgpt on pharmacy students' learning. *Exploratory Research in Clinical and Social Pharmacy 15*, 100478.

Thomson, S. R., B. A. Pickard-Jones, S. Baines, and P. C. Otermans (2024). The impact of AI on education and careers: What do students think? *Frontiers in Artificial Intelligence 7*, 1457299.

Tourangeau, R. and T. Yan (2007). Sensitive questions in surveys. *Psychological bulletin 133*(5), 859.

Tveita, L. J. and E. Hustad (2025). Benefits and challenges of artificial intelligence in public sector: A literature review. *Procedia Computer Science 256*, 222–229.

Tversky, A. and D. Kahneman (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology 5*(2), 207–232.

Tyson, L. D. and J. Zysman (2022). Automation, AI & work. *Daedalus 151*(2), 256–271.

Urban, M., F. Děchtěrenko, J. Lukavský, V. Hrabalová, F. Svacha, C. Brom, and K. Urban (2024). Chatgpt improves creative problem-solving performance in university students: An experimental study. *Computers & Education 215*, 105031.

Vartiainen, H. and M. Tedre (2023). Using artificial intelligence in craft education: crafting with text-to-image generative models. *Digital Creativity 34*(1), 1–21.

Velazquez-Garcia, L., A. Cedillo-Hernandez, M. D. P. Longar-Blanco, and E. Bustos-Farias (2025). Enhancing educational gamification through AI in higher education. In *Proceedings of the 2024 16th International Conference on Education Technology and Computers*, ICETC '24, New York, NY, USA, pp. 213–218. Association for Computing Machinery.

Videla, R., S. Penny, and W. Ross (2025, August). 'if you can't dance your program, you can't write it': Challenges and implications for AI in education. *ACM Transactions on Computing Education*. Just Accepted.

Wei, M. and Z. Zhou (2023). Ai ethics issues in real world: Evidence from AI incident database. In *56th Annual Hawaii International Conference on System Sciences*, pp. 4923–4932.

Yan, L., R. Martinez-Maldonado, and D. Gasevic (2024). Generative artificial intelligence in learning analytics: Contextualising opportunities and challenges through the learning analytics cycle. In *Proceedings of the 14th learning analytics and knowledge conference*, pp. 101–111.

Yin, Y., S. Karumbaiah, and S. Acquaye (2025). Responsible AI in education: Understanding teachers' priorities and contextual challenges. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '25, New York, NY, USA, pp. 2705–2727. Association for Computing Machinery.

A Figures

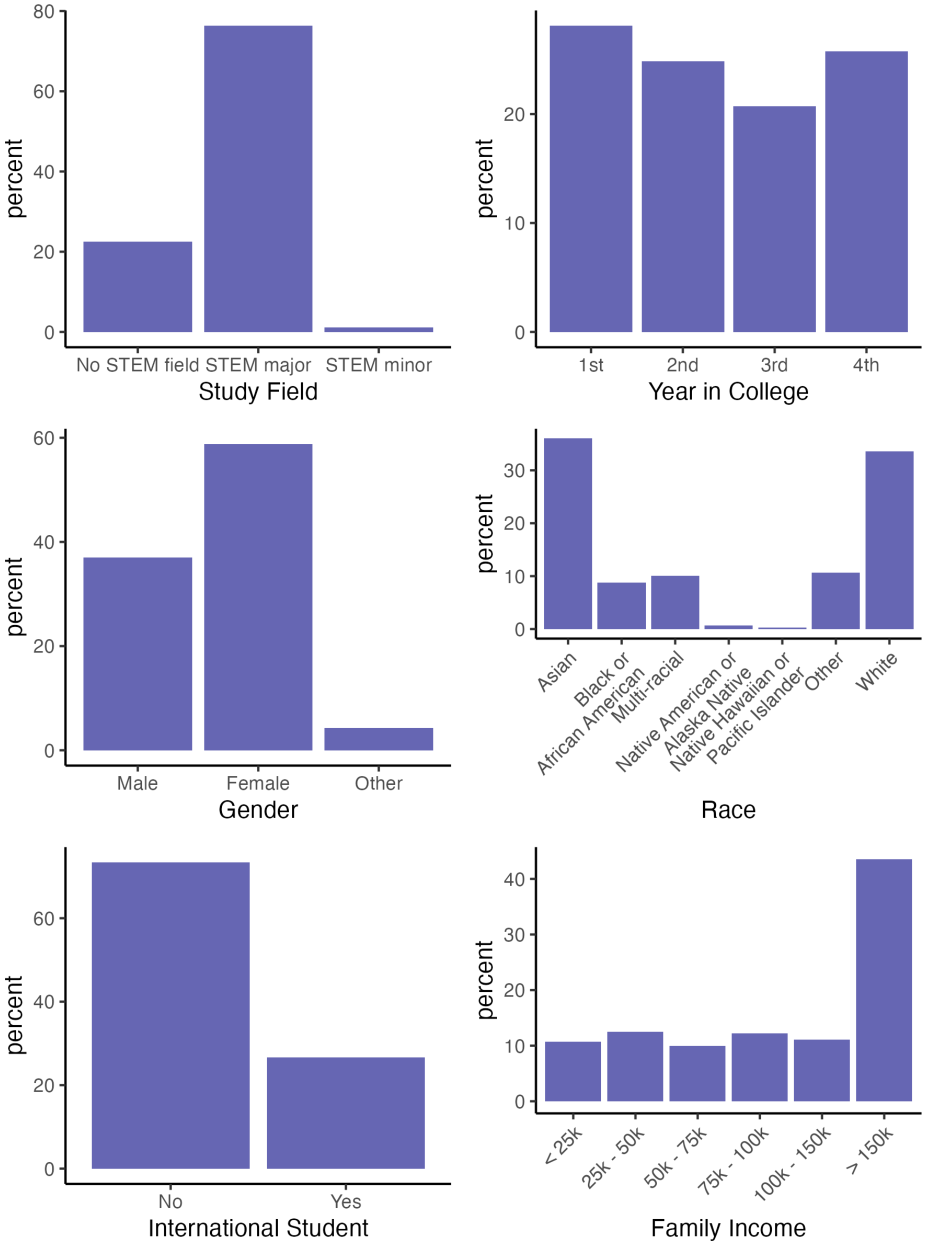
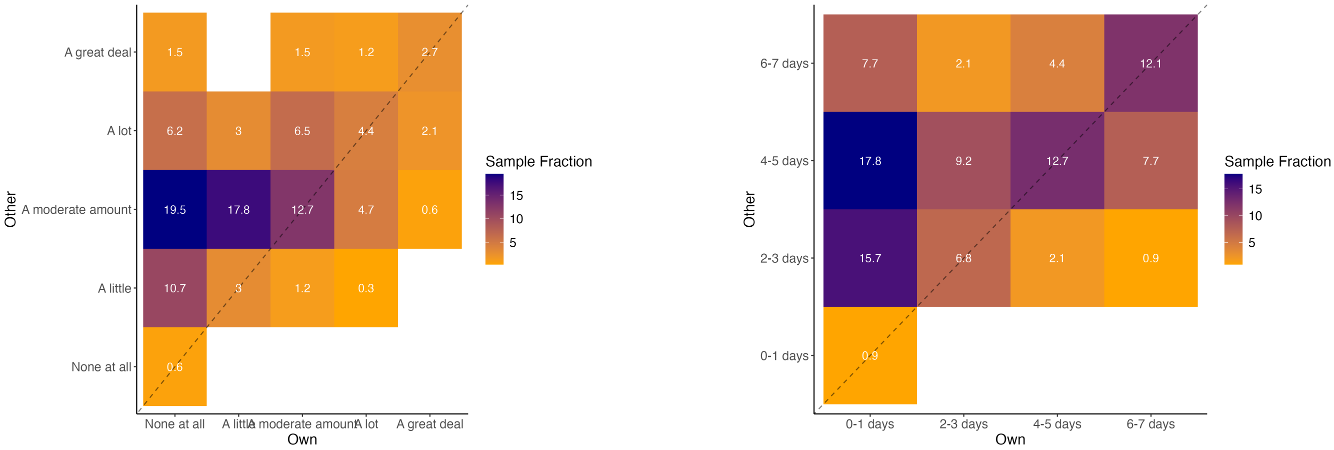


Figure 9: Reported demographics in the representative undergraduate survey.

A.1 Survey Results - Heatmaps

We emphasize the single-choice measure for Other reliance to be consistent with how Own use was elicited. Figures 10a and 10b reflects the reporting gap on a more granular level. The numbers in each cell reflect the sample fractions of each Own×Other combination. No subject report using AI for themselves but no AI use for others, and the “sample distributions" are generally above the 45 degree line, the “equal" reporting benchmark. This suggest that the reporting gap is not a result of aggregation.



(a) Reliance

(b) Frequency

Figure 10: Proportions of participants in each own-use × other-use combination.

B Bayesian Ordinal Analysis

We also performed a Bayesion ordianl analysis using the ordered levels of AI usage. We separately ran the regressions using `AI Use`$\sim$ `Own`$\times$`Gender`$\times$`STEM+(1|participant)` for both measures of AI usage: reliance and frequency, and fit the model predictions to sample data. The model predicts a probability distribution over levels of AI usage for each entry (i.e., each participant× Own/Other combination) within each simulation draw.

B.1 Reliance

B.1.1 Marginal Contrast

We evaluate the mean probability for each level of AI usage across all participants within each simulation draw to calculate the marginal contrasts. For example, focusing on AI usage with 5 levels (same as reliance) and one draw, suppose there are 5 participants, and the probabilities predicted for the lowest level are 0.8, 0.8, 0.6, 1, and 0.8 for Own, and 0.2, 0.2, 0.4, 0, and 0.2 for Other, we calculate the marginal contrast for this draw as

$$\overline{\text{Own}} - \overline{\text{Other}} = \frac{0.8 + 0.8 + 0.6 + 1.0 + 0.8}{5} - \frac{0.2 + 0.2 + 0.4 + 0.0 + 0.2}{5} = 0.6$$

Specifically, we consider the probability of any AI usage for marginal contrasts. Hence, the levels we focus on are any level other than zero usage: levels except “None at all" for reliance and “0-1 days" for frequency. We further contrasts along the gender and the STEM dimentions similar to a Difference-in-Difference set-up, where we further calculate the difference of difference between Own and Other between females and males (Male $\times$ ($\overline{\text{Own}} - \overline{\text{Other}}$) $-$ Female $\times$ ($\overline{\text{Own}} - \overline{\text{Other}}$)), or STEM-majors and non-STEM-majors (STEM $\times$ ($\overline{\text{Own}} - \overline{\text{Other}}$) $-$ nonSTEM $\times$ ($\overline{\text{Own}} - \overline{\text{Other}}$)).

Figure 11 shows the baseline Own-Other marginal contrasts across draws, while Figures 12a and 12b capture the further contrasts by gender or STEM. Figure 11 reflects the reporting gap: the model almost entirely predict lower probabilities of AI usage for Own than Other. Figures 12a and 12b display both positive and negative differences across draws with 2.5% and 97.5% quantile intervals $[-0.06, 0.13]$ and $[0.02, 0.28]$ respectively. This suggests that STEM-majors may produce larger reporting gaps than non-STEM-majors after aggregation.

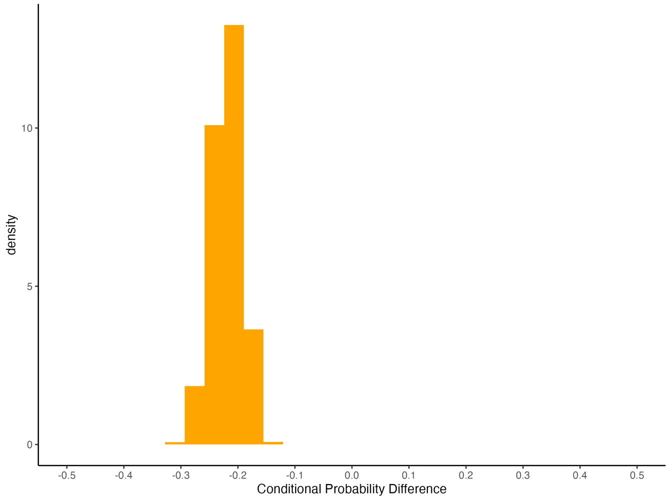
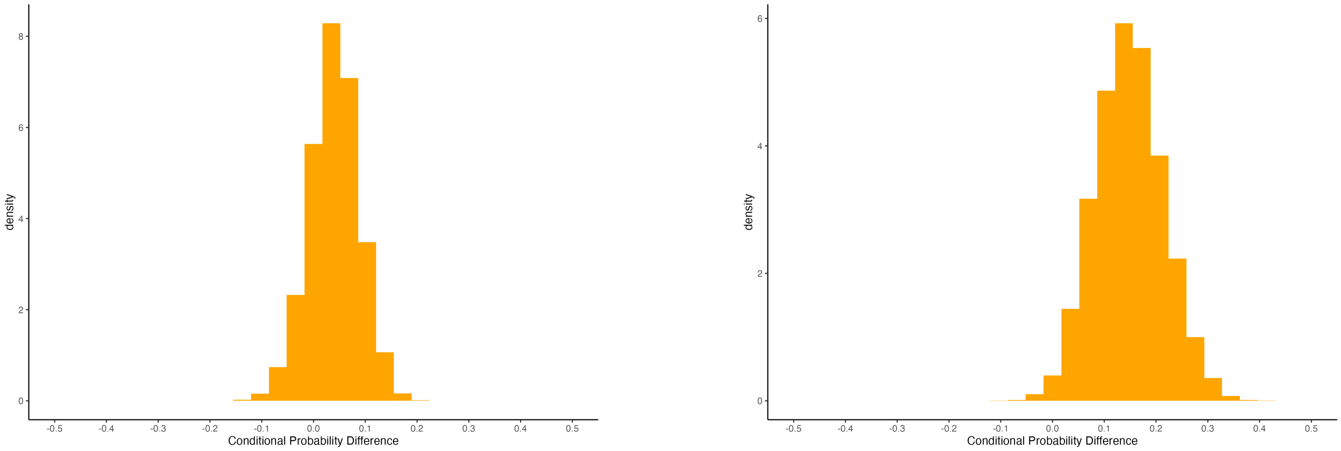


Figure 11: Reliance - Marginal contrast using own vs. other predicted probability of AI use.



(a) Gender - using marginal contrast in male participants vs. female participants

(b) STEM - using marginal contrast in STEM-major participants vs. non-STEM-major participants

Figure 12: Reliance - Further Marginal contrasts

B.1.2 Prediction

In this section we evaluate the model predictions directly by presenting the mean and 95% quantile intervals for each level of AI usage across participants and draws, separately for Own and Other. We also combine sample data with these predictions to reflect model fit. We also examine differences in predictions across gender and STEM-majors.

Figures 13, 14, and 15 consistently reflects the reporting gap in both sample data and model predictions. Similarly, we emphasize on the shaded regions of “None at all", and all areas outside the shaded region represent AI use. While model predictions also relfect a reporting gap as seen in the data, the figures show varying degrees of model misfit, potentially because participants only responded to Own and Other AI reliance once and a random effect estimation could not be properly supported.

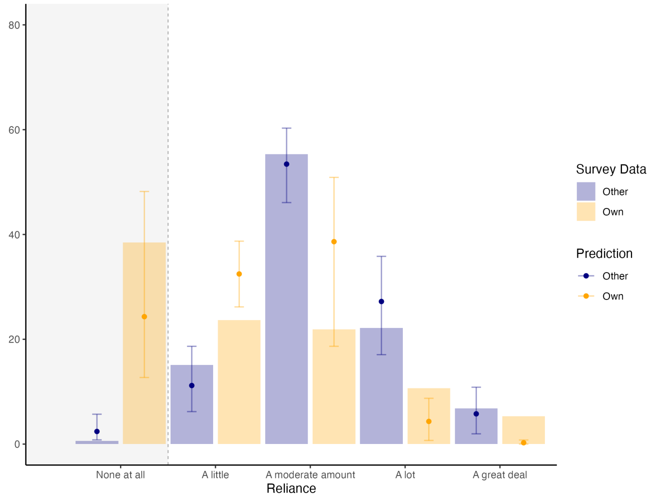
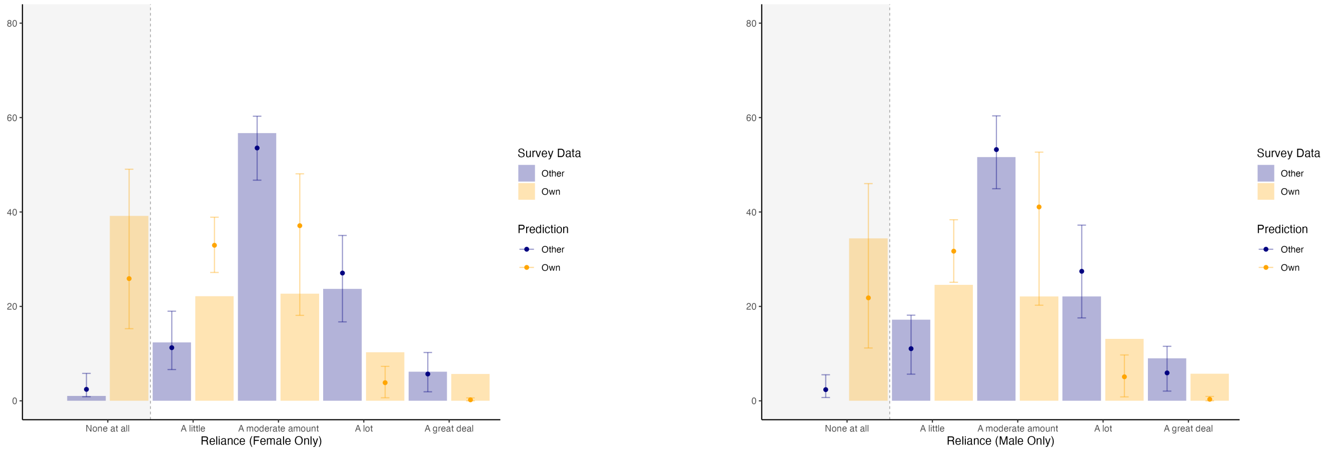


Figure 13: Reliance - Model predicted probability at each level of reliance



(a) Female

(b) Male

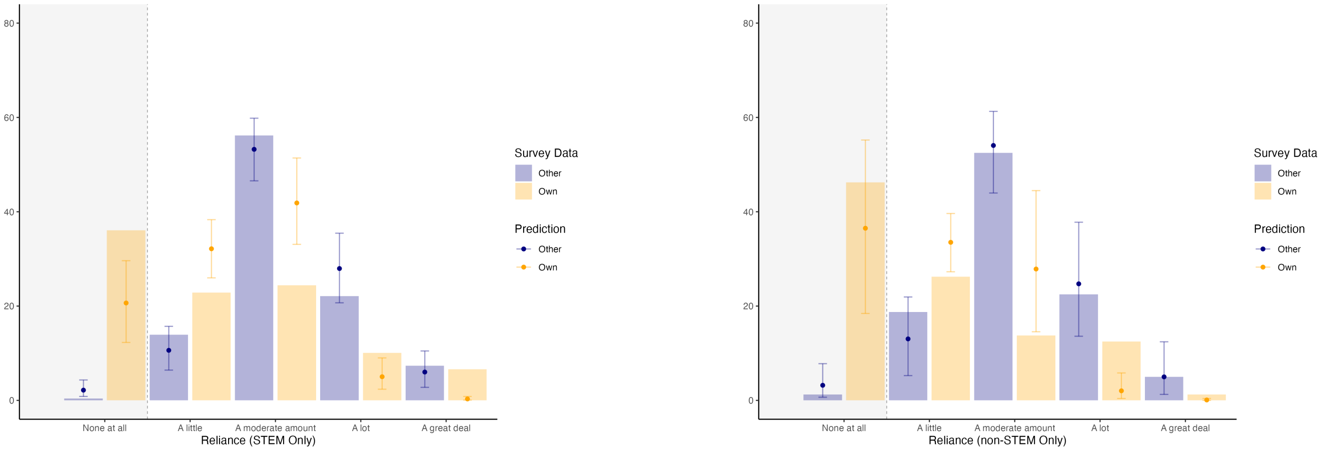
Figure 14: Reliance - Model predicted probability at each level of reliance by gender

B.2 Frequency

We perform the same analyses on the metric of AI frequency.

B.2.1 Marginal Contrasts

Figures 16, 17a, and 17b show a similar trend as reliance. In the base case, the model almost entirely predicts higher probability of AI usage for Other. The 2.5% and 97.5% quantile interval for the gender contrast is $[0.021, 0.20]$ and for the STEM contrast is $[0.001, 0.25]$. These suggests male respondents and STEM-majors may produce larger reporting gaps.



(a) STEM

(b) non-STEM

Figure 15: Reliance - Model predicted probability at each level of reliance by STEM-major

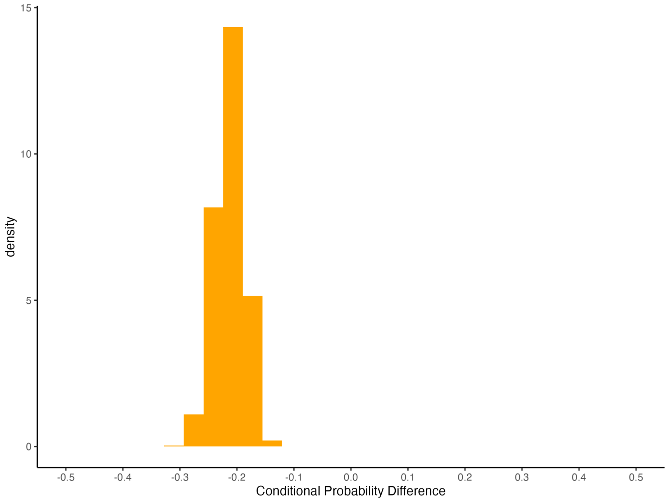
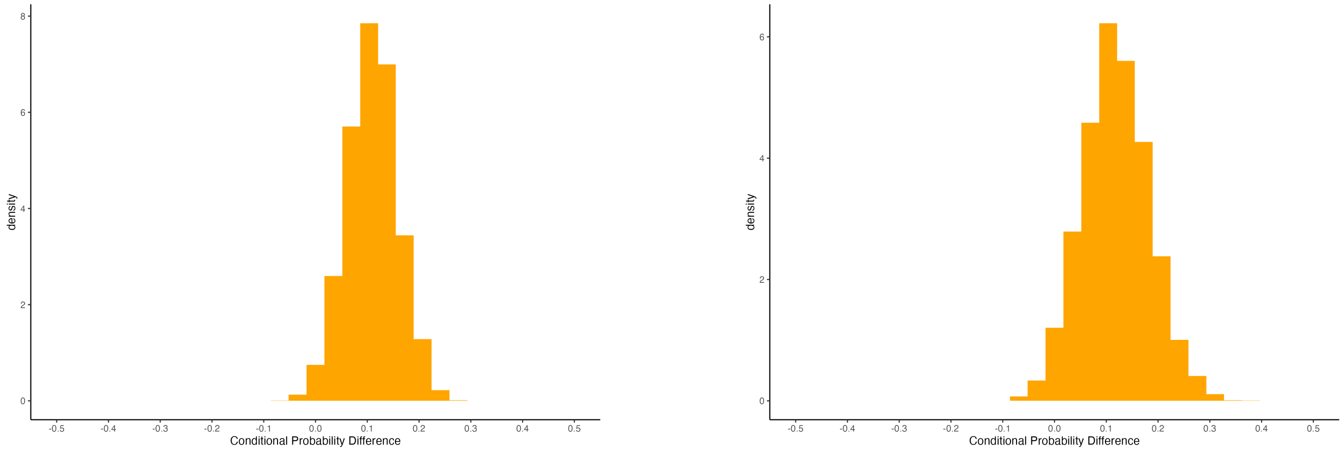


Figure 16: Frequency - Marginal contrast using own vs. other predicted probability of AI use.



(a) Gender - using marginal contrast in male participants vs. female participants

(b) STEM - using marginal contrast in STEM-major participants vs. non-STEM-major participants

Figure 17: Frequency - Further Contrasts

B.2.2 Prediction

Similar to reliance, we also observe reporting gaps in both sample data and model predictions in Figures 18, 19, and 20. Varying levels of model misfit still persist, potentially due to the same reason that each participant only responded to Own and Other AI frequency once.

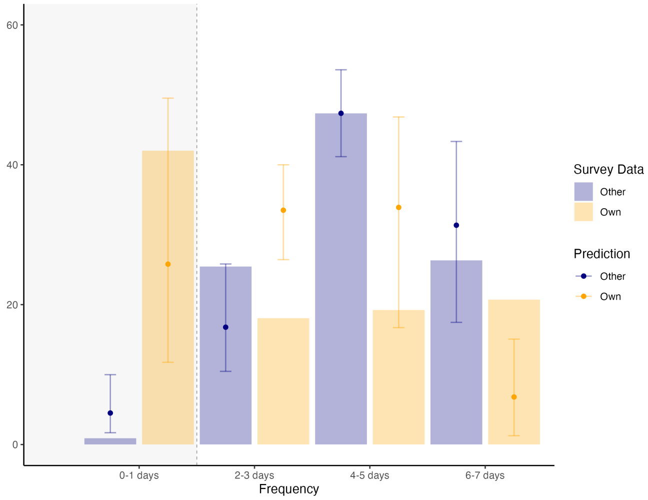
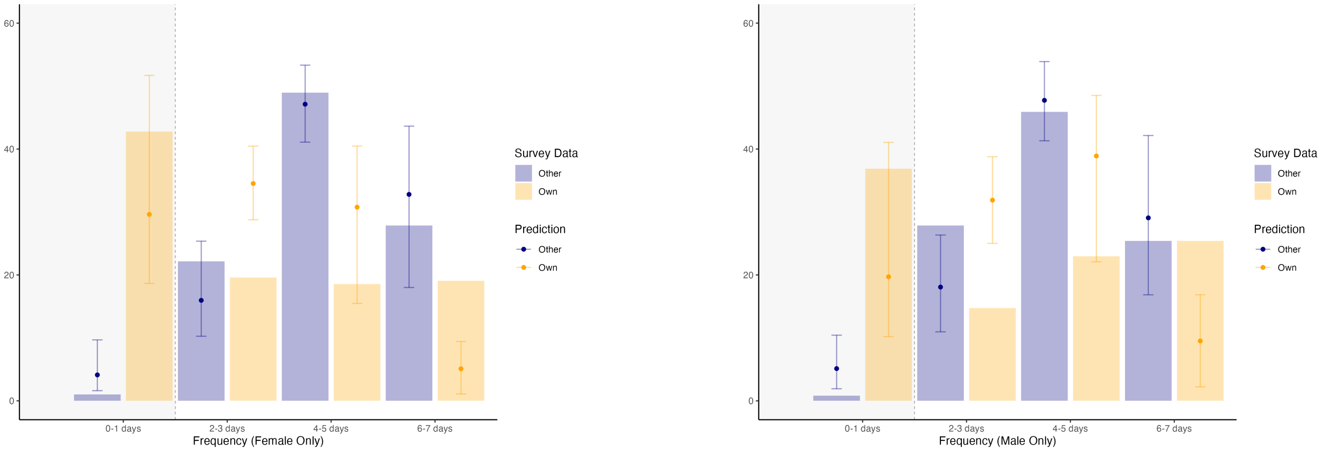


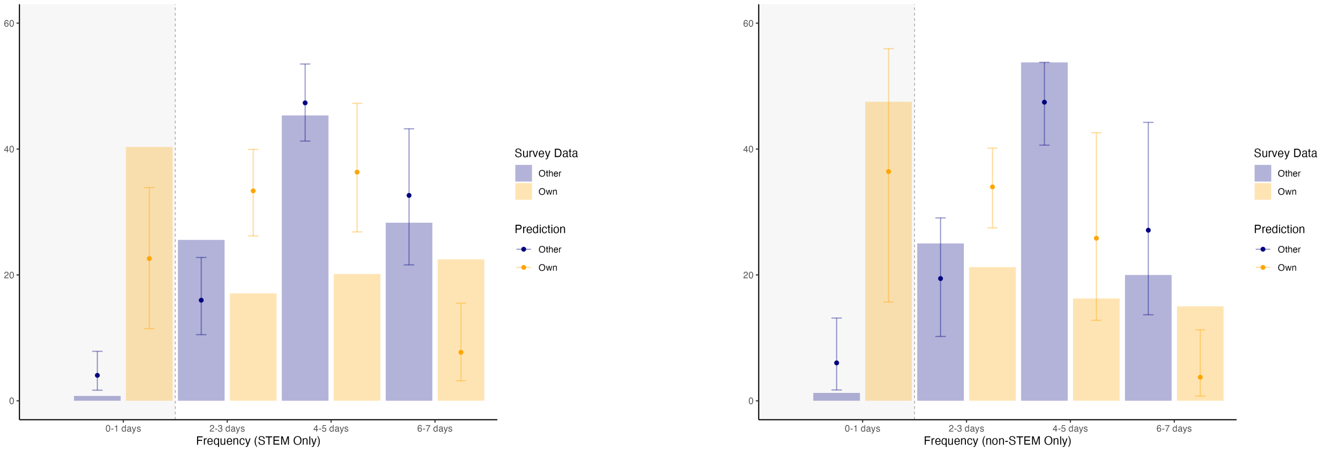
Figure 18: Frequency - Model predicted probability at each level of frequency



(a) Female

(b) Male

Figure 19: Frequency - Model predicted probability at each level of frequency by gender



(a) STEM

(b) non-STEM

Figure 20: Frequency - Model predicted probability at each level of frequency by STEM-major