Will Artificial Intelligence Get in the Way of Achieving Gender Equality?*

Daniel Carvajal† Catalina Franco‡ Siri Isaksson§

March, 2025

Abstract

We conduct two surveys with preregistered experiments to examine gender differences in generative AI adoption and potential labor market consequences. First, we document a substantial gender gap in AI adoption among students at a top business school in Norway, with female students, particularly top students, opting out of AI use. Second, a survey of managers shows that acquiring AI skills would significantly enhance job prospects for top female students currently avoiding AI. Finally, we provide causal evidence on policy tools to close the gender gap in AI adoption. Our findings show that while generative AI could widen existing gender gaps in the labor market, with the right policies, it can also be leveraged to help close them.

JEL Codes: I24; J16; J24; O33

Keywords: Generative AI, ChatGPT, gender, labor market, technology adoption

*We would like to thank Anna Dreber Almenberg, Katie Baldiga Coffman, Andrea Bocchino, Francesco Capozza, Finn Casey, Nisvan Erkal, Iver Finne, Anders Humlum, Dorothea Kübler, Danielle Li, Akhsay Moorthy, Erika Povea, María Recalde, Johanna Rickne, Mikko Silliman, Frode Skjeret, Emilie Steinmark, Stig Tenold, Heidi Thysen, Bertil Tungodden, Kata Urban, Tom Wilkening and Adam Zylbersztejn. Audiences at the ASSA 2024 meeting, Bergen-Berlin Behavioral Economics Workshop, FAIR, King's College, NHH, U. of Strathclyde, U. of Melbourne, WZB Berlin, Nordic AI-BEST workshop, Telenor, Paris School of Economics, LISER Luxembourg, KIT Karlsruhe, Aalto University, the Swedish Institute for Social Research (SOFI), MCV Workshop Madrid, GATE lab at Lyon University 2, NTU-GATE Workshop, ASFEE 2024, PET Lyon 2024, 2024 Asia Pacific ESA, EALE 2024, ESA European Meeting 2024 and EEA-ESEM Congress 2024 provided useful comments and feedback. This study was approved by the NNH IRB (NHH-IRB 42/22) and preregistered with pre-analysis plans in the AEA RCT registry (AEARCTR-0012452). We acknowledge support from the project "Management Practices and Gender gaps: Mechanisms behind the Gender Gap in Career Progression" (335448) financed by the Norwegian Research Council.

†Aalto University, Helsinki, Finland. Email: Daniel.Carvajal@aalto.fi

‡Center for Applied Research (SNF) and FAIR at NHH Norwegian School of Economics, Bergen, Norway. Email: Catalina.Franco@snf.no

§FAIR at NHH Norwegian School of Economics, Bergen, Norway. Email: Siri.Isaksson@nhh.no

1 Introduction

The advent of generative artificial intelligence (AI) is predicted to reshape the labor market. Recent surveys of employers in the US and globally find that over 90% expect their organizations to use AI by 2028 (Amazon Web Services, 2024) and 66% state they would not hire someone without AI skills (Microsoft & LinkedIn, 2024). Experimental research shows how access to AI can provide substantial productivity boosts (Capraro et al., 2024; Caplin et al., 2024), spanning domains such as professional writing (Noy and Zhang, 2023), customer support tasks (Brynjolfsson et al., 2025), and coding (Peng et al., 2023). Although exact economic impacts are hard to predict and depend on the policies adopted (Brynjolfsson and Unger, 2023), generative AI proficiency is likely to shape labor market paths and success in the near future. Moreover, those who actively participate in this technological revolution may be better positioned to reap its benefits. This raises a key question: will differences in AI use bridge or expand existing labor market inequalities?

We focus on gender differences in generative AI adoption. Previous technological breakthroughs, such as the introduction of the internet, have shown that gender plays a significant role in technology usage patterns, a phenomenon known as the digital gender divide (Bimber, 2000; OECD, 2018).[1] If women are less likely than men to adopt productivity-enhancing technologies such as generative AI, they may miss out on the promise of these technologies and fall behind in a labor market increasingly demanding and rewarding AI skills. Our findings show that female students adopt these technologies less frequently than their male counterparts, particularly among those with higher academic skills. Our study was the first to document gender gaps in generative AI adoption and several subsequent studies (summarized in Otis et al., 2024) have validated our findings.[2] Our study remains unique in identifying where these gaps emerge in the skill distribution, how they can be closed, and

[1] In general, a large literature documents that women are less inclined to and have lower participation rates in technology-related fields (Buser et al., 2014, 2017; Cimpian et al., 2020). Women also report higher technological stress than men (Kotek and Vranjes, 2022).

[2] Otis et al. (2024) conduct a meta-analysis of 16 studies across 26 countries and shows that gender gaps in AI adoption are nearly universal. Humlum and Vestergaard (2024) were the second study after ours to document a similarly-sized gap in a large representative sample of Danish workers across 11 occupations. Later work provides evidence from a nationally representative U.S. survey showing that men are more likely than women to use ChatGPT at work and at home (Bick et al., 2024), and from internet traffic data reporting that only 33% of global ChatGPT users are female (Liu and Wang, 2024).

their implications for labor market outcomes.

We target both the supply and demand sides of the labor market by focusing on college students who will be facing a rapidly evolving labor market due to generative AI along with their potential future employers. On the supply side, between November 2023 and April 2024, we surveyed 595 students at a top business school in Norway about their generative AI use, as well as their preferences, perceptions and exposure to generative AI. In addition, in a between-subject vignette experiment we assessed the impact of two types of policies —either explicitly allowing or banning the use of generative AI— on the students' intended use of AI for schoolwork. The experiment contributes to the ongoing global discussion among educators and institutions regarding their stance on the use of AI tools in the classroom, whether to ban or allow them.[3] We then assess whether the gender gap in adoption could potentially affect labor market outcomes. To do this, we turn to the demand side by surveying managers to examine gender gaps in job candidate ratings based on their AI expertise: can female students enhance their job prospects by acquiring generative AI skills? In June 2024, we conducted a survey experiment on 1,143 managers in Norway working in occupations that typically employ graduates from this business school and ask them to rate profiles of hypothetical candidates with and without generative AI skills. The design of the manager survey experiment allows us to isolate the effect of acquiring AI skills for a male vs. a female job candidate and of using AI in the workplace.

Our results are obtained using survey experiments and hypothetical scenarios.[4] The reason for this is the inherent challenges in detecting generative AI use. While direct observation and revealed-preference methods are typically preferred in economics, field experiments or observational data on AI adoption are impractical because of several reasons, making our chosen method a first best. First, AI use is not observable. Even if researchers provided students with computers to track use, they could still use personal devices to bypass these restrictions. Second, reliable AI detection is not feasible as AI detection technologies are prone to false positives and easily circumvented (Sadasivan et al., 2023). This lack of reli-

[3]Example institutions where bans have been enacted include Sciences Po in Paris (Business Insider, 2023) In contrast, Harvard Business School pays for each student to have a plus account (Magazine, 2024).
[4]We include a preregistration and pre-analysis plan (PAP) for the experiments in our study.

able detection means that field experiments would need to fully control access to devices over extended periods of time, which is practically unworkable and ethically questionable. We believe our chosen method allows us to assess adoption patterns in a natural setting, providing insights into early adopters and their motivations.

We provide five key insights. First, we document a substantial and significant gender gap in AI adoption: male students are 25% more likely to report a high use of ChatGPT or similar AI tools. Defining high use as engaging with AI tools either occasionally or all the time as opposed to never or only a few times, 60.7% of female students and 75.7% of male students fall into this category, with an overall average of 68.9%. When we ask students a more objective revealed-preference question of whether they have a paid account, or the more limited free account, the gender gap widens. Male students are more than twice as likely as female students (23.3% vs. 10.7%) to have a paid subscription.

Importantly, the gender gap in AI adoption is not uniform across the academic ability spectrum. Our second key finding shows that it is largely driven by the top women, those with the highest admission grades. While female students with lower admission grades are just as likely as male students to use AI frequently (over 75% report high use), top women are opting out and using it only about half as much. Put differently, male students use AI frequently regardless of their academic skill whereas top women opt out from using it. We know that this gender gap is not explained by gender gaps in ability of using AI since top women are just as good at a prompting task as men in our sample. In the top quintile of admission grades, female students have a success rate of 46% vs. 39% for top male students when asked to prompt.

The pronounced gender gap in AI adoption at the top is in line with previous research showing that gender differences are particularly wide at the top: top women often fail to recognize that they are better than average and act accordingly. This is noteworthy since the labor market gender gaps are also most evident at the top (e.g., Bertrand et al., 2019). Large gaps have been documented among top women compared to top men in several domains: high-achieving women are less likely to compete (Niederle and Vesterlund, 2007), less likely to speak up (Coffman, 2014), and less likely to claim credit for their contributions in successful group-work (Isaksson, 2019; Kinnl et al., 2023). Strikingly, throughout our analysis, the strongest results stem from the top women in our sample.

Our third key insight is that the gender gap in AI adoption is primarily driven by attitudes and beliefs about the technology. To better understand the reasons behind the gap, we examined three preregistered sets of factors influencing AI adoption: preference-based, belief-based, and experience/exposure-based factors. For example, in terms of belief-based factors, male students are more likely than female students to disagree with statements that using AI for learning (88% vs. 77%) and for course assignments (64 vs. 50%) constitutes cheating. In terms of preference-based factors, male students are more likely to report finding AI use enjoyable (81%) relative to female students (71%). These differences in perceptions on cheating and enjoyment highlight a broader pattern: the gender gap in AI adoption is closely linked to differences in preferences, perceptions, and exposure to the technology. When we control for the full set of these measures, the gender gap in adoption is fully explained and becomes insignificant, driven by preference-based and, in particular, belief-based factors.

Our descriptive analysis suggests that the large gender gap among top students is primarily driven by lower enjoyment and intrinsic motivation for using AI, along with ethical concerns. One interpretation of this finding is that top-performing women may be imposing self-restrictions on AI use. While these differences in perceptions could both contribute to and result from the observed gap, we next present the results from an experiment that addresses this potential endogeneity by virtue of randomization. Specifically, we explore how explicit policies about AI use can make female students overcome these self-imposed bans, demonstrating that the gap can be reduced with encouragement and permission to use AI in the classroom.

Our fourth key insight is that the main mechanism underlying gender gaps in AI adoption is the lack of clear policies on AI use. We show that the gender gap in intended use completely closes when generative AI tools are allowed in class. But also that the flip-side is true: if AI is banned, a substantial gender gap in intended use emerges. In our vignette experiment, male students intend to use AI tools regardless of the policy. In contrast, female students adjust their behavior based on the policy. Specifically, when it is allowed, over 80% of both men and women intend to use it. However, forbidding AI opens a large and statistically significant gap in intended use. While male students respond to the ban with a decrease of 20.7 pp, from 87.3% intending to use when allowed to 66.7% when forbidden, the response of female students is much larger at 37.2 pp, from 82.8% when allowed to 45.6% when forbidden, on average. This shows how seemingly innocuous university policies on AI use could have large unintended gendered consequences. However, clear and explicit policies encouraging generative AI use can close the gap. These insights are crucial as universities and workplaces are currently formulating guidelines and policies around AI use. As we show next, choosing these policies can have real labor market consequences for the top women currently opting out.

Our fifth key insight is that generative AI skills are valued in the labor market, in particular for top female job candidates. We find that female candidates with top grades who signal generative AI skills are evaluated 7.6% higher for an entry-level job than their female counterparts without AI skills, while male candidates do not receive a similar premium. An exploratory analysis suggests that the signal of AI skills is more informative and beneficial for women than for men which is driven by managers who correctly anticipate that women have lower AI adoption rates than men. This part of the study provides critical insights into how generative AI skills are rewarded in the labor market.[5] Our findings imply that top women, in particular, could significantly enhance their job prospects by acquiring generative AI skills.

Taken together, the supply and demand sides of our study show a clear picture: top women are opting out of AI, despite being the very candidates who stand to gain the most from adopting it.[6] Strikingly, this gap is driven by differences in attitudes and beliefs rather than ability to prompt. We also demonstrate that with the right policies, top women are willing and able to use AI tools. While the gender disparities in AI adoption are concerning, our findings also convey optimism by showing how adopting and developing AI skills can

[5]In a correspondence study, Drydakis (2024) reports similar callback rates for men and women showing training in business AI skills in their CV. The AI skills mentioned in that study encompass programming languages, machine learning, reinforcement learning, and natural language processing. These skills are fundamentally distinct from the generative AI focus discussed in our paper.

[6]In a follow-up study conducted in the lab, Franco et al. (2025) show no overall negative effects of AI use on learning. If anything, top women's learning is more likely to benefit from using AI tools.

help level the playing field for men and women. Thus, our study highlights the potential of generative AI to advance career opportunities for both genders, rather than impede progress towards gender equality.

The remainder of the paper is structured as follows. In Section 2, we provide an overview of the two survey instruments used in the study, along with their target sample: (i) the university student survey and (ii) the manager survey. In Section 3, we report our findings on the gender gap in generative AI adoption. Section 4 analyses different sets of potential drivers of the gender gap, including descriptive evidence for our set of pre-registered factors, as well as the policy experiment on explicit rules in generative AI use. In Section 5, we outline the experiment in the manager survey and present our results on the value of generative AI skills in hiring decisions. Finally, Section 6 concludes and proposes future directions.

2 Setting and Data Overview

Our design is guided by the two objectives of the paper: to study (i) whether there are gender differences in the use of generative AI from the supply side of the labor market (current students who will be looking for jobs within the next 2-4 years), (ii) what may be behind the gap and promising avenues for closing it, and (iii) whether AI skills are valued by the demand side (their potential employers). We use two complementary instruments: a survey of university students and a survey of managers. In this section, we provide a general overview of the survey instruments, recruitment, and sample. Both survey instruments, as well as a series of hypotheses regarding our main outcomes, were preregistered in the AEA RCT Registry (AEARCTR-0012452) with pre-analysis plans (PAP) for the experiments. More details on the preregistrations, as well as deviations, are discussed in detail in Appendix E.

2.1 Student survey

The first study focuses on current students who will be facing a labor market that is rapidly changing due to the expansion of AI technology. The survey instrument was administered to 595 bachelor's and master's students at NHH Norwegian School of Economics. NHH offers a tuition-free, five-year program consisting of three years of a bachelor's program in economics and business administration followed by two years of a master's program in either economics and business administration or international management. The bachelor's program at NHH is the most popular program in Norway, listed as the first choice by most applicants to higher education. Regarding NHH's stance on the use of generative AI by students, there was no specific policy at the time of our first data collection in November 2023. The school released a policy in December 2023 aimed at increasing transparency but did not offer specific guidelines for instructors, leaving it up to them to decide how to incorporate AI tools in their courses.

The survey collected self-reported measures of the use of generative AI, perceptions, preferences, and exposure to the technology, as well as a measure of prompting skills. In addition, the survey included a policy experiment aimed at exploring the impact of different policies regarding the use of generative AI on the gender gap. Questions about background characteristics, such as demographic and academic background were also collected. We measured risk and time preferences through survey questions following Falk et al. (2018). The motivation to add risk and time preferences is that there is some evidence of gender differences in these measures (Croson and Gneezy, 2009; Charness and Gneezy, 2012; Bettinger and Slonim, 2007; Castillo et al., 2011) and they could drive or correlate with generative AI adoption. Students were asked about their university admission grade (only requirement for admissions in Norway), with 328 students providing valid responses out of the 595 respondents (55% of the sample).

Recruitment and sample characteristics. Students were recruited during lecture hours in November 2023 and early 2024.[7] We approached students taking three of the mandatory courses of the bachelor's program (one for each year of the bachelor's program), as well as one of the core courses in the master's program. Importantly, all students are following the same study program and most of the subjects are mandatory as opposed to elective. Hence, any differences we find are not driven by self-selection into fields of study or specific subjects

[7]In November 2023, when most of our student sample was gathered, we focused solely on ChatGPT, as other platforms were either unavailable or not widely used at that time. By 2024, we expanded the questions to include ChatGPT along with similar platforms, providing examples of alternative options.

that lend themselves more or less to the use of AI. The anonymous survey was implemented and supervised by the research team in the classroom using a QR code.

Almost 55% of our sample is male, which is close to the historical male student representation at NHH of about 60% (Hirshman and Willén, 2022). In addition, over 90% of the sample is in the bachelor's program. Male students in the sample are statistically more willing to take risks and forgo something beneficial today to benefit more in the future than female students. While only 55% of the sample provided a valid answer for their admission grade, there are no gender differences in the likelihood of reporting the grade or in the grade itself. On average, the admission grade is 5.6 (median equal to 5.7) for both men and women, and the distributions are quite similar (see Figure A1).[8] The full questionnaire of the student survey is in Appendix F.1. Students took on average 8 minutes to respond the survey (7.9 minutes for women and 8.2 minutes for men, not statistically different).

We also note that our sample corresponds to the upper tail of the generative AI-use distribution among higher education students in Norway. Across the Studiebarometeret survey of 997 institution-programs in 2023 (Ministry of Education, 2024), the bachelor's and master's programs at NHH are at the 95th percentile in generative AI use, with an average score of 2.17 on a scale from 0 (does not use) to 3 (uses frequently). This highlights the fact that students in this school are highly exposed to the technology.

Finally, we point out a few strengths of our sample. First, the size of the typical cohort is 500, so considering that most of our sample is from the bachelor's program (around 1,500 students in total), we reach almost 50% response rates. Second, our sample is quite homogeneous given that the school offers a single major and admissions are very competitive, guaranteeing that those who get in have fairly similar backgrounds. Third, as the students were recruited in class from the mandatory courses, we believe our results are not simply driven by gender differences in the choice of subjects that are more or less amenable to the use of generative AI.[9]

[8]Higher education in Norway requires admissions to be based on an admission score determined through standardized testing and performance in high school. The maximum grade is 6, but it is possible to retake subjects so that there could be values above 6. The admission grade provides us with a comprehensive measure of academic performance, which we exploit for heterogeneity analysis in our results.

[9]In the bachelor's program, students take 4 subjects every semester, for a total of 24 subjects, of which only 6 are

2.2 Manager survey

The second study aims to understand the implications of gender gaps in AI adoption by students through assessing whether generative AI skills are valued by employers making hiring decisions. To achieve this, we conducted a survey on a sample of 1,143 managers in Norway. To measure the value of generative AI skills in hiring, we implemented a conjoint-type experiment, where managers evaluate and score hypothetical job candidates applying for an entry-level job at their company. The survey also included questions regarding managers' own use of generative AI, attitudes and exposure towards the technology at their company, and their perception of gender gaps in AI use by students. Finally, we collected information on background characteristics such as gender, age, level of education and tenure at the company.

Recruitment and sample characteristics. Managers were recruited through the survey provider Norstat between May 30th and June 18th, 2024. Respondents were screened based on two characteristics. First, whether the respondent has been involved in managerial tasks involving hiring or promotion in their current job. Second, whether the respondent works in one of four pre-selected occupations. The survey was sent to 2,030 respondents in the Norstat panel.

We aimed to obtain managers from companies in the sectors that NHH graduates typically find jobs. An NHH report indicates that almost 90% of their graduates start their first job after graduation in one of the following lines of business: consulting, auditing, banking/insurance/finance, energy, IT/telecom and accounting (NHH, 2024). To most closely match the labor market where graduates from NHH will find jobs, we pre-selected managers working in the following occupations: administration/personnel, banking/accounting/finance, consulting, and management services. The choice of sector rather than companies where NHH graduates are hired reflects high costs of screening based on whether a specific company hires NHH graduates.

elective. There are no electives in the Autumn semester of the first year (from which a third of our sample is recruited), and one elective thereafter except in the last semester of the program in which students can choose two electives. Subjects in the master's programs involve 6 subjects and a master's thesis, where at least 3 of the 6 subjects must be selected from a list of mandatory subjects.

About 60% of the sample is male, and 52% of the managers worked in administration, 18% in banking/accounting/finance, 9% in consulting, and 21% in public service and management. Around 30% of the managers worked in companies that allow and encourage the use of generative AI at work. The full questionnaire of the manager survey is in Appendix F.2. Managers took a median of around 7 minutes to complete the survey.

2.3 Anonymity and Participant Incentives

Our aim is to elicit truthful responses in both survey instruments. Given the controversy and ongoing debate surrounding the ethics of generative AI use, we opted for ensuring full anonymity of responses. This approach minimizes the risk of misrepresentation of generative AI use due to experimenter demand effects or social desirability bias.

Incentivizing the reporting of the measures collected and the prompting task, or linking survey data to administrative data, would have required collecting personally identifying information. To preserve anonymity, we opted for unincentivized measures. We also chose to conduct the survey in the classroom to prevent students from seeking external help (from someone else or from generative AI applications) to get the correct prompt.

Validation exercises have found strong similarities in the use of hypothetical and unincentivized measures relative to incentivized elicitations and real-world behavior across different domains (Hainmueller et al., 2015; Brañas-Garza et al., 2021, 2023; Enke et al., 2022; Falk et al., 2023). At the same time, there has been an increase in the use of unincentivized measures in economics research (Ameriks et al., 2020; Bernheim et al., 2022; Stango and Zinman, 2023; Almås et al., 2023; Andre et al., 2022). Given the restrictions in our scenario and the concerns over potential effects of incentives on reporting actual capabilities, we opted for the use of unincentivized questions.

Ex-post, the role of incentives may have been minimal. The reason for this is that we cannot incentivize students to report truthfully on their AI adoption because we can never know their true usage of AI due to the nature of generative AI. The only measures we could have incentivized were performance in the prompting exercise, confidence in the quality of their prompt, second-order beliefs about peer usage, and risk and time preferences—none of which play key roles in the paper. For the prompting exercise, offering incentives would have required collecting personal information, which could have influenced how students responded to questions about usage or subscriptions even if these cannot be incentivized, as they would be aware that we knew their identities. Thus, setting up an anonymous and thus unincentivized study is the best way to go about answering adoption questions since students can report truthfully without the fear of potential repercussions in class.

3 Gender gap in generative AI use

3.1 Main outcomes

We investigate two main outcomes related to use. First, we focus on students' answers to the question "*How familiar are you with generative AI?*" In the analysis we use a binary variable equal to zero for *low use* if the student indicated "not heard about it," "heard about it but not using it myself" or "used it a few times," which indicates none or limited use, and equal to one for *high use* if the student indicated "use it occasionally" or "use it all the time," which indicates a more regular use. Second, we use an objective, revealed-preference measure of use, namely whether the student had a free or paid subscription to an AI chatbot such as ChatGPT. Participants also selected the types of tasks they "typically ask AI to help with."

We also measure skill proficiency in the use of generative AI, where we present an image of the "Ebbinghaus illusion" to the students, and asked them to write in a text box the query/prompt they would provide to ChatGPT to arrive at the correct official name of the visual phenomenon represented by the image.[10] We use three outcome measures based on the prompting exercise: time spent writing the prompt, the number of characters written, and the success rate of the prompt. The success rate was computed by entering a prompt 50 times in ChatGPT, and calculating the proportion of times it gives the correct answer.[11]

[10]The students wrote the prompt as a response to the survey question and not directly on ChatGPT. We developed this prompting task aiming for an objective and non-trivial task. Ex-post it was evident that the task belongs to the retrieving information category that students state they use generative AI most for. In this sense, it is a relevant task for the student population. Their prompting exercise was supposed to give the answer: Ebbinghaus-Titchener illusion, which is an optical illusion where context affects perceptions of size. Gender differences in visual recognition of images (Phillips et al., 2004) may have disproportionately affected the ability of women to write a successful prompt. We provide details in Appendix C.1.

[11]As large language models (LLMs) output is the result of probabilities and prediction, a different output is generated after every prompt. To address the noise of the process, we input the prompt a sufficiently large

3.2 Econometric specification

We estimate the gender gap in AI use using an indicator for whether the participant is a male student:

$$y_i = \alpha_0 + \alpha_1 \text{Male}_i + X_i \gamma + \varepsilon_i \tag{1}$$

The coefficient α_1 provides an estimate of the gender gap in the outcomes of interest. We present raw gaps in our main results tables, and complement the analysis by controlling for a series of controls X_i including background characteristics and pre-specified factors that may influence adoption such as AI-related preferences, perceptions, and experience in additional tables. We pre-specified the hypothesis that men have higher use of generative AI than women.

3.3 Main results

Generative AI adoption. Figure 1 shows the proportion of responses in each level of AI use split by gender, with the height of the bars adding up to 100% within gender. Female students are much more likely to be represented in low use categories. 9.6% of female while 2.2% of male students state that they have heard about generative AI but do not use it. 29.6% of female and 21.5% of male students have used it few times. In contrast, male students are overrepresented in the use all the time category with 44.3% relative to 30% of female students in this category. The proportions in the use occasionally category are similar with 30.7% of female and 31.4% of male students. A non-parametric Wilcoxon rank-sum test shows that the distribution of answers is different for women and men at the 1% significance level. Overall, the raw gender gap in adoption is estimated at 15 pp or 25% over a base of 60.7% of female students using AI occasionally or all the time (Column 1 in Table 1, Panel A).

In terms of having a free or paid account to a generative AI chatbot, about a third of female students declare having a free subscription, while less than 11% have a paid subscription (Columns 2 and 3 of Panel A, respectively). Male students are more than twice as likely to

amount of times (50).

have a paid subscription, which we interpret as evidence that they have a higher willingness to pay for a more comprehensive generative AI toolkit.[12]

Figure 2 lists the tasks for which students typically get AI help along with the fractions of female and male students who select each of the tasks.[13] The most popular task is "retrieving information" followed by "writing tasks." 65% of male students selected retrieving information as one of the tasks where they typically use AI, relative to 50% of female students, while 55% of men selected writing tasks, relative to 46% of women. We also see gender differences in coding tasks, but not in solving math questions and other tasks, which include brainstorming.

Heterogeneity by academic skill. Our unique data on admission grades, allows us to explore whether the gap is uniform across the academic skill spectrum. We plot the high use variable by quintile of admission grade, a measure of relative academic skill in Figures 3a and 3b.[14][15]

The fraction of men with high AI use (Figure 3b) is between 75% in the second highest quintile up to 87% in the middle quintile, so it is quite homogeneous across quintiles. In contrast, the fraction of women with high AI use is strongly and negatively correlated with admission grade quintile. In the bottom two quintiles, the fraction of women with high use is similar to the fraction of men (88%), while for the three top quintiles, the fraction of women with high use is below 55% (Figure 3a). A regression estimating the correlation between the raw admission grade and the high baseline use indicator yields a negative and significant coefficient for both men and women, but it is over six times larger for women (-0.316) than for men (-0.05). A test of the difference between the two correlations yields a p-value of 0.000.

The finding that women at the top of the skill distribution are less likely to use generative

[12]For chatbots with paid subscriptions, e.g., ChatGPT and Claude, the monthly price of a standard paid subscription is around US$20 as of February, 2025.

[13]The fractions are computed across all students, assigning a zero to those who do not use generative AI.

[14]Quintiles are calculated pooling men's and women's admission grades. The admission grade densities by gender are plotted in Figure A1.

[15]We did not pre-register hypotheses related to academic ability in our PAPs so these analyses are exploratory. Admission grades tend to be correlated with college GPA, which in turn increases hiring interest by employers (Kessler et al., 2019). They are also less likely to be affected by differences in AI use than college grades since they were obtained before the proliferation of generative AI.

AI is particularly interesting in light of the work by Brynjolfsson et al. (2025), who find that using AI help reduces the quality of work for the most experienced workers at a technical-support firm. One might conjecture that for the best students, using generative AI could reduce the quality of their output rather than improve it. If this is the case, top women would perform better in school and on the job because they do not use generative AI, and the gender gaps in the labor market could be reduced. However, recent studies indicate that access to generative AI in educational settings can improve learning outcomes for high-GPA women, and would not adversely affect overall learning (Franco et al., 2025). Hence, top female students opting out may not necessarily mean that they are learning more or better than their peers who use AI.

Prompting skill proficiency. Our measure of prompting skills allows us to assess whether top female students face greater challenges than men in interacting with AI chatbots, which could justify targeted training as a policy (Humlum and Vestergaard, 2024). Alternatively, if top female students who opt out AI are just as capable as their male peers, training may be less critical, and other barriers to female AI adoption may be more important. We find evidence for the latter with our prompting task: Our prompting task supports the latter: Top women achieve similar success rates with their prompts as top men. However, when considering all students, male students, on average, perform better at prompting than female students.

Table 1, Panel B, Column 1 quantifies the raw gap in prompt success rates. Overall, male students write prompts with a 34% higher success rate relative to female students. Column 2 shows that, on average, everyone spends about 129 seconds writing their prompt. Column 3 shows that male students write about 31.6 more characters in their prompt relative to a mean of 145 characters among female students. However, it is important to note that women with top admission grades perform just as well in the prompting exercise as their male counterparts. In Figures 4a and 4b, students at the top of the admission grade distribution have higher success rates regardless of gender (39-46% in the top two quintiles). As with the high use outcome, male students have more homogeneous success rates across quintiles than female students. Even though women in quintile 1 have the highest use, their success rate (17%) is the lowest overall and is only half of the success rate for men in the same quintile (34%), who have similar levels of use. These patterns suggest that women at the top of the academic skill distribution might be missing out on reaping the productivity benefits of generative AI, as they are equally skilled as their male counterparts.

While our crude measure of prompting is useful to have a feasible test of this type of skill within a short survey, it has some drawbacks. For example, the gender gap in proficiency may be partially attributed to differences in recognizing the Ebbinghaus illusion (Phillips et al., 2004), rather than differences in the quality of prompts written conditional on correctly recognizing the illusion (see Appendix C.1 for details). Nevertheless, the analysis provides insights into potential barriers and challenges in AI implementation. First, high-skill women may not necessarily benefit from training. Second, if generative AI were implemented at scale in schools and workplaces, lower-skill individuals might benefit more from training.

In sum, our results on adoption are in line with previous findings on gender differences in choices, particularly at the top of the skill distribution. Top women are less likely to compete then their male counterparts (Niederle and Vesterlund, 2007). Gender gaps in speaking up do not close with expertise (Coffman, 2014). The women who contribute the most to their team also underestimate themselves the most (Isaksson, 2019; Kinnl et al., 2023). Female students who do not achieve a top grade in a foundational course are less likely to choose a major related to that field (Rask and Tiefenthaler, 2008; Ost, 2010; Avilova and Goldin, 2018; Kugler et al., 2021; Ugalde, 2022). As in most of these previous studies, we find that gender gaps are driven by top women. We care about top women in this setting because they are the ones who have the highest prospects to become influential in the business sector, which traditionally lacks female representation even in a high-equality country such as Norway (Bertrand et al., 2019).

4 Drivers of the gender gap in generative AI use

4.1 Descriptive evidence

To understand what may be behind the gender gaps reported in the previous section, it is crucial to assess the role of various factors that may influence or correlate with adoption. We elicited students' attitudes regarding generative AI, which we pre-specified and classified into three categories: (i) preference-based factors, (ii) belief-based factors, and (iii) exposure/experience.

Preference-based factors. Preference-based factors measure intrinsic utility or disutility derived from AI use, as well as persistence in its use. Figure 5b shows the means by gender of our preference-based measures. We find that male students exhibit stronger preferences for generative AI by finding its use enjoyable (81%), relative to women (71%). Men also find it less difficult to use (lower disutility) than female students, though the difference is only significant at the 10% level. In terms of persistence when using generative AI, 71% of men attempt at least two times interacting with an AI chatbot when it does not provide a desired response, compared to 55% of women. The higher persistence could allow for increased engagement and skill development through prompting experience. To summarize, the evidence suggests that women have lower intrinsic motivation to use generative AI tools.

Belief-based factors. Belief-based factors capture perceptions regarding generative AI, including its perceived usefulness, whether its use is considered cheating, trust in AI-generated information, and confidence in one's ability to use AI. In Figure 5a we show whether gender differences emerge in our belief-based factors. First, students might not adopt the technology if they perceive its use as unethical or cheating. While most students disagree that ChatGPT use to solve assignments constitutes cheating, there is a notable gender gap: men are 12 pp more more likely to disagree with this statement than women. Interestingly, over 75% of men and women believe ChatGPT as a learning aid is not cheating, although the gender gap in percentage points is the same as in the question on use to solve assignments. Second, confidence in AI use also differs, with 81% of men expressing some level of confidence in whether their prompts would give the correct answer, compared to 60% of women. Male students are also 7 pp more overconfident in their prompts being correct than female students. Third, there are strong gender differences in perceptions of usefulness regarding the technology as follows (fraction of male vs. female students in parentheses): believing that using AI improves grades in a course (28% vs. 15%), increases accuracy or work quality (38%

vs. 26%), and improves the learning of course methods (56% vs. 43%). Finally, we measure trust in ChatGPT's accuracy in providing information where we find no gender differences. All in all, women may be more reluctant than men to use AI due to hesitations about its usefulness and ethical concerns.

Experience or exposure. Differences in AI adoption may stem from varying exposure to the technology, either through personal experience or peers. We ask students to estimate the percentage people that they believe use generative AI for three different groups: their friends, students in the course and professors at NHH. We find that they perceive that about 75% of their friends and peers, and 45% of professors at NHH, use ChatGPT, with no significant gender differences. However, men report higher direct experience with AI, which we measure by asking whether they have encountered inaccurate information provided by generative AI. In sum, all students seem well aware of their peers' AI use, but men appear to have more experience, suggesting greater engagement with the technology. This aligns with our findings that men have higher intrinsic motivation to use AI.

Revisiting the gender gap in adoption. We now aim to understand which factors correlate with or influence more the gender gap in AI adoption. To do this, we add different sets of controls in the regression of the main outcomes: academic variables and risk and time preference measures, preference-based factors, belief-based factors and experience/exposure. Table 2 presents the results after adding the controls to the raw estimates presented in Column 1. The group of controls added in each subsequent column is specified at the bottom. For illustration purposes, Figure A3 shows a selected set of factors used as controls, by quintiles of the admission grades distribution, with multiple instances showing gender differences in the top three quintiles.

In terms of the gender gaps in adoption using the high use and paid subscription outcomes, we see that the raw gaps go from 15 pp without controls to 0.8 pp with all controls in high use and from 12.6 pp to 3.5 pp in paid subscription (see Columns 1 and 6 of Panels A and B in Table 2). The raw gaps, initially both statistically and economically significant, become negligible in both respects after incorporating the full set of controls. Columns 2-5, which add the groups of controls individually in each column, suggest that the belief-based factors (cheating, overconfidence, trust and usefulness) and preference-based factors (intrinsic utility or disutility of use) are the ones that help to reduce the gender gap the most for both the high use and subscription outcomes. Zooming in the belief-based factors, we find that perceptions of usefulness and, notably, the belief that generative AI constitutes cheating account for nearly half of the reduction in the observed gender gap (see Table **??**).

While most of these controls may not be exogenous since they could both be consequences as well as causes of students' use and proficiency with generative AI, we see that the more exogenous controls already play an important role explaining the gap. The academic controls (year in college, admission grade and missing indicator for admission grade) along with the risk and time preferences measures in Column 2 already reduce the gaps in the two adoption outcomes by about half.[16] This is not surprising given that the gap is driven by women at the top and there are larger differences in adoption among first-year students (see Table A2).

Based on the descriptive evidence presented in this subsection, we posit that women—particularly high-achieving women—may be imposing self-restrictions on AI use, primarily due to ethical concerns. These restrictions may stem from conflicting messages in ongoing debates about whether AI should be allowed or banned in educational institutions, as well as from a lack of clear regulations. To address the endogeneity of the controls discussed above, the next subsection presents an experimental approach that serves three key purposes. First, the randomization is orthogonal to any beliefs or preferences students may already have, allowing us to provide evidence of causal mechanisms. Second, it allows us to examine how external restrictions, such as explicit regulations, influence AI adoption in contrast to self-imposed restrictions. Third, we can speak more directly to the mechanism driving the gender gap in adoption, offering insights into how this gap can be addressed and ultimately closed.

4.2 Experimental evidence: allowing/forbidding generative AI use

Motivated by the policy discussions around the world on whether to ban or allow generative AI use by students as part of formal education, we included in the student survey a policy

[16] If we define adoption as using AI "all the time," the gender gap already becomes statistically insignificant when adding academic controls and risk and time preferences (see Table A1).

experiment to assess student responses to such policies. Moreover, associated to the presence of self-imposed bans, it is important to understand whether actual restrictions affect use. Random assignment allows us to causally study the role of explicit restrictions on intended adoption.

4.2.1 Experiment design and main outcomes

We rely on a hypothetical vignette experiment as follows.[17] Students were presented with a hypothetical scenario describing a course they would be hypothetically enrolled in. The course description indicates how it is evaluated and we experimentally vary a statement of whether the professor explicitly allows or forbids the use of ChatGPT in the course as follows:

> *Imagine you are enrolled in a course on Environmental Policy and Economic Impact. This course explores the intersection of environmental regulations, economic incentives, and their effects on industry practices and sustainability. The professor explicitly allows/forbids the use of ChatGPT during coursework. It is an 8-week course with final evaluation given by a final in-person written exam.*

Subsequently, students were asked: "*Given this scenario, how likely are you to use ChatGPT throughout the course?*," where the choice consists of indicating intended use in a 5-point scale from "Very unlikely" to "Very likely." The wording throughout the course was intended to convey consistent use as opposed to using it in a single assignment.

Stratifying by gender, we randomly allocated students into one of two treatment conditions: (i) the professor explicitly *allows* the use of ChatGPT, and (ii) the professor explicitly *forbids* the use of ChatGPT. This enables us to causally study the effects of the allow/forbid policy on intended use.[18]

[17] Unfortunately, randomizing this type of policy in real institutions would not be feasible as we suspect few institutions would like to be part of such experiment and the number of institutions required to estimate the effects is likely large.

[18] A second layer of randomization was the type of evaluation of the course, which could be either an in-person exam or a home exam. Given that there are no strong differences in the findings by type of evaluation, we present the results for both types pooled. A discussion on the types of exam can be found in Appendix E.

4.2.2 Econometric specification

In this analysis, we estimate the gender gap in the policy reaction to allowing/forbidding ChatGPT in the hypothetical course presented in the vignette experiment:

$$y_i = \beta_0 + \beta_1 \text{Male}_i + \beta_2 \text{ChatGPT forbidden}_i + \beta_3 \text{Male}_i \times \text{ChatGPT forbidden}_i + X_i \gamma + \epsilon_i \quad (2)$$

The outcome y_i is equal to 1 for students who state that they are likely or very likely to use ChatGPT during the course. The coefficient β_1 provides the estimated gender gap when ChatGPT is allowed, β_2 represents the policy response (from allowed to forbidden) among women, and β_3 measures the differential change in the policy response for men relative to women. Similarly as in specification 1, we add different types of controls X_i that help us understand the influence of the preregistered factors on our results.

4.2.3 Main results

Figure 6 plots the raw gender gaps in intended use (likely or very likely to use) when ChatGPT is allowed or forbidden. When it is allowed, over 80% of both men and women intend to use it. However, forbidding ChatGPT opens a large and statistically significant gap in intended use. While male students respond to the ban with a decrease of 20.7 pp, from 87.3% intending to use when allowed to 66.7% when forbidden, the response of female students is much larger at 37.2 pp, from 82.8% when allowed to 45.6% when forbidden (see also Table 3, column 1). The point estimate for the gender gap in intended use following specification 2 is in Table 3, Column 1. When ChatGPT is explicitly allowed, the gap is 4.5 pp and not statistically significant. A gender gap in intended use equal to 16.6 pp opens up as a result of the forbidding policy (see interaction coefficient). Overall, female students react more strongly to policies banning ChatGPT use.

One of our most interesting results is the reaction to policies across the academic skill distribution in Figure 6. When the use of ChatGPT is explicitly allowed, men and women across the distribution state a similar level of intended use, and gaps do not emerge at any level of academic skill. Resembling the previous finding on current adoption, Figure 7a shows that the top female students would be the ones reacting more strongly to the forbidding policy. In Figure 7b we see that male students respond to the forbidding policy quite homogeneously across all quintiles and to a much lesser extent than female students.

We discuss a few points regarding our results on policy responses. First, the heterogeneity in responses based on academic skill among female students weakens the argument that our results are influenced by social desirability bias. One would have to make complicated assumptions on how social desirability bias interacts with relative academic skill and gender to explain the results. Second, inattention in vignette experiments is often a pervasive problem (Mas and Pallais, 2017) but, again, for it to generate our results, one must make assumptions on how inattention differs by gender and level of academic skill. We also note that in the vignette experiment female and male students spent 32 and 31 seconds, respectively, a difference not statistically significant. This suggests that inattention does not seem to be differential by gender.[19] Third, the gender gap in responses to policy remains the same even after adding the set of controls including background characteristics, and preferences, perceptions and exposure/experience regarding generative AI (Columns 2-6 in Table 3). Our interpretation of this result is that inclinations towards rule-following, obedience to authority, and trust in the professor's recommendations may play crucial roles in shaping the divergence in intended use, consistent with previous work showcasing gender differences in cheating (Kajackaite and Gneezy, 2017), and rule-following behavior (Kimbrough et al., 2024). These differences may be particularly relevant for top female students, who might place greater importance on how they are perceived by their professors. Finally, in Appendix C.3 we discuss the introduction of a transparency policy allowing the use of generative AI at NHH, where students surveyed before the policy show a strong gender gap while after the policy no gender gap emerges. This empirical fact aligns with our findings in the vignette experiment.

Our findings from the policy experiment align with the descriptive analysis in Section 4. As shown in Section 3.3, women exhibit lower adoption rates compared to men and are more likely to perceive generative AI use as cheating. We interpret these patterns in the context of NHH's AI policy, which, until December 2023, did not exist. In the absence of

[19] Hainmueller et al. (2015) find no gender differences in stated behavior versus real behavior in their validation exercise of survey experiments in comparison to behavioral benchmarks.

clear guidelines, students relied on their own interpretations, with some perceiving the lack of policy as not the technology being not encouraged or allowed.

Notably, when rules are explicit and ChatGPT is unambiguously allowed, top-performing women intend to use the technology at the same rate as men. This finding supports the idea that self-imposed restrictions play a crucial role in AI adoption. Our experiment demonstrates that establishing clear policies can significantly reduce these restrictions, reinforcing the importance of explicit institutional guidelines in preventing the emergence of gender gaps in AI adoption.

Two crucial implications emerge from the findings on policy responses. First, the explicit permission by the authoritative figure to use ChatGPT—in this case, the professor—closes the gender gap in use, suggesting the potential of the policy to prevent the emergence of disparities in the use of the technology. Second, there are potential unintended consequences of banning ChatGPT in the classroom. Such a prohibition, intended to maintain a level playing field or address concerns by educators, might inadvertently contribute to a gender gap in AI adoption. By restricting access to this technology, female students could be placed at a disadvantage compared to their male peers, hindering their exposure to and familiarity with AI tools. Taken together, explicit policies can have important implications on students' adoption of AI and potentially influence their prospects for success in a rapidly evolving labor market. Next, we analyze whether this is likely to be the case.

5 Value of generative AI skills in the labor market

To assess the potential labor market consequences of the gender gap in the adoption of generative AI, we examine whether its use is valued by managers in hiring decisions. If employers value the use of generative AI in job candidates, gender gaps in adoption could potentially translate into gender gaps in labor market outcomes. Finding a job after college is a key measure of labor market success, and employers typically cannot directly evaluate how much students have learned in college but rather observe the grades they obtained. While the gender gap in adoption could affect various outcomes, such as how much students learn or the grades they obtain, we focused on a measurable and observable labor market outcome: how managers evaluate job candidates, specifically comparing those who signal possessing generative AI skills to those who do not. Since employers cannot observe the counterfactual of how candidates' grades or learning would have been without generative AI use in college, our survey experiment with managers is uniquely suited to explore how gender differences in AI adoption affect early career success.[20]

5.1 Experiment design and main outcomes

We study hiring decisions in a conjoint-type experimental design, where managers must evaluate hypothetical candidates represented by a short profile.[21] The candidates are applying for a typical job for recent graduates in the managers' company. The profiles contain basic information about the candidates, including gender, signaled through name, grade in a core course of the bachelor's program, skills, degree and age (see Figure A4 for an example of a profile). All job candidates presented to managers are recent NHH graduates as we were interested in knowing the job market prospects for students as similar as possible as those who answered our student survey. The managers were asked the following: "*Please give each candidate a score between 0 and 12 based on how well-qualified you think they are for a typical job for recent graduates in your department/company.*" Thus, our main outcome corresponds to a score from 0 to 12, where 0 corresponds to an average candidate, 6 to a good candidate, and 12 to an exceptional candidate. We opted to start the scale with average as NHH is the top business school in Norway and the students in the profiles have either average or good grades.

Three main dimensions in the profiles were manipulated. First, gender was represented by assigning either a male or a female name. Second, we varied whether the candidate has

[20]A hypothetical vignette experiment was also conducted to determine whether managers would support for promotion workers who are more productive through the use of generative AI. The results suggest that adopting generative AI, when it leads to productivity gains, increases a worker's chances of being promoted. However, due to space constraints and its limited direct relevance to our setting involving individuals soon to enter the labor market, we do not present these findings in the main text. Instead, the results are detailed in Appendix D.3.1.

[21]In a study validating hypothetical survey experiments, Hainmueller et al. (2015) found that, in the context of voting on residence rights of foreigners in Switzerland, the choices stemming from evaluating short hypothetical profiles predict real-life decisions.

generative AI skills. This was represented as a bullet point indicating one of the following skills: either (i) Expertise in MS Office or (ii) Expertise in generative AI (e.g., ChatGPT).[22] Finally, the profile contained the grade and class distribution for a relevant course named "Data Analysis for Economists." Each candidate has one of two possible grade levels: (i) high grades, which are students in the top 30% of their class (represented by grades A or B), and (ii) low grades, which are students below the top 30% with grade C.[23]

As the gender gap in adoption of generative AI among students emerged only at the top of the academic skill distribution, we focus on estimating the returns of signaling generative AI expertise for students with top grades. At the same time, we are interested in quantifying whether an average male student could compensate for lower grades with AI expertise in comparison to women with higher grades but no AI expertise. Therefore, the managers were randomly presented two profiles out of five possible pre-specified types:

1. **Top Woman No AI**: a female candidate with high grades and no generative AI skills.
2. **Top Woman AI**: a female candidate with high grades and generative AI skills.
3. **Top Man No AI**: a male candidate with high grades and no generative AI skills.
4. **Top Man AI**: a male candidate with high grades and generative AI skills.
5. **Low Man AI**: a male candidate with average grades and generative AI skills.[24]

The manager must give a score to each of the two candidates presented. After assigning scores, the managers indicated which of the two they would select for an interview. In addition, for the selected candidate, the managers indicated what percentage they believe the candidate would be able to negotiate on top of the initial salary offer. We use these latter two outcomes for exploratory analysis.

[22]A recent report documents an increase in LinkedIn members globally adding AI skills like ChatGPT and Copilot to their profiles. In 2023, between 10 to 15% of LinkedIn members in industries from Administrative and Support Services to Construction added AI aptitudes to their LinkedIn profiles, with certain professions such as Content Writer reaching over 30% (Microsoft & LinkedIn, 2024).

[23]The decision to show the grades and the distribution was made to mimic the way real applications are presented in Norway, through a transcript where the grade of the student and the class distribution are shown. Moreover, to generate variation in the characteristics, we presented grades A and B as top students, where the distribution was different, but both signaled a student in the top 30%.

[24]We did not include low woman treatments because we did not have a specific hypothesis to test for that group, and opted for having fewer treatments to maximize statistical power.

To examine the value of generative AI skills for interview invitations, the randomization procedure ensured that the majority of participants faced one candidate with AI skills and one candidate without AI skills. As several elements of the profile were manipulated simultaneously, we do not worry about potential experimenter demand effects, as it is unclear for the managers which of the characteristics is the most meaningful for the experimenter (Stantcheva, 2023).[25]

As companies expect that the use of generative AI will become widespread in the near future (Amazon Web Services, 2024), we also measured managers' expected scores for some profiles if the candidates were applying to a job at their company in three years. In this three-year exercise, managers were only presented with one candidate out of two: (i) Top Woman No AI, or (ii) Low Man AI. The goal was to measure whether, in the near future, generative AI skills could compensate for lower grades relative to female candidates with high grades but no AI skills.

5.2 Econometric Specification

Our econometric specification aims to estimate the advantage of signaling generative AI skills in hiring decisions in the conjoint-type experiment:

$$y_i = \beta_0 + \beta_1 \text{Top Woman AI}_i + \beta_2 \text{Top Man No AI}_i + \beta_3 \text{Top Man AI}_i + \beta_4 \text{Low Man AI}_i + X_i \gamma + \epsilon_i \quad (3)$$

The outcome y_i is the score given to the candidate (either in the present or in three years). Note that the baseline group corresponds to Top Woman No AI. The coefficient β_1 provides the score premium for top female candidates with generative AI skills, $\beta_3 - \beta_2$ represents the AI-skill score premium among top male candidates, and β_4 measures the differences in score between a female candidate with high grades and no AI skills relative to a male candidate with low grades and AI skills. As the level of observation in our analysis is each hypothetical candidate evaluated, and each manager evaluates two candidates simultaneously, we include manager fixed effects.

[25]For more details on the randomization, see Appendix D.

5.3 Main Results

Managers gave an average score of around 6.5 to the hypothetical candidates, with 6 representing a "Good candidate" in the scale. Since each manager evaluates two candidates, we ensure that in all our main analyses, the characteristics of the other candidate are balanced on average (see Figure A5). Panel A of Table 4 reports the estimated coefficients of equation (3). Column 1 compares the scores of the present hiring decision across profile types. The estimated coefficient on "Top Woman AI" indicates that a premium in scores exists for top female candidates with generative AI skills, who are evaluated with a score 7.6% higher than female candidates with a similar profile but without generative AI skills. For male candidates, the premium in scores of signaling generative AI expertise—given by the difference between the coefficients on "Top Man AI" and "Top Man No AI"—is close to zero and not statistically significant. Note that the scores for top male candidates with and without AI skills and for top female candidates without AI skills are statistically the same. Thus, our results suggest that female candidates with AI expertise would have an advantage in hiring decisions relative to other top candidates.[26]

We also find that male candidates with average grades and AI skills ("Low Man AI") are graded 11.1% lower than female candidates without AI skills. Grades still play an important role in the evaluation, and men cannot compensate for their lower grades with their AI skills when competing with top female students without AI expertise. Furthermore, we observe that, in expectation, this difference persists at the same level in three years' time (column 2).

An explanation for the premium benefiting only female candidates with generative AI skills is that the generative AI expertise signal might be more informative in women than in men. If adoption and skills in AI are perceived to be rooted in interest and experience in technology or STEM disciplines and women are less likely to be represented in STEM fields (e.g., Breda et al., 2023), differences in beliefs about who uses generative AI may affect the infor-

[26]The premium of using generative AI for top women is lower in magnitude to the returns of higher grades for both men and women. For example, going from grade B to A—both representing students at the top 30% in the class grade distribution—generates an increase in scores of 0.64 points (9.9%) and 0.53 points (8.5%) for female candidates and male candidates, respectively. For men, candidates with grade B have a score 0.44 points (7.6%) higher than candidates with grade C. The increase in score for men of going from grade C (5.86) to A (6.84) is of 16.7% (see Table A4).

mativeness of the signal. In our setting, if a manager believes that female students are not as likely as male students to use generative AI, the signal of expertise might be more informative for women than for men. This is consistent with Bohren et al. (2019) who study evaluations of men and women's contributions on a large online platform in a field experiment. They find that initially, without prior information, there is discrimination in evaluations against women, generating different expectations towards men and women. However, once new information signals come in the form of objective reputation on the platform, a reversal takes place. Women with high enough objective reputation receive higher subjective evaluations than their male counterparts, due to different initial expectations.

To test this hypothesis, we asked managers about their perceptions of the gender gap in generative AI use among students: *"Do you think that male and female students use AI tools to the same extent?"* with choices: "Yes, to the same extent," "No, male students use them more," "No, female students use them more," and "Don't know." Around 30% of managers had correct perceptions of the gap, i.e., that male students use AI tools more than women, while 38% believe men and women have similar use, 29% indicate they do not know, and only 2% believe women use it more than men. In an exploratory analysis, Table A5 shows a breakdown of equation (3) estimates by the subsamples of managers with correct and incorrect perceptions, with the score given to the hypothetical candidate as the dependent variable. A positive premium in scores for women signaling AI skills emerges for managers who have correct perceptions of the gap, yet this is not the case for managers with incorrect perceptions. Moreover, managers with both correct and incorrect perceptions show no significant premium for men. The findings are consistent with the hypothesis that for managers who expect women to use AI less than men, the generative AI signal from a female candidate is stronger than from a male candidate.

Our results in scores over hypothetical candidates hold even when controlling for other characteristics of the profiles and the managers, such as the grade and course distribution of the hypothetical candidate, candidate order, and characteristics of the comparison candidate fixed effects (see Table A6). In Appendix D we discuss additional evidence indicating that managers are more likely to call the candidate with AI skills for an interview when faced with one candidate with and the other without generative AI, holding their grades constant. We also provide suggestive evidence that candidates with generative AI skills can negotiate their salary more than candidates without those skills.

Finally, we validate the findings on the value of generative AI in hiring using two additional survey questions. First, managers indicated their level of agreement with the statement: *"I would prefer to hire a graduate with generative AI skills rather than a similar candidate without generative AI skills,"* with 44.9% of managers agreeing relative to 18.4% disagreeing. For the second statement: *"Having generative AI skills can help a graduate earn a higher salary in their first job,"* 40.9% of managers agreed relative to 18.8% disagreeing (see Figure A6).[27]

Taken together, the evidence suggests that signaling generative AI skills is valuable in hiring decisions, specifically for top women. If top women do not possess generative AI skills, and thus do not signal them, they might be missing opportunities to increase their chances of success in the labor market. Moreover, given recent evidence on the presence of a gender gap in self-promotions (Exley and Kessler, 2022; Murciano-Goroff, 2022), we expect female students o be less likely to signal generative AI skills relative to men with the same level of experience with the technology. Therefore, we consider our findings a lower bound.

6 Discussion and Conclusion

We conducted two surveys with students at NHH Norwegian School of Economics and managers of companies in the sectors where NHH graduates are often employed. We find large gender disparities in adoption of generative AI among students, and evidence that explicit policies banning/allowing generative AI in educational institutions would further widen/close the gender gap. Our findings are mainly driven by female students at the top of the grade distribution. The manager survey indicates that top female students would greatly benefit from acquiring generative AI skills as these are rewarded in hiring decisions. Overall, even though AI skills are valued by employers, our results suggest that low levels of adoption among top women has an effect on their labor market entry prospects.

[27]A significant share of between 35% and 40% of managers responded "Neither agree nor disagree" in both questions.

The implications of these findings could significantly impact the career trajectories of female students. Recent work indicates that while male and female graduates may start with similar earnings in their first jobs, as is the case for NHH graduates (Bertrand et al., 2019), men tend to earn more and advance faster over time (Bertrand et al., 2010; Cortés et al., 2023). Although extensive research has explored why high-skill women's careers lag behind men's (Bertrand, 2020; Goldin, 2014), our findings suggest that generative AI skills can provide a crucial advantage for top female students entering the labor market, potentially mitigating these negative career trends. More broadly, gender disparities in generative AI usage can create additional barriers during the transition to the labor market. These include women not applying for jobs requiring AI skills, not being selected due to a lack of such skills, or missing out on promotions and career advancement opportunities. These outcomes could not only impact individual career prospects but also perpetuate gender imbalances, hindering diversity and inclusion efforts.

As the rapid increase in adoption and capabilities of generative AI technology has prompted companies and institutions to discuss regulations or policies regarding its use, our findings highlight the importance of carefully designing explicit policies for the technology's use. We demonstrate that explicit policies are consequential, as they can either mitigate or exacerbate disparities in usage. In industries where workers benefit from the use of generative AI in the workplace, employers should implement active policies encouraging its use to prevent the emergence of inequalities at work.

Our results also have wider implications regarding whether AI will reduce or exaggerate existing inequalities between high- and low-skill workers. The results from early work suggest that AI can reduce inequalities between workers. An experiment with customer support agents shows that low-skill agents using an AI tool that provides conversational guidance are able to increase the number of issues resolved per hour to the level of high-skill agents, but that high-skill agents reduce their work quality (Brynjolfsson et al., 2025). In education, where human capital skills such as critical thinking and problem solving are being developed, it may be harder for lower-skill students to catch up to top students through the use of AI. However, we still lack evidence on whether AI adoption affects students' learning or grades. While we cannot rule out that top students who use AI would see their quality of schoolwork or learning reduced, we emphasize the key role of policies on how generative AI could be used as a learning aid. Policies aimed at using AI as a complementary tool for learning with clear guidelines on how to use it and how its use would be evaluated could prevent becoming extremely reliant on AI and using it as a replacement of one's own thought or learning processes.

While it is likely too early to draw definitive conclusions, the evidence we present in this paper, along with accumulating data on productivity gains, suggests that generative AI can be beneficial for students entering the labor market. Although some suggest that gender disparities in adoption will naturally disappear over time, our data suggests further consideration. The persistent gender gap in paid subscriptions, even among older student cohorts that have had more time to adapt to the technology, suggests that these disparities may not resolve on their own, especially if paid versions become the ones that best complement students' and workers' own efforts. Instead of relying on organic changes, well-designed policies can reduce the potential for learning losses and top female students with AI skills could gain an advantage in hiring and promotions.

Finally, we suggest three avenues for further research. First, although our sample benefits from a high degree of homogeneity among students, it remains important to explore how our findings generalize to other educational programs and institutions. Second, our study relies on hypothetical experiments due to the challenge of identifying exogenous variation in relevant policies in a large scale. While we believe our findings reflect emerging trends in the value of generative AI across education and labor markets, future research could focus on validating these results in other contexts. Third, it remains unclear whether selective use of generative AI tools, as opposed to intensive use, leads to better output quality for students and workers. Further research is needed to analyze how varying levels of usage intensity affect productivity and output quality.

References

Almås, I., O. Attanasio, and P. Jervis (2023). Economics and measurement: New measures to model decision making. Technical report, National Bureau of Economic Research.

Amazon Web Services (2024). Accelerating AI Skills: Preparing the Workforce for Jobs of the Future. `https://resources.awscloud.com/ai-skills`. [Online; accessed March 2024].

Ameriks, J., J. Briggs, A. Caplin, M. D. Shapiro, and C. Tonetti (2020). Long-term-care utility and late-in-life saving. *Journal of Political Economy 128*(6), 2375–2451.

Andre, P., C. Pizzinelli, C. Roth, and J. Wohlfart (2022). Subjective models of the macroeconomy: Evidence from experts and representative samples. *The Review of Economic Studies 89*(6), 2958–2991.

Avilova, T. and C. Goldin (2018). What can uwe do for economics? In *AEA Papers and Proceedings*, Volume 108, pp. 186–90.

Bernheim, B. D., D. Björkegren, J. Naecker, and M. Pollmann (2022). Causal inference from hypothetical evaluations. Technical report, National Bureau of Economic Research.

Bertrand, M. (2020). Gender in the twenty-first century. In *AEA Papers and proceedings*, Volume 110, pp. 1–24. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Bertrand, M., S. E. Black, S. Jensen, and A. Lleras-Muney (2019). Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway. *The Review of Economic Studies 86*(1), 191–239.

Bertrand, M., C. Goldin, and L. F. Katz (2010). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics 2*(3), 228–55.

Bettinger, E. and R. Slonim (2007). Patience Among Children. *Journal of Public Economics 91*(1-2), 343–363.

Bick, A., A. Blandin, and D. J. Deming (2024). The rapid adoption of generative ai. Technical report, National Bureau of Economic Research.

Bimber, B. (2000). Measuring the gender gap on the internet. *Social science quarterly*, 868–876.

Bohren, J. A., A. Imas, and M. Rosenberg (2019). The dynamics of discrimination: Theory and evidence. *American Economic Review 109*(10), 3395–3436.

Brañas-Garza, P., L. Estepa-Mohedano, D. Jorrat, V. Orozco, and E. Rascón-Ramírez (2021). To pay or not to pay: Measuring risk preferences in lab and field. *Judgment and Decision Making 16*(5), 1290–1313.

Brañas-Garza, P., D. Jorrat, A. M. Espín, and A. Sánchez (2023). Paid and hypothetical time preferences are the same: Lab, field and online evidence. *Experimental Economics 26*(2), 412–434.

Breda, T., J. Grenet, M. Monnet, and C. Van Effenterre (2023). How Effective are Female Role Models in Steering Girls Towards STEM? Evidence from French High Schools. *The Economic Journal*, uead019.

Brynjolfsson, E., D. Li, and L. Raymond (2025). Generative AI at work. *The Quarterly Journal of Economics*.

Brynjolfsson, E. and G. Unger (2023). The Macroeconomics of Artificial Intelligence. https://www.imf.org/en/Publications/fandd/issues/2023/12/Macroeconomics-of-artificial-intelligence-Brynjolfsson-Unger. [Online; accessed 05-December-2023].

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics 129*(3), 1409–1447.

Buser, T., N. Peter, and S. C. Wolter (2017). Gender, competitiveness, and study choices in high school: Evidence from switzerland. *American Economic Review 107*(5), 125–130.

Business Insider (2023). Chatgpt is banned in schools and colleges as plagiarism and misinformation fears spread. Accessed: 2024-09-02.

Caplin, A., D. J. Deming, S. Li, D. J. Martin, P. Marx, B. Weidmann, and K. J. Ye (2024). The abc's of who benefits from working with ai: Ability, beliefs, and calibration. Technical report, National Bureau of Economic Research.

Capraro, V., A. Lentsch, D. Acemoglu, S. Akgun, A. Akhmedova, E. Bilancini, J.-F. Bonnefon, P. Brañas-Garza, L. Butera, K. M. Douglas, et al. (2024). The impact of generative artificial intelligence on socioeconomic inequalities and policy making. *PNAS nexus 3*(6).

Castillo, M., P. J. Ferraro, J. L. Jordan, and R. Petrie (2011). The Today and Tomorrow of Kids: Time Preferences and Educational Outcomes of Children. *Journal of Public Economics 95*(11-12), 1377–1385.

Charness, G. and U. Gneezy (2012). Strong Evidence for Gender Differences in Risk Taking. *Journal of Economic Behavior & Organization 83*(1), 50–58.

Cimpian, J. R., T. H. Kim, and Z. T. McDermott (2020). Understanding persistent gender gaps in STEM. *Science 368*(6497), 1317–1319.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics 129*(4), 1625–1660.

Cortés, P., J. Pan, L. Pilossoph, and B. Zafar (2023, 05). Gender Differences in Job Search and the Earnings Gap: Evidence from Business Majors. *The Quarterly Journal of Economics*.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–474.

Drydakis, N. (2024). Artificial intelligence capital and employment prospects.

Enke, B., R. Rodriguez-Padilla, and F. Zimmermann (2022). Moral universalism: Measurement and economic relevance. *Management Science 68*(5), 3590–3603.

Exley, C. L. and J. B. Kessler (2022). The gender gap in self-promotion. *The Quarterly Journal of Economics 137*(3), 1345–1381.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics 133*(4), 1645–1692.

Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science 69*(4), 1935–1950.

Franco, C., N. Irmert, and S. Isaksson (2025). Does AI help or hurt student learning?

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American economic review 104*(4), 1091–1119.

Hainmueller, J., D. Hangartner, and T. Yamamoto (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences 112*(8), 2395–2400.

Hirshman, S. D. and A. Willén (2022). Does increasing risk widen gender gaps? *NHH Dept. of Economics Discussion Paper* (20).

Humlum, A. and E. Vestergaard (2024). The adoption of chatgpt. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2024-50).

Isaksson (2019). It takes two: Gender differences in group work. *Working Paper*.

Kajackaite, A. and U. Gneezy (2017). Incentives and cheating. *Games and Economic Behavior 102*, 433–444.

Kessler, J. B., C. Low, and C. D. Sullivan (2019). Incentivized resume rating: Eliciting employer preferences without deception. *American Economic Review 109*(11), 3713–3744.

Kimbrough, E. O., E. L. Krupka, R. Kumar, J. M. Murray, A. Ramalingam, S. Sánchez-Franco, O. L. Sarmiento, F. Kee, and R. F. Hunter (2024). On the stability of norms and norm-following propensity: A cross-cultural panel study with adolescents. *Experimental Economics*, 1–28.

Kinnl, K., J. Möller, and A. Walter (2023). Borrowed plumes: The gender gap in claiming credit for teamwork. *Department of Economics Working Paper Series* (345).

Kotek, M. and I. Vranjes (2022). Exploring the antecedents of technostress at work: a meta-analysis.

Kugler, A. D., C. H. Tinsley, and O. Ukhaneva (2021). Choice of majors: Are women really different from men? *Economics of Education Review 81*, 102079.

Liu, Y. and H. Wang (2024). Who on earth is using generative ai?

Magazine, H. (2024). Ai policy and regulation: A new frontier. Accessed: 2024-09-02.

Mas, A. and A. Pallais (2017). Valuing alternative work arrangements. *American Economic Review 107*(12), 3722–3759.

Microsoft & LinkedIn (2024). 2024 Work Trend Index Annual Report. https://assets-c4akfrf5b4d3f4b7.z01.azurefd.net/assets/2024/05/2024_Work_Trend_Index_Annual_Report_6_7_24_666b2e2fafceb.pdf. [Online; accessed July 2024].

Ministry of Education (2024). 2023 studiebarometeret. https://www.studiebarometeret.no/no/artikkel/5. [Online; accessed 15-Jul-2024].

Murciano-Goroff, R. (2022). Missing women in tech: The labor market for highly skilled software engineers. *Management Science 68*(5), 3262–3281.

NHH (2024). 2023 job market survey. https://www.nhh.no/globalassets/studier/jobb-og-karriere/arbeidsmarkedsundersokelsen-nhh-2023.pdf. [Online; accessed 19-April-2024].

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Noy, S. and W. Zhang (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*.

OECD (2018). Bridging the digital gender divide: Include, upskill, innovate. *OECD*.

Ost, B. (2010). The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review 29*(6), 923–934.

Otis, N. G., B. Haas, K. Cranney, S. Delecourt, and R. Koning (2024). Global evidence on gender gaps and generative ai. Technical report, Center for Open Science.

Peng, S., E. Kalliamvakou, P. Cihon, and M. Demirer (2023). The impact of ai on developer productivity: Evidence from github copilot. *arXiv preprint arXiv:2302.06590*.

Phillips, W. A., K. L. Chapman, and P. D. Berry (2004). Size perception is less context-sensitive in males. *Perception 33*(1), 79–86.

Rask, K. and J. Tiefenthaler (2008). The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Economics of Education Review 27*(6), 676–687.

Sadasivan, V. S., A. Kumar, S. Balasubramanian, W. Wang, and S. Feizi (2023). Can ai-generated text be reliably detected? *arXiv preprint arXiv:2303.11156*.

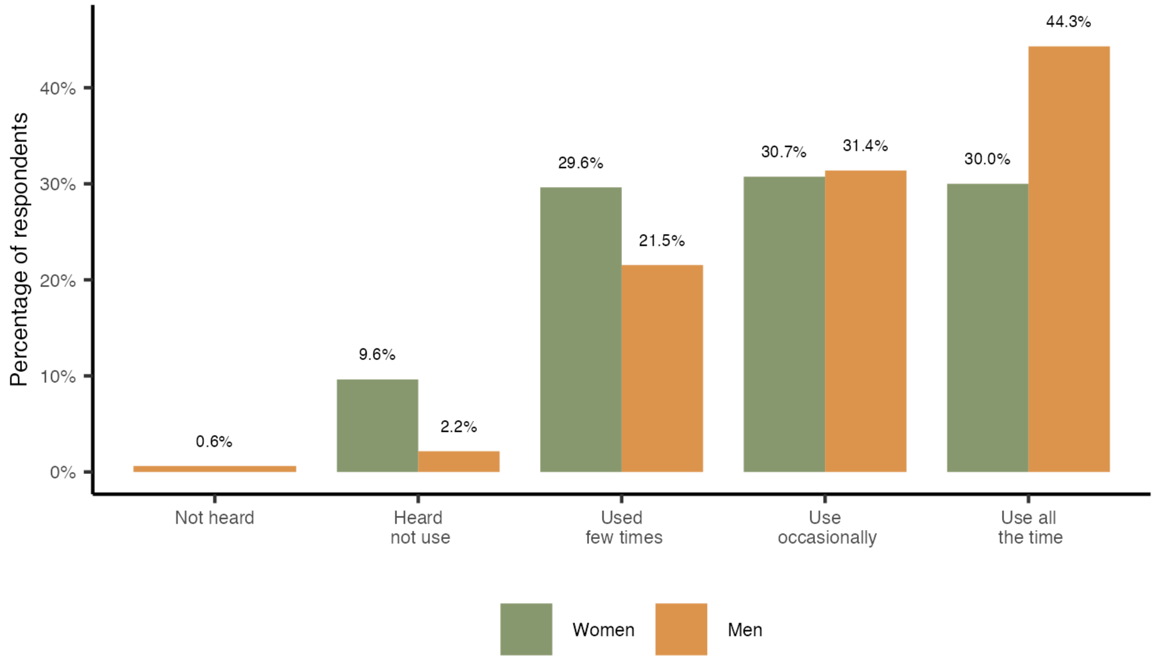
Stango, V. and J. Zinman (2023). We are all behavioural, more, or less: A taxonomy of consumer decision-making. *The Review of Economic Studies 90*(3), 1470–1498.

Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics 15*(1), 205–234.

Ugalde, M. P. (2022). Gender, grade sensitivity, and major choice.

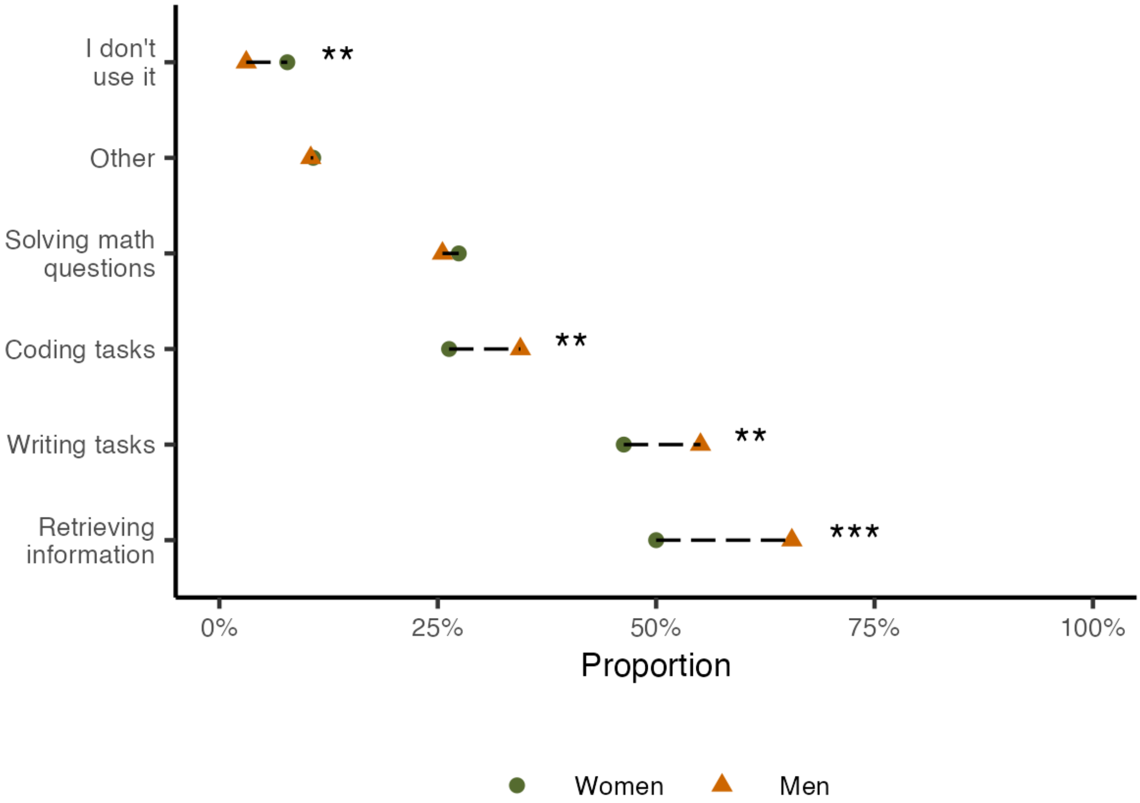
7 Figures

Figure 1: Gender differences in adoption



Notes: The figure shows a bar plot with the percentage of women and men indicating each answer to the question "How familiar are you with ChatGPT or similar tools?." Within gender the percentages across categories add up to 100%. A Wilcoxon rank-sum test shows that the distribution of answers for men and women are different (p<0.01).

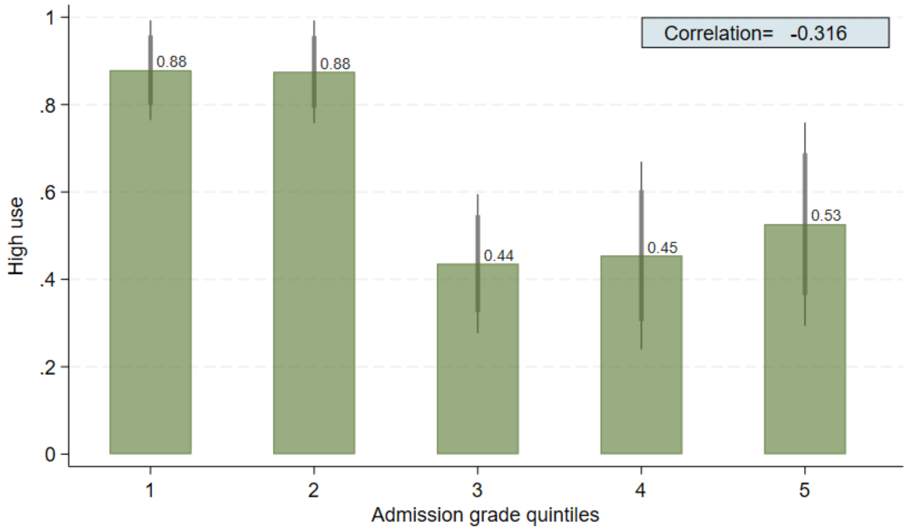
Figure 2: Tasks for which students typically get AI help by gender



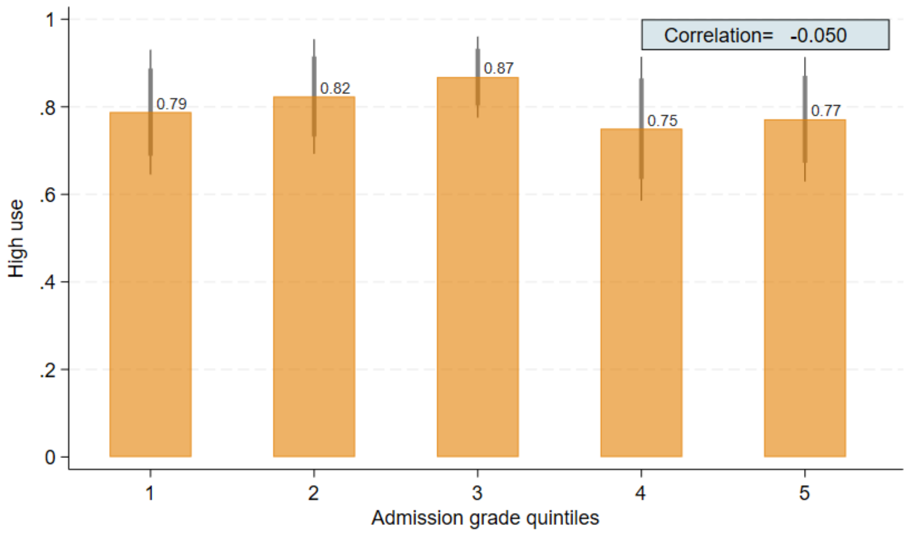
Notes: The figure plots the fraction of students who select the options on the vertical axis in the question: "What type of tasks do you typically ask ChatGPT to help with? (Please select up to the most common three)." The option "I don't use it was added so students who do not use AI could answer the question. The main use they report in the "Other" category is brainstorming. The stars reflect whether the raw gender gap is statistically significant. $* \; p < 0.10$, $** \; p < 0.05$, $*** \; p < 0.01$.

Figure 3: Gender differences in baseline use by admission grade quintiles

(a) Women



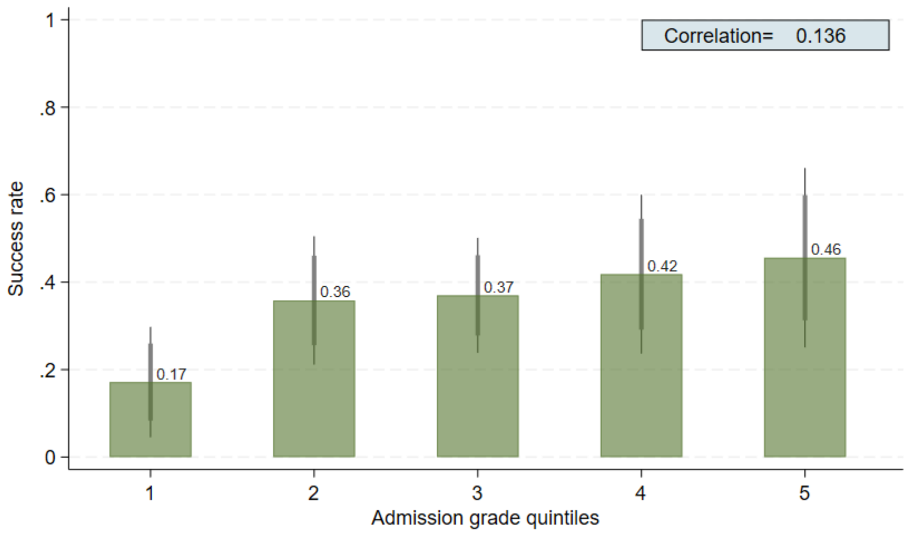
(b) Men



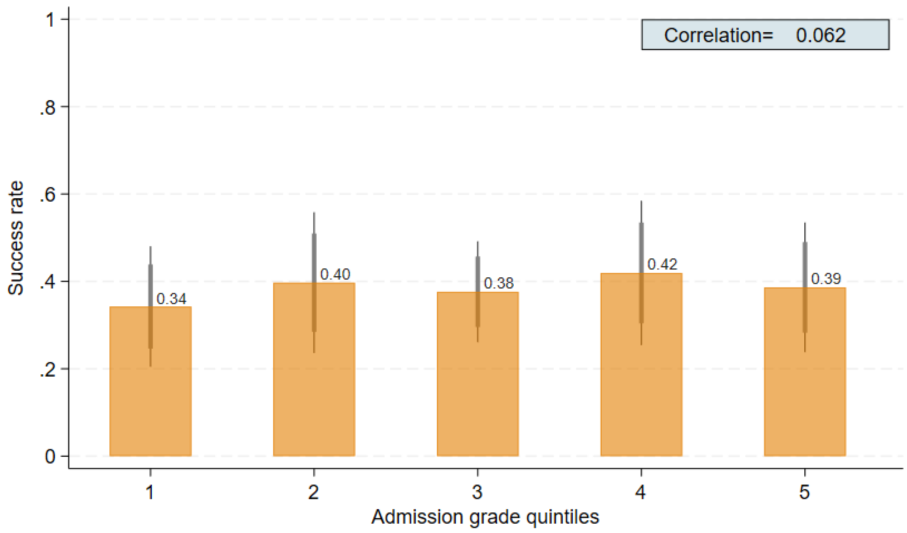
Notes: Panels (a) and (b) show the proportion of women and men, respectively, with high use of generative AI (use occasionally or all the time) across the self-reported admission grade quintiles (328/595 respondents, of which 145 are female and 183 are male). The plots present two sets of confidence bars: 95% (longer) to determine whether the means by quintile are statistically different from zero, and 83% (shorter) to determine whether the means across quintiles are different from each other. A chi-square test with 1 degree of freedom testing the equality of the two correlations gives a statistic equal to 19.25 (p-value=0.000).

Figure 4: Gender differences in prompt succcess by admission grade quintiles

(a) Women

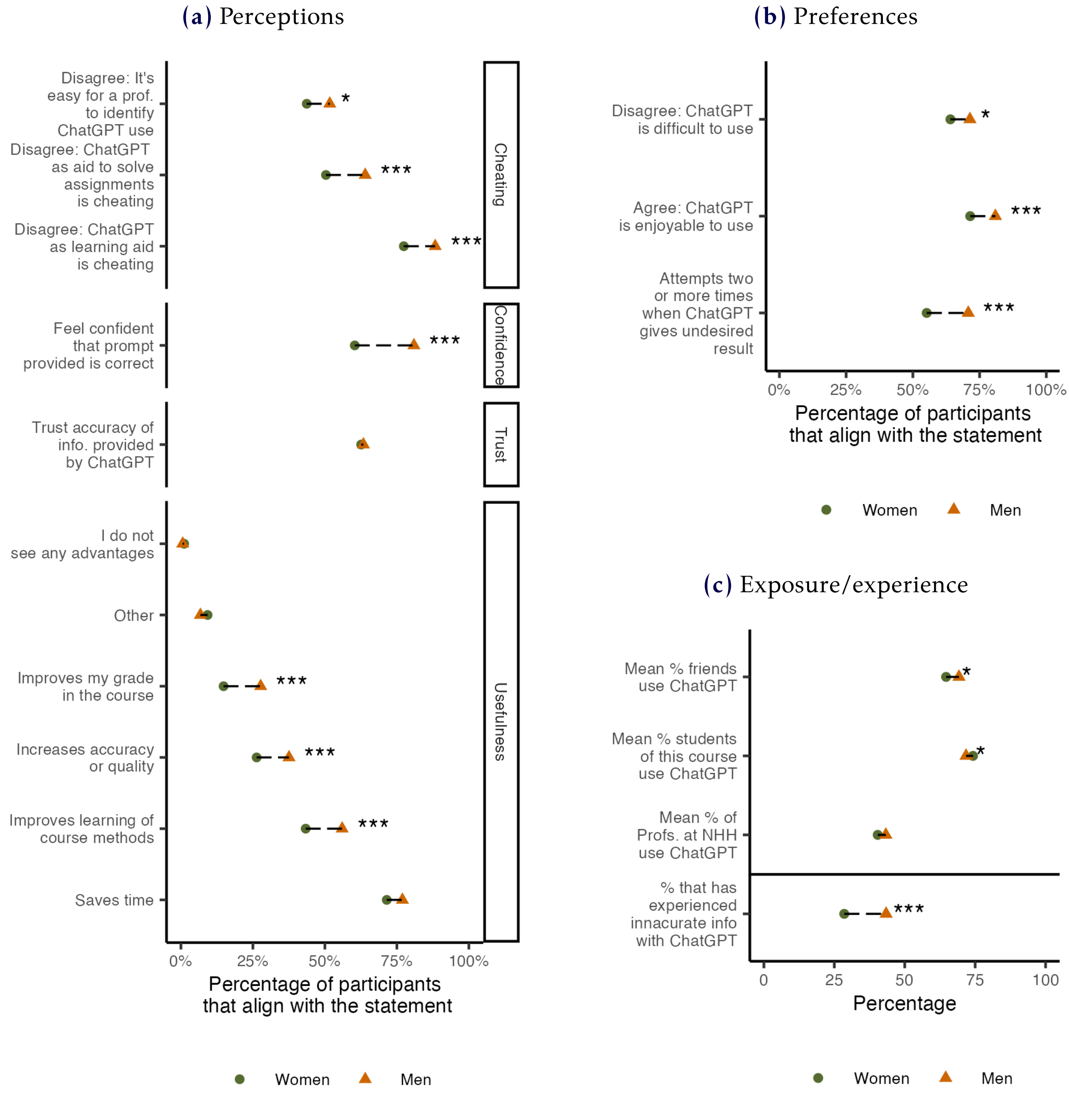


(b) Men



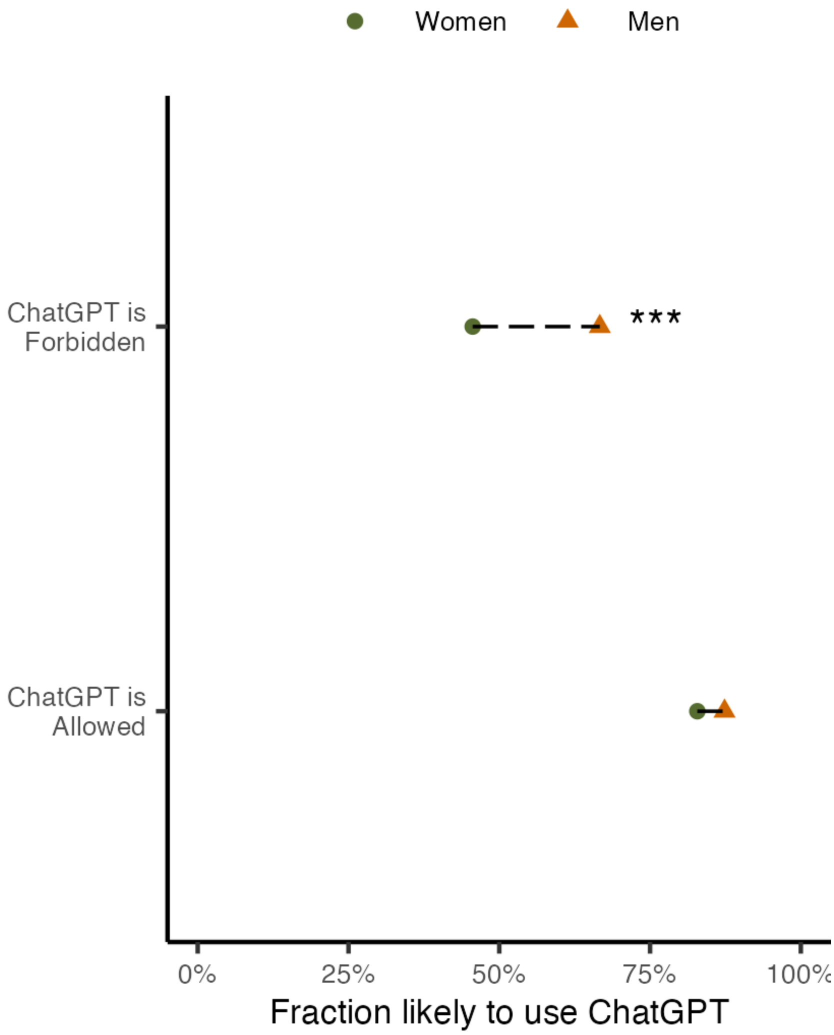
Notes: Panels (a) and (b) show the average success rate in the prompting task for women and men, respectively, across the self-reported admission grade quintiles (328/595 respondents, of which 145 are female and 183 are male). The success rate is calculated running each student's prompt 50 times on ChatGPT and recording how many times the prompt gets the correct answer. The plots present two sets of confidence bars: 95% (longer) to determine whether the means by quintile are statistically different from zero, and 83% (shorter) to determine whether the means across quintiles are different from each other. A chi-square test with 1 degree of freedom testing the equality of the two correlations gives a statistic equal to 0.77 (p-value=0.38).

Figure 5: Potential factors influencing use and skill: gender differences in attitudes



Notes: Panels (a) and (b) show, by gender, the percentage of participants whose answer aligns with each statement on the left of the corresponding graph. Panel (a) shows the results for the statements related to perceptions, while Panel (b) for the statements related to preferences. Panel (c) shows the variables capturing the exposure/experience channel, where the first three rows indicate, by gender, the mean estimate of the percentage of individuals that the participant believes use ChatGPT within the three indicated groups. The last row shows the percentage of participants that indicated to have experienced inaccurate information from ChatGPT. All gender gaps are raw estimates, without any controls. $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

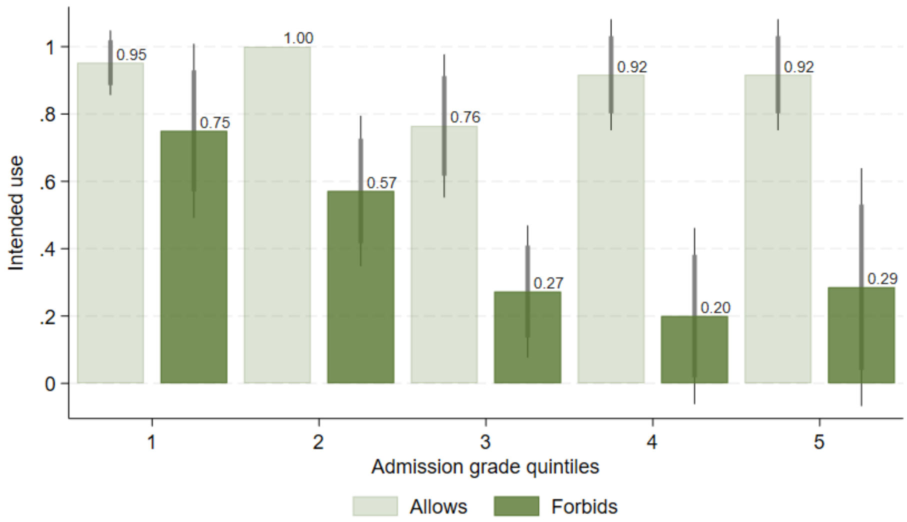
Figure 6: Policy responses



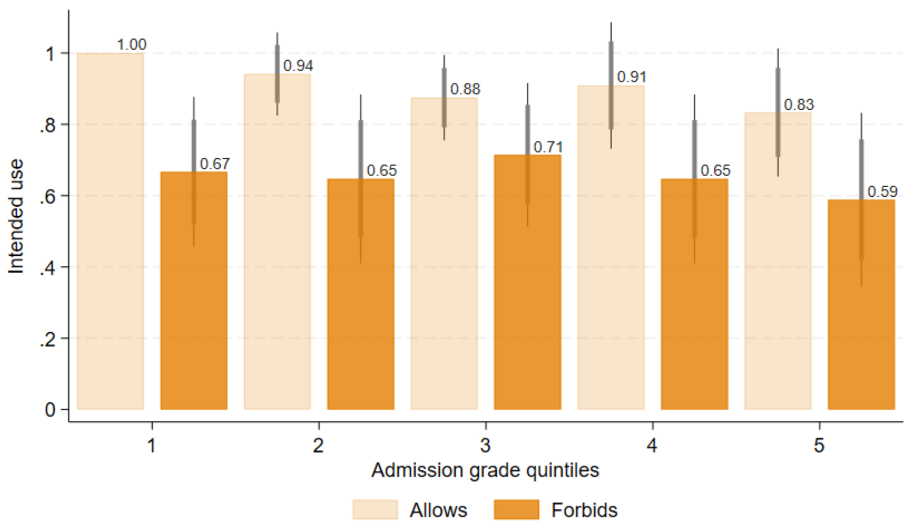
Notes: The figure shows, by gender, the fraction of students that indicated “Somewhat likely” or “Very likely” to the question of how likely would they use ChatGPT in the hypothetical course presented in the vignette experiment. We show the estimates for the two randomly assigned scenarios: professor “forbids” and “allows” treatment.

Figure 7: Gender differences in policy response by admission grade quintiles

(a) Women



(b) Men



Notes: Panels (a) and (b) show the proportion of students who indicated "Somewhat likely" or "Very likely" to the question of how likely would they use ChatGPT in the hypothetical course presented in the vignette experiment for women and men, respectively, and across the self-reported admission grade quintiles (328/595 respondents, of which 145 are female and 183 are male). In brighter colors is the intended use in the professor "allows" treatment, whereas in darker colors is the intended use in the "forbids" treatment. The plots present two sets of confidence bars: 95% (longer) to determine whether the means by quintile are statistically different from zero, and 83% (shorter) to determine whether the means across quintiles are different from each other.

8 Tables

Table 1: Gender differences in use

Panel A: Use ChatGPT occasionally or all the time (adoption)

	Use occasionally/ all the time	Has a subscription	
	(1)	(2) Free	(3) Paid
Male	0.150***	-0.071*	0.126***
	(0.038)	(0.037)	(0.030)
Constant	0.607***	0.326***	0.107***
	(0.030)	(0.029)	(0.019)
Controls	No	No	No
Observations	595	595	595

Panel B: Prompting skills

	Success rate	Time spent (seconds)	No. of characters
	(1)	(2)	(3)
Male	0.094***	1.073	31.646***
	(0.034)	(5.774)	(9.903)
Constant	0.278***	129.000***	145.363***
	(0.024)	(4.417)	(7.162)
Controls	No	No	No
Observations	595	595	595

Notes: Estimates from specification 1 without controls. Panel A Column 1 shows the gender gap in adoption using the binary variable equal to 1 if the students report using AI occasionally or all the time and 0 if has only used it few times or never. Panel A Columns 2 and 3 show whether students self-report having a free or paid subscription to a generative AI chatbot such as ChatGPT. Panel B Column 1 reports gender gaps in the success rate of the prompts in getting the correct name of the visual phenomenon. The success rate is calculated running each student's prompt 50 times on ChatGPT and recording how many times the prompt gets the correct answer. Panel B Columns 2 and 3 show, respectively, time spent writing the prompt and number of characters written. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Gender difference in adoption and skill adding controls

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Use ChatGPT occasionally or all the time (adoption)						
Male	0.150***	0.074*	0.050	0.034	0.080**	0.008
	(0.038)	(0.039)	(0.033)	(0.038)	(0.034)	(0.036)
Constant	0.607***	0.623	0.598***	0.166	0.221***	0.147
	(0.030)	(0.380)	(0.143)	(0.188)	(0.074)	(0.339)
Controls	None	Academic, risk & time	Preferences	Perceptions	Exposure/ experience	All
Observations	595	595	595	595	595	595
Panel B: Paid subscription						
Male	0.126***	0.075**	0.081***	0.046	0.108***	0.035
	(0.030)	(0.034)	(0.029)	(0.031)	(0.030)	(0.033)
Constant	0.107***	-0.586*	0.042	0.340	-0.136**	-0.732*
	(0.019)	(0.334)	(0.086)	(0.222)	(0.061)	(0.384)
Controls	None	Academic, risk & time	Preferences	Perceptions	Exposure/ experience	All
Observations	595	595	595	595	595	595
Panel C: Prompting success rate						
Male	0.094***	0.079**	0.085**	0.108***	0.085**	0.085**
	(0.034)	(0.039)	(0.035)	(0.031)	(0.036)	(0.035)
Constant	0.278***	-0.729*	0.521***	0.867***	0.453***	0.174
	(0.024)	(0.382)	(0.187)	(0.128)	(0.078)	(0.341)
Controls	None	Baseline use, academic, risk & time	Preferences	Perceptions	Exposure/ experience	All
Observations	595	595	595	595	595	595

Notes: Panels A and B show point estimates on gender gaps in self-reported adoption and having a paid subscription. Panel C shows point estimates on gender gaps in the success rate of the prompt provided by students. Each column indicates what control variables are included in the regression at the bottom of the column. Column 1 presents raw estimates and Column 6 includes all controls. Columns 2-5 add groups of controls one by one. Academic controls include year in college, admission grade and an indicator for whether the admission grade is missing. Risk and time preferences are collected using the survey questions from the World Preferences Survey. Preferences include questions on whether students enjoy or find it difficult to use ChatGPT, as well as a measure of persistence in using ChatGPT. Perceptions include views on whether ChatGPT is equivalent to cheating, how useful it is, trust and overconfidence in own ChatGPT skills. Exposure/experience refers to what fraction of their friends, other students in their class and NHH professors use ChatGPT. $* p < 0.10$, $** p < 0.05$, $*** p < 0.01$.

Table 3: Policy responses to forbidding or allowing ChatGPT

	Intended use					
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.045	-0.050	-0.028	-0.053	-0.001	-0.082*
	(0.042)	(0.041)	(0.040)	(0.042)	(0.040)	(0.045)
ChatGPT forbidden	-0.372***	-0.376***	-0.391***	-0.382***	-0.356***	-0.384***
	(0.054)	(0.048)	(0.051)	(0.049)	(0.050)	(0.049)
Male × ChatGPT forbidden	0.166**	0.182***	0.198***	0.190***	0.173***	0.204***
	(0.071)	(0.066)	(0.067)	(0.065)	(0.065)	(0.065)
Constant	0.828***	0.640*	0.640***	0.308	0.529***	0.183
	(0.033)	(0.376)	(0.201)	(0.261)	(0.078)	(0.535)
Controls	None	Baseline use, academic, risk & time	Preferences	Perceptions	Exposure/ experience	All
Observations	595	595	595	595	595	595

Notes: The table shows point estimates from specification 2 on gender gaps in responses to the professor "allows" or "forbids" policies. Each column title indicates what control variables are included in the regression. Column 1 presents raw estimates and Column 6 includes all controls Columns 2-5 add groups of controls one by one. Academic controls include year in college, admission grade and an indicator for whether the admission grade is missing. Risk and time preferences are collected using the survey questions from the World Preferences Survey. Preferences include questions on whether students enjoy or find it difficult to use ChatGPT, as well as a measure of persistence in using ChatGPT. Perceptions include views on whether ChatGPT is equivalent to cheating, how useful it is, trust and overconfidence in own ChatGPT skills. Exposure/experience refers to what fraction of their friends, other students in their class and NHH professors use ChatGPT. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: The value of generative AI knowledge in the labor market

Panel A: Hiring decisions

	Score (1)	Score (3y) (2)
Top Woman AI	0.483***	
	(0.143)	
Top Man No AI	-0.010	
	(0.160)	
Top Man AI	0.053	
	(0.153)	
Low Man AI	-0.712***	-0.708***
	(0.156)	(0.163)
Top Woman No AI (mean)	6.386	6.562
Men AI premium (p-value)	0.646	-
Fixed Effects	Manager	None
Observations	2,286	1,143

Panel B: Promotion decisions

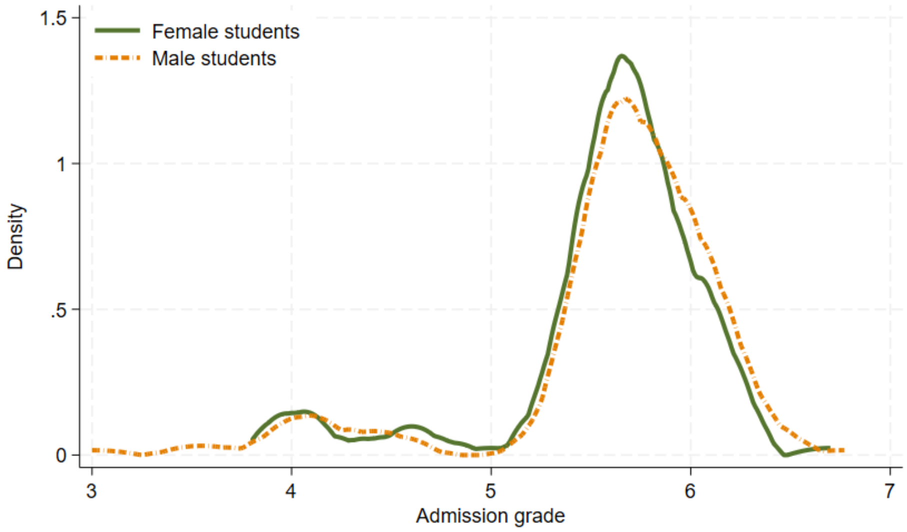
	Fastest worker selected (1)	(2)
Constant	0.740***	0.774***
	(0.018)	(0.031)
Gen AI use: Known	-0.181***	-0.077
	(0.028)	(0.047)
Policy: Not Encouraged		-0.049
		(0.038)
Known × Not Encouraged		-0.154***
		(0.058)
Share Known > 50% (p-value)	0.005	-
Observations	1,143	1,143

Notes: Panel A shows estimates from equation (3), with scores as a dependent variable (values 0 to 12). Column 1 shows the comparison of scores towards the 5 types of hypothetical candidates, represented by an indicator variable for each type of candidate, and with "Top Woman No AI" as the benchmark. As each manager evaluated two candidates, we include manager fixed effects. We report the p-value a two sided test with H_0: $\beta_3 - \beta_2 = 0$, which tests whether there is a non-zero premium of generative AI skills in score for male job candidates. Column 2 shows the comparison of the expected scores given in three years time towards the candidates Top Woman No AI and Low Man AI. Panel B reports the estimates from equation (4) (column 1) and equation (5) (column 2) without controls. We report the p-value for a two-sided test that H_0: $\beta_0 + \beta_1 = 0.5$, that tests whether the proportion that selected the fastest candidate in the "Known" treatment is higher than 50%. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Online Appendix

A Appendix Figures

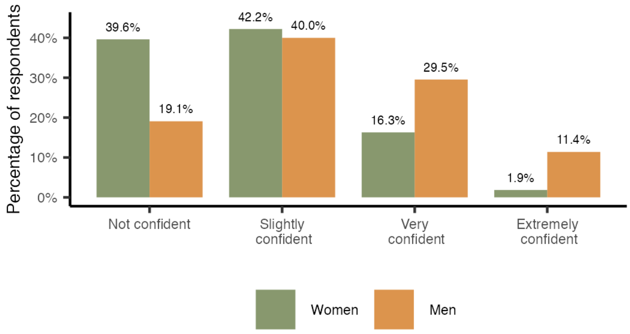
Figure A1: Distribution of admission grades by gender



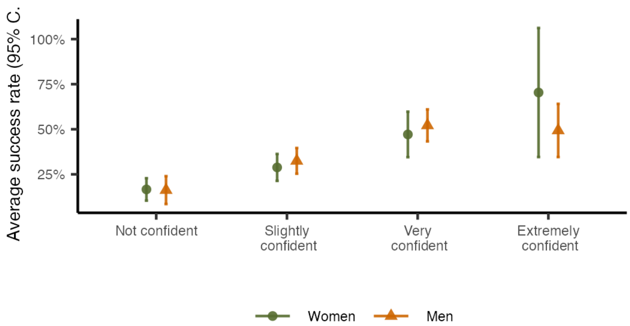
Notes: The plot shows the density of self-reported admission grades separately for male and female students. Admissions take into account the high school GPA that is between 1 and 6. If not admitted to their intended program fresh from high school, students can take extra credits and increase their GPA to apply again. This is the reason why the admission grade can be above 6. NHH admits students through two quotas: 50% get in as first-time applicants and 50% in the ordinary quota which allows taking extra credit courses.

Figure A2: Confidence in own prompt and success rates by level of confidence, and overconfidence

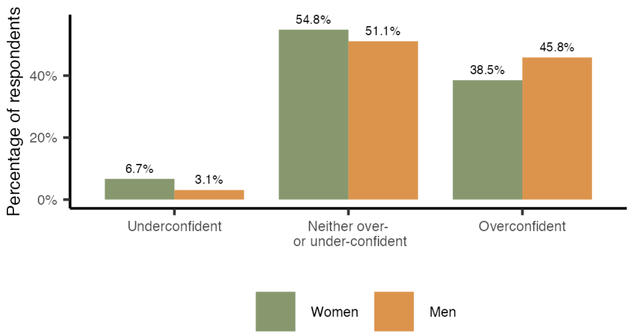
(a) Confidence in own prompt



(b) Success rate by level of confidence



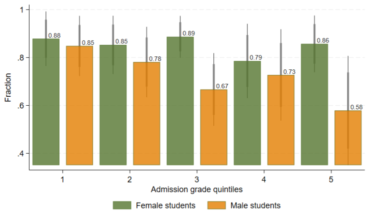
(c) Under or overconfidence in their prompt being successful



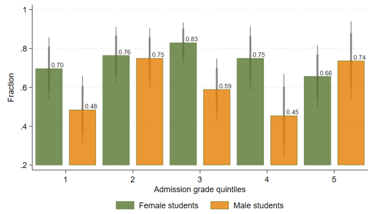
Notes: Panel (a) shows a bar plot with the percentage of women and men indicating each answer to the question "How confident do you feel that the query you just provided will make ChatGPT get the information you need?", which they answered after the prompting skills task. Panel (b) shows the average success rate for each answer option in the confidence question. Panel (c) plots, by gender, a categorical variable where students were classified as follows: (i) underconfident if they indicated "Not confident at all" and had a success rate higher than 0.5, (ii) neither over- nor underconfident if they indicated some level of confidence and their success rate was higher than 0.5, and (iii) overconfident if they indicated some level of confidence but their success rate was lower than 0.5.

Figure A3: Influencing factors across the admission grade quintiles, by gender

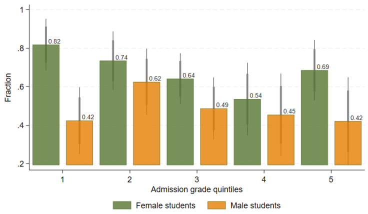
(a) Agree: Enjoyable to use



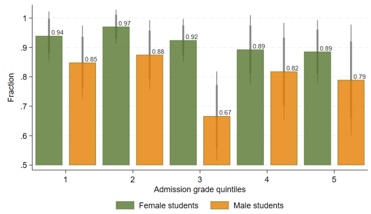
(b) Disagree: Difficult to use



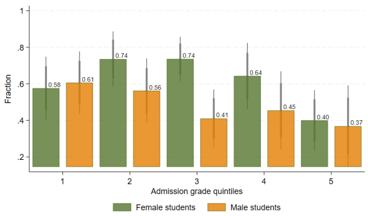
(c) Disagree: Using in assignments is cheating



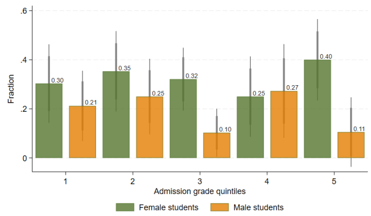
(d) Disagree: Using in learning is cheating



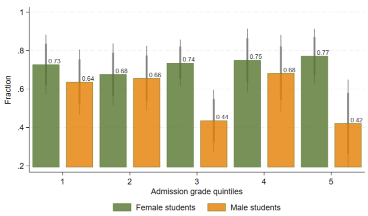
(e) Helps with learning



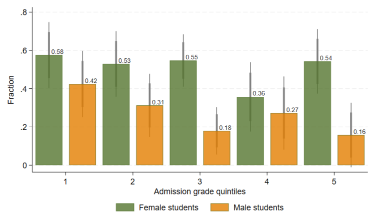
(f) Helps increase grades



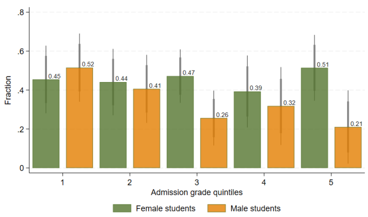
(g) Persistence: ≥ two attempts



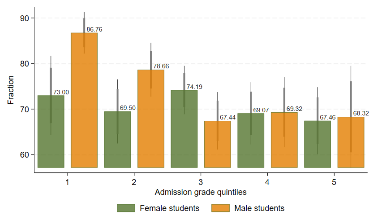
(h) Got inaccurate info. many times



(i) Overconfident in prompt



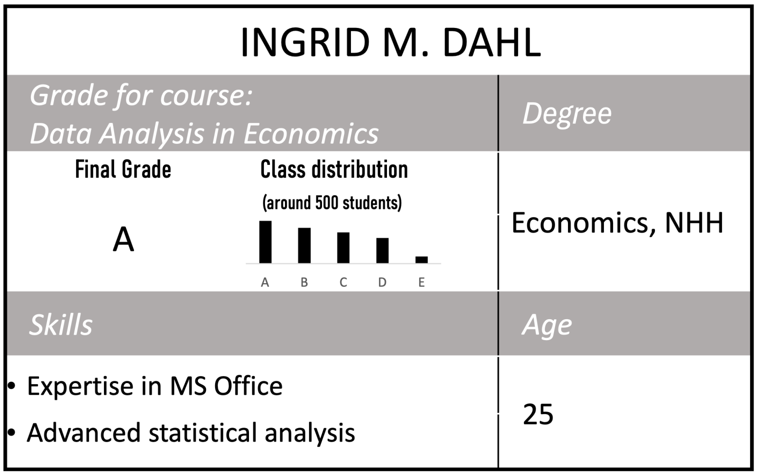
(j) Belief: % peers using AI



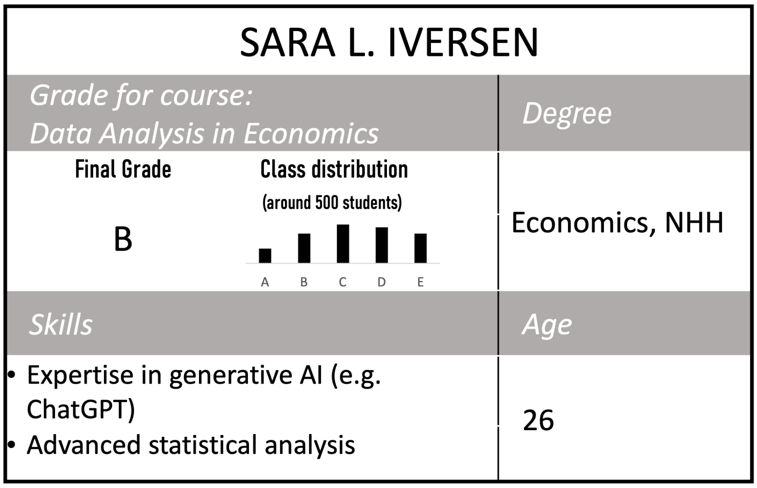
All plotted variables are binary. Agree/disagree include responses somewhat and completely agree/disagree. Persistence measures whether the student makes at least two additional attempts when ChatGPT does not provide the desired answer at first. Overconfidence measures whether the success rate of the prompt is less than 50% and the student reported being very or extremely confidence about the prompt.

Figure A4: Example of short profiles

(a) Profile for: Top Woman No AI

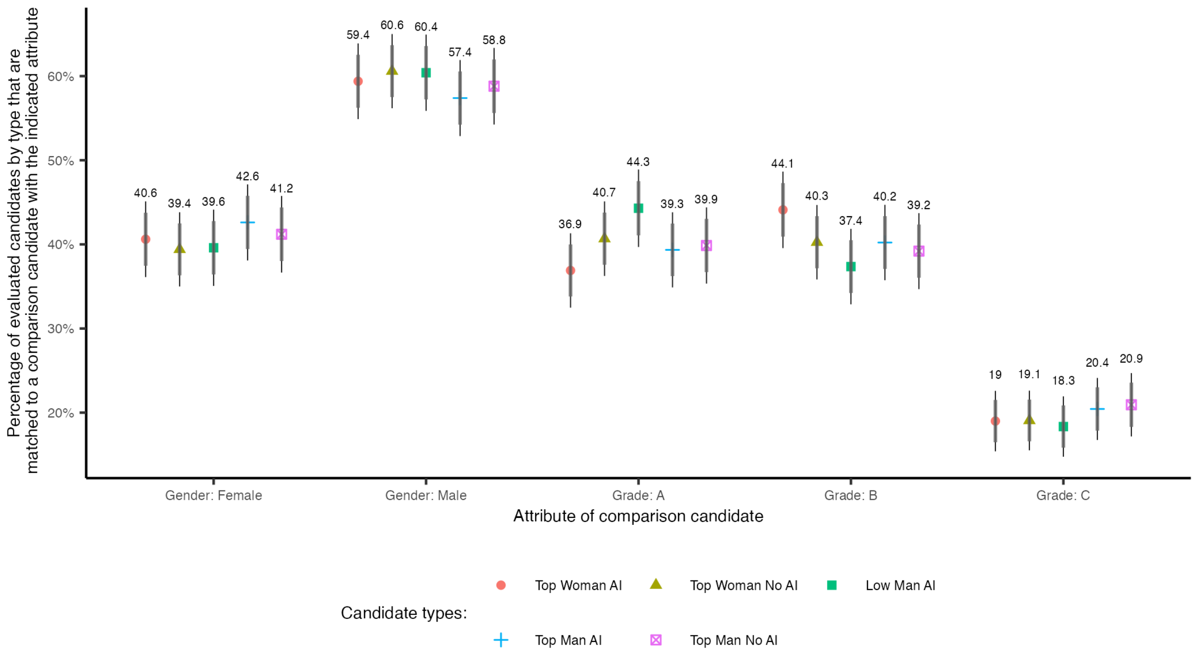


(b) Profile for: Top Woman AI



Notes: The figure shows two examples of profiles of top-performing women—top 30% in class distribution of the course—presented to the managers. In the experiment, the name of the participant, the skills, the grade and grade distribution, and the age are randomly varied. In Figure (a), we show a high-skill female candidate without generative AI skills, and in Figure (b), we show a high-skill female candidate with generative AI skills. The profiles shown to managers were in Norwegian. The grade distribution is usually shown in the transcripts that employer evaluate when hiring new graduates.

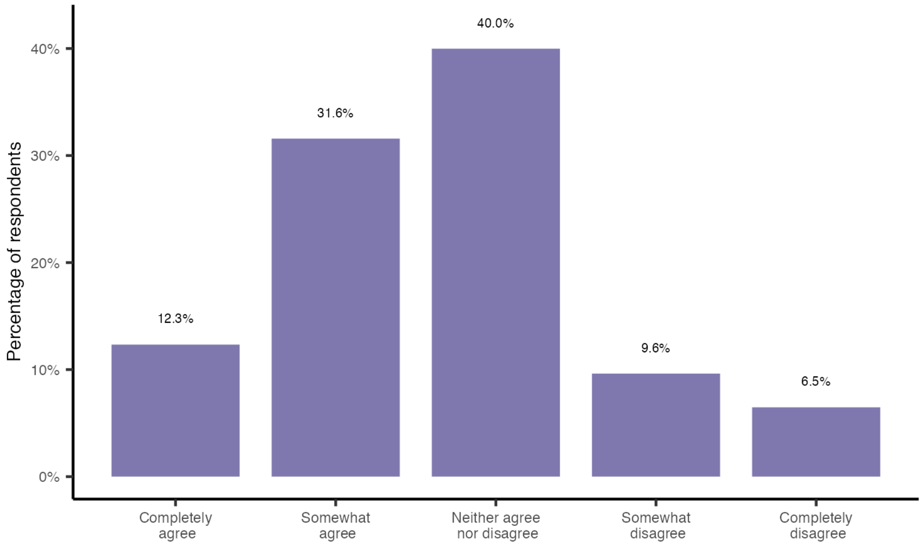
Figure A5: Balance analysis for the characteristics of the other candidate not being specifically evaluated



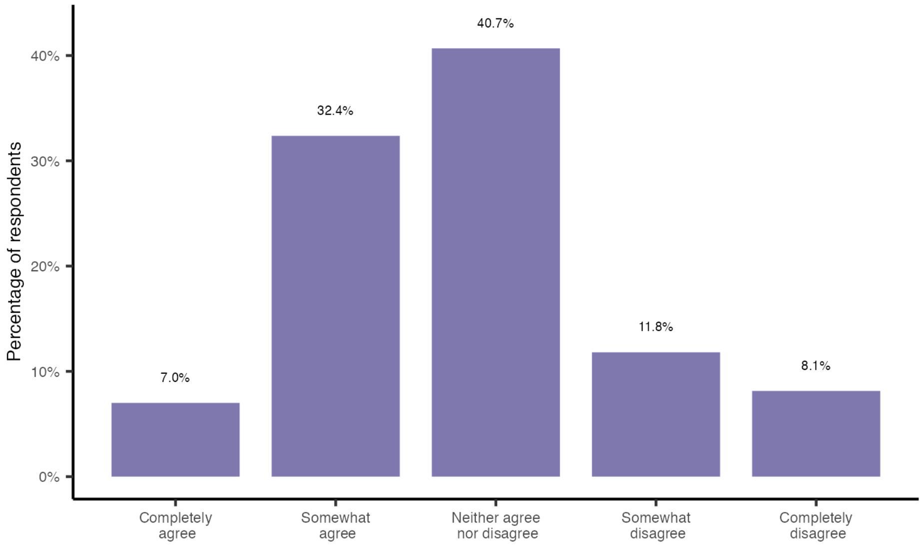
Notes: As the manager evaluates two candidates, we compare the characteristics of the "comparison candidate"—who is not being specifically evaluated—across the different "evaluated candidate" types. The figure shows for each evaluated candidate type the percentage of evaluated candidates who have a comparison candidate with the attribute indicated in the x-axis. The plots present two sets of confidence bars: 95% (longer) to determine whether for each attribute the proportion by type are statistically different from zero, and 83% (shorter) to determine whether for each attribute the proportion by type are are different from each other. The relevant comparisons for our analysis are the preregistered comparisons of "Top Woman AI vs Top Woman No AI", "Top Man AI vs Top Man No AI" and "Top Woman No AI vs Low Man AI". The lowest p-value testing the difference in proportions in all relevant comparisons across attributes correspond to the difference between "Top Woman AI and Top Woman No AI" for the attribute Grade: A, which is p=0.23.

Figure A6: Value of generative AI skills in hiring (agreement with statements)

(a) I would prefer to hire a graduate with generative AI skills rather than a similar candidate without generative AI skills.

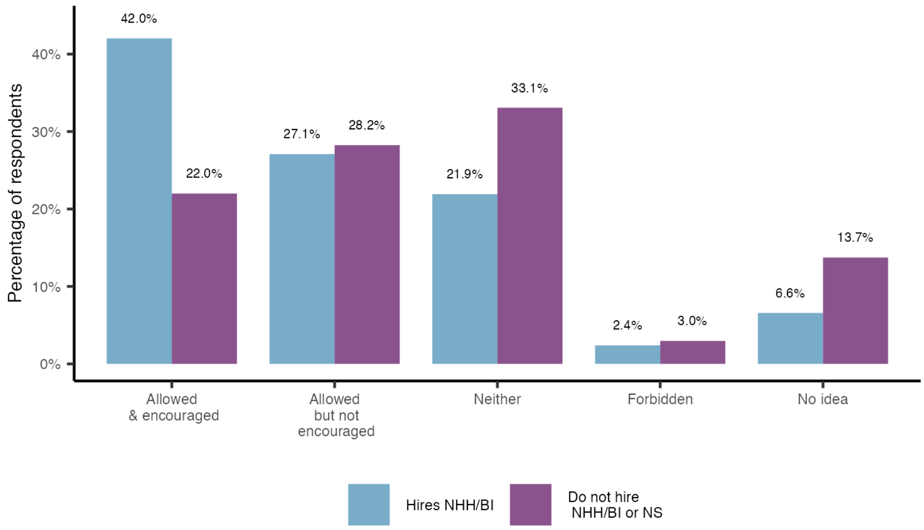


(b) Having generative AI skills can help a graduate earn a higher salary in their first job.



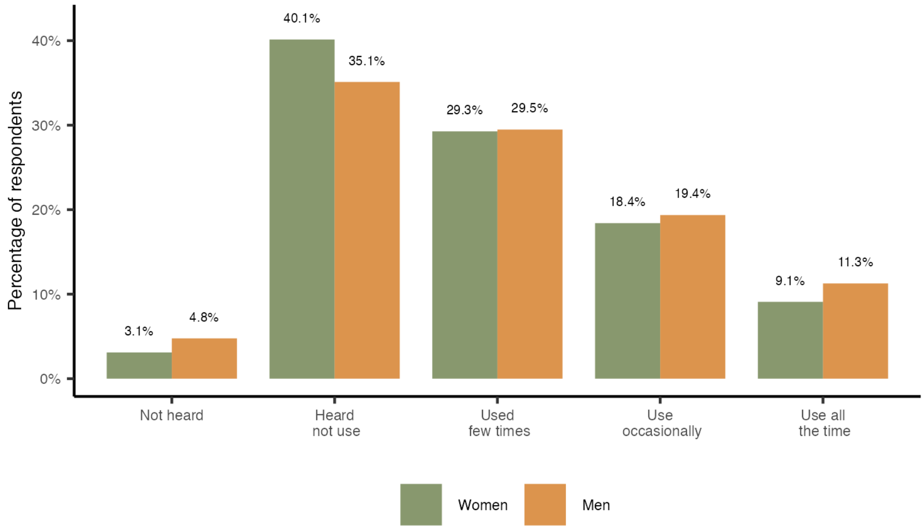
Notes: Panels (a) and (b) show the distribution of the answers of the extent to which managers agree/disagree the statements indicated in subcaptions.

Figure A7: Company's policy by whether the company usually hires NHH/BI students



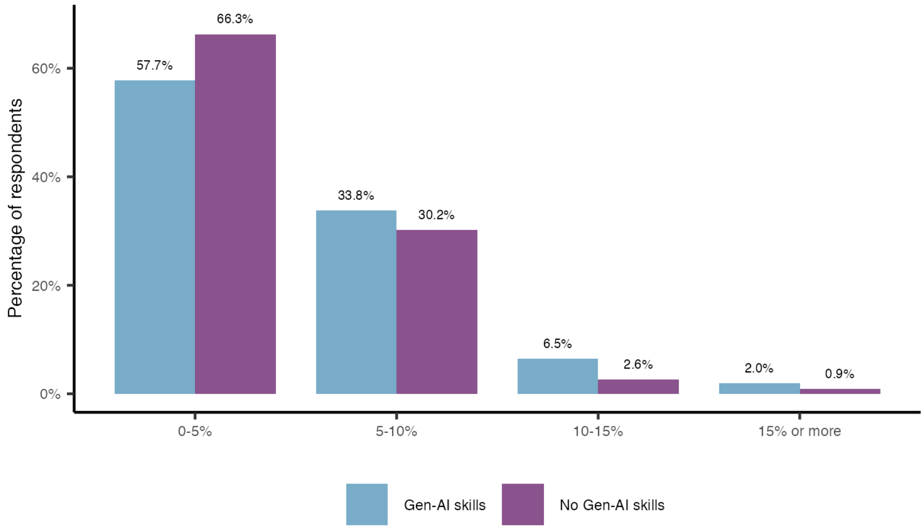
Notes: Managers are asked in the survey: "Does your company/your department employ newly graduated candidates with a master's degree in economics and administration (for example, candidates with a master's degree from NHH or BI)?" We split the sample into two, for participants who answered "Yes" and participants who answered either "No" or "Don't know". For each subsample, we plot their answers to the question: "What is your company's attitude towards the use of generative AI tools at work?" Within each subsample (each color), the bars add up to 100%.

Figure A8: Familiarity of managers with generative AI, by gender of the manager



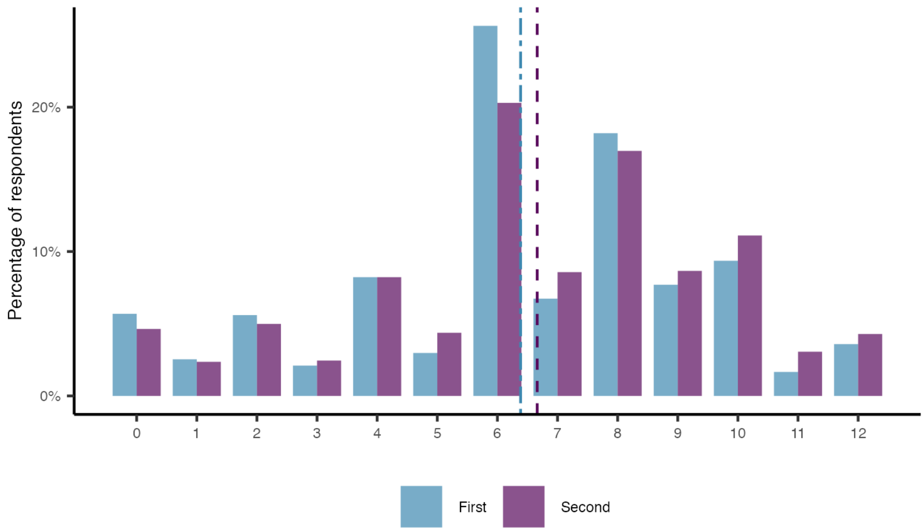
Notes: The figure shows a bar plot with the percentage of female and male managers indicating each answer to the question "How familiar are you with ChatGPT or similar tools?." Within gender the percentages across answer options add up to 100%.

Figure A9: Salary negotiation potential of a candidate with and without generative AI skills



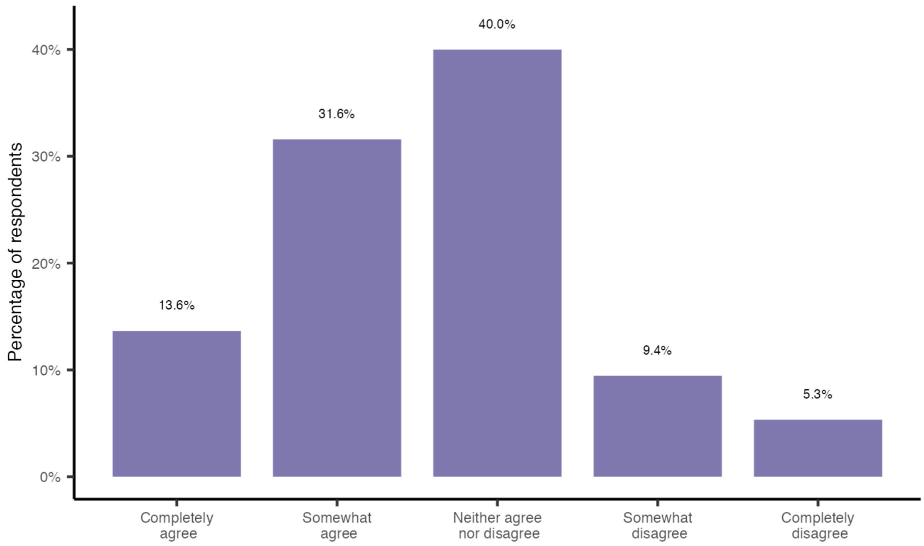
Notes: We consider the subsample of candidates whose evaluating manager faced two candidates with high grades where one candidate has generative AI skills and the other does not have generative AI skills. The bar plot shows the salary negotiation potential conditional on the candidate being selected for the interview for the indicated subsample. The plot represents the answers to the question: "Imagine that the selected candidate is offered the position and receives an offer of a starting salary. The candidate can negotiate the starting salary. What do you think is the maximum starting salary this candidate will be able to get in this job?." In the plot, we show two distributions: (i) when the candidate selected has generative AI skills and (ii) when the candidate selected does not have generative AI skills. Within candidate type, the bars sum up to 100%.

Figure A10: Distribution of scores by whether the candidate was shown first or second



Notes: Each manager evaluates two hypothetical candidates and the order of appearance is randomized. The bar plot shows the distribution of scores given to all hypothetical candidates, by the order whether the candidate was the first or the second candidate the manager evaluated. The dashed lines correspond to the mean score for first and second candidates. The two means are statistically different at the 5% level.

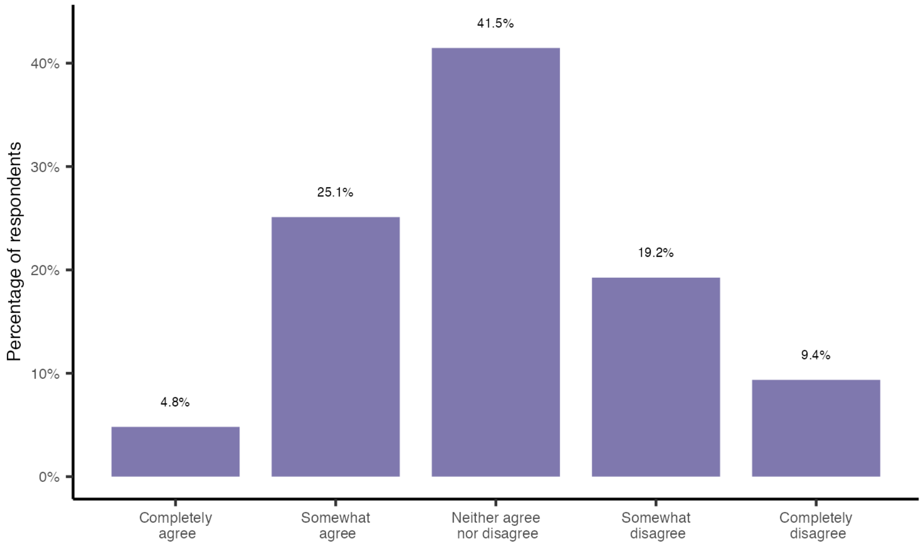
Figure A11: Level of agreement to the following statement: "For my company, I think the advantages of generative AI outweigh the disadvantages"



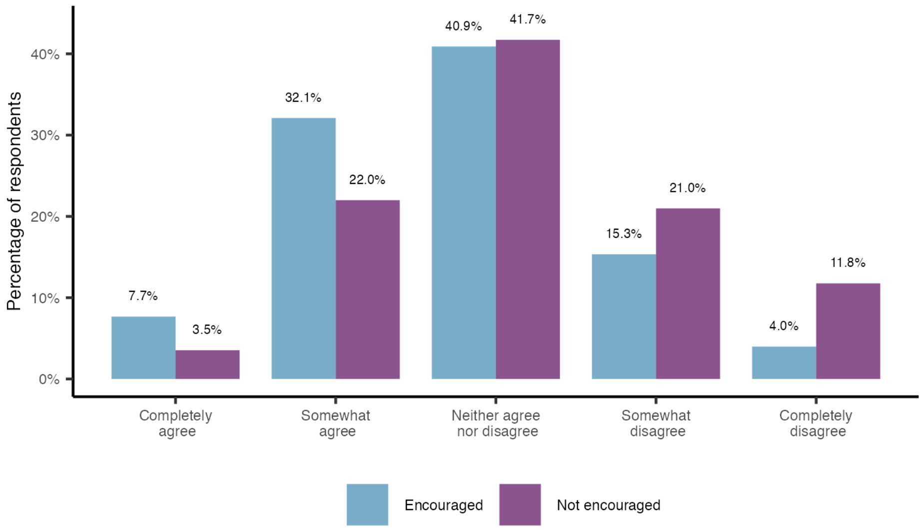
Notes: The plot shows the distributions of the answers to the statement for the full sample of managers. The bars add up to 100%.

Figure A12: Level of agreement to the following statement: "If a student achieves higher grades by using generative AI, it is because the AI tools effectively improve learning, rather than replace individual effort"

(a) For the full sample



(b) By the policy at the manager's company



Notes: Panels (a) and (b) show the distribution of the answers for the statement indicated. Panel (a) shows the distribution of the answers for the full sample. Panel (b) shows the distribution of answers after splitting the sample in two, according to whether in the company where the manager works the use of generative AI in the workplace is allowed and encouraged, or not. For each subset of managers, the bars add up to 100%.

B Appendix Tables

Table A1: Robustness: Gender gap in using generative AI all the time

	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.143***	0.069	0.049	-0.010	0.096***	-0.019
	(0.039)	(0.042)	(0.036)	(0.039)	(0.037)	(0.039)
Constant	0.300***	-0.383	0.423***	0.173	-0.151**	-0.661
	(0.028)	(0.412)	(0.122)	(0.193)	(0.075)	(0.424)
Controls	None	Academic, risk & time	Preferences	Perceptions	Exposure/ experience	All
Observations	595	595	595	595	595	595

Notes: The outcome is defined as 1 if the student uses generative AI all the time and 0 otherwise (including when using occasionally). Each column indicates what control variables are included in the regression at the bottom of the column. Column 1 presents raw estimates and Column 6 includes all controls. Columns 2-5 add groups of controls one by one. Academic controls include year in college, admission grade and an indicator for whether the admission grade is missing. Risk and time preferences are collected using the survey questions from the World Preferences Survey. Preferences include questions on whether students enjoy or find it difficult to use ChatGPT, as well as a measure of persistence in using ChatGPT. Perceptions include views on whether ChatGPT is equivalent to cheating, how useful it is, trust and overconfidence in own ChatGPT skills. Exposure/experience refers to what fraction of their friends, other students in their class and NHH professors use ChatGPT. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Generative AI adoption by course and year in which survey was administered

	Adoption		Has a subscription	
	(1)	(2)	(3)	(4)
	Use occasionally/ all the time	Use all the time	Free	Paid
Panel A: Students in first year course (2023)				
Male	0.273***	0.160***	0.014	0.117***
	(0.058)	(0.047)	(0.052)	(0.032)
Constant	0.338***	0.125***	0.243***	0.022*
	(0.041)	(0.028)	(0.037)	(0.013)
Controls	No	No	No	No
Observations	280	280	280	280
Panel B: Students in third year course (2023)				
Male	-0.031	0.151**	-0.133**	0.095
	(0.045)	(0.070)	(0.065)	(0.063)
Constant	0.897***	0.437***	0.368***	0.241***
	(0.033)	(0.053)	(0.052)	(0.046)
Controls	No	No	No	No
Observations	206	206	206	206
Panel C: Students in second year and master's courses (2024)				
Male	0.036	-0.021	-0.199**	0.152**
	(0.066)	(0.097)	(0.094)	(0.072)
Constant	0.851***	0.553***	0.489***	0.106**
	(0.052)	(0.073)	(0.074)	(0.045)
Controls	No	No	No	No
Observations	109	109	109	109

Notes: Students were approached in one of the core courses in each of the three years of the bachelor's program and in the master's program. Students get the "siviløkonom" degree by doing three years of bachelor's courses and two years of master's. We obtained most of the sample from students in first and third year courses, who answered the survey at the end of November 2023. We approached the second year and master's course in March-April 2023, but few students were present when we conducted the survey. Students who took the survey twice are only counted in the 2023 data. 28 master's students who answered the survey in 2023 are combined with the master's students answering in 2024 in Panel C. Each column contains the estimates from specification 1 using as outcomes the variables in the column labels as defined before. $* p < 0.10$, $** p < 0.05$, $*** p < 0.01$.

Table A3: Gender differences in success rates adding controls for prompt characteristics

	Prompt success rate			
	(1)	(2)	(3)	(4)
Male	0.094^{***}	0.045	0.033	0.009
	(0.034)	(0.031)	(0.030)	(0.028)
Constant	0.278^{***}	0.056^{**}	0.220^{***}	0.070^{***}
	(0.024)	(0.023)	(0.022)	(0.020)
Controls	None	N. Char.	Keywords	Both
Observations	595	595	595	595

Notes: The table shows point estimates from specification 1 on gender gaps in success rates of the prompts provided by students. The first column replicates the main result with no controls presented in Tables 1 and 2. In Tables 2, none of the control variables helped explain the gender gap in success rates of the prompt, so we add here controls for the characteristics of the prompt: number of characters written (column 2), keywords according to the methodology explained in appendix C.2 (column 3), and both controls together (column 4). $* \, p < 0.10$, $** \, p < 0.05$, $*** \, p < 0.01$.

Table A4: OLS estimates of a regression on the score given to a candidate by grade and gender of candidate

	Score	
	(1)	(2)
Woman: grade A	0.642***	0.629***
	(0.159)	(0.157)
Man: grade A	0.371**	0.354**
	(0.151)	(0.149)
Man: grade B	-0.162	-0.172
	(0.155)	(0.154)
Man: grade C	-0.605***	-0.748***
	(0.147)	(0.146)
Manager FEs	Yes	Yes
AI skills FEs	No	Yes
Woman: grade B (mean)	6.465	6.465
Men A-B premium (p-value)	0.000	0.000
Men B-C premium (p-value)	0.004	0.000
Diff Men vs Women A-B (p-value)	0.600	0.618
Diff Men A vs Women A (p-value)	0.073	0.067
Observations	2,286	2,286

Notes: The table shows the estimates from a regression with the dependent variable being the score given to a candidate (0 to 12). The explanatory variables correspond to indicator variables that take value 1 if the candidate was a woman (man) with grade A (B or C), and 0 otherwise. We incorporate manager fixed effects as managers evaluate two candidates. In column 2 we also incorporate a fixed effect of whether the candidate evaluated has AI skills. We report the p-value of a series of two-sided tests for linear combinations of the parameters. First, a test for the significance of the premium of having grade A relative to B for male students (H_0: $\beta_2 - \beta_3 = 0$), second, a test for a premium of having grade B relative to C for male candidates (H_0: $\beta_3 - \beta_4 = 0$). We also provide two additional tests that compare grade returns in scores for men and women. First, we compare the A-B premium for men versus women (H_0: $\beta_2 - \beta_3 - \beta_1 = 0$), and second, we compare whether men and women with grade A are scored differently (H_0: $\beta_1 - \beta_2 = 0$). $* p < 0.10$, $** p < 0.05$, $*** p < 0.01$.

Table A5: Role of perceptions of gender gaps in explaining hiring decisions

Perceptions of Gender Gap:	Correct		Incorrect	
	(1)	(2)	(3)	(4)
Top Woman AI	0.877***	0.826***	0.263	0.205
	(0.279)	(0.274)	(0.164)	(0.160)
Top Man No AI	0.262	0.254	-0.160	-0.148
	(0.319)	(0.316)	(0.184)	(0.179)
Top Man AI	0.573*	0.514	-0.241	-0.323*
	(0.320)	(0.314)	(0.171)	(0.170)
Low Man AI	-0.770**	-0.798**	-0.730***	-0.735***
	(0.344)	(0.348)	(0.170)	(0.166)
Manager FEs	Yes	Yes	Yes	Yes
Grade Distribution FEs	No	Yes	No	Yes
Gender Manager FEs	No	Yes	No	Yes
Order of Candidate FEs	No	Yes	No	Yes
Top Woman No AI (mean)	6.227	6.227	6.462	6.462
Men AI premium (p-value)	0.219	0.304	0.621	0.284
Observations	680	680	1,556	1,556
R^2	0.81	0.81	0.85	0.86

Notes: We report a breakdown of the analysis on evaluations of candidates by whether the managers have correct or incorrect perceptions over the gender gap in generative AI use by students. Columns 1 and 2 show the estimates of equation (3) for the subsample of managers that indicated in a survey question that "Male students use generative AI tools more than female students". Columns 3 and 4 correspond to the same analysis for the subsample of managers who did not indicate that male students use generative AI tools more than female students. Columns 1 and 3 correspond to the preregistered specification with manager fixed effects, whereas columns 2 and 4 include other sets of fixed effects as in our analysis in Table A6. We report the p-value a two sided test with H_0: $\beta_3 - \beta_2 = 0$, which tests whether there is a premium of generative AI skills in score for male students. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table A6: Robustness: OLS estimates of a regression on the score given to a candidate by type of candidate under different sets of controls

	Score					
	(1)	(2)	(3)	(4)	(5)	(6)
Top Woman AI	0.483***	0.415***	0.490***	0.424***	0.483***	0.461***
	(0.143)	(0.141)	(0.142)	(0.140)	(0.143)	(0.140)
Top Man No AI	-0.010	-0.017	0.002	-0.005	0.702***	-0.025
	(0.160)	(0.158)	(0.158)	(0.157)	(0.146)	(0.158)
Top Man AI	0.053	-0.015	0.035	-0.029	0.765***	0.036
	(0.153)	(0.154)	(0.152)	(0.153)	(0.148)	(0.152)
Low Man AI	-0.712***	-0.729***	-0.714***	-0.730***		
	(0.156)	(0.155)	(0.155)	(0.155)		
Manager FEs	Yes	Yes	Yes	Yes	Yes	Yes
Grade Distribution FEs	No	Yes	No	Yes	No	No
Order of candidate FEs	No	No	Yes	Yes	No	No
Comparison cand. gender FEs	No	No	No	No	Yes	No
Comparison cand. grade FEs	No	No	No	No	No	Yes
Men AI premium (p-value)	0.646	0.986	0.809	0.862	0.646	0.656
Observations	2,286	2,286	2,286	2,286	2,286	2,286
R^2	0.84	0.84	0.84	0.84	0.84	0.84

Notes: The table shows the estimates from specification 3 under different sets of fixed effects. Our baseline specification (column 1) includes manager fixed effects, as each manager evaluates two hypothetical candidates. We consider three additional sets of fixed effects in our analysis, which are: the grade distribution of the class of the candidate (column 2), and whether the candidate was presented first or second (column 3). Additionally, column 4 corresponds to the regression including all aforementioned fixed effects. Finally, column 5 and 6 correspond to fixed effects corresponding to the characteristics of the other candidate that is not specifically evaluated, regarding gender and grade, respectively. We report the p-value of a two sided test with H_0: $\beta_3 - \beta_2 = 0$, which tests whether there is a premium of generative AI skills in score for male students. $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

Table A7: Differences on treatment effect of knowing whether the fastest worker uses generative AI by gender of the worker

	Fastest in Promotion
	(1)
Constant	0.811***
	(0.023)
Gen AI use: Known	-0.251***
	(0.038)
Fastest worker: Male	-0.140***
	(0.036)
Known × Male	0.138**
	(0.055)
Observations	1,143
R^2	0.05

Notes: The table reports the estimates from pre-specified equation that evaluates the gender of the fastest worker in the decision to choose which of two workers to recommend for the promotion track. The specification is similar to equation (5), but we change the explanatory variable "Encourage" for an indicator variable "Male" which takes value 1 if the gender of the fastest worker is male and 0 otherwise. The regression is estimated without controls. $* p < 0.10$, $** p < 0.05$, $*** p < 0.01$.

Table A8: The additional treatment effect of the policy experiment when the evaluation is a home exam

	Intended use (1)
Constant	0.878***
	(0.038)
Male	-0.001
	(0.052)
ChatGPT forbidden	-0.417***
	(0.069)
Home Exam	-0.112*
	(0.067)
Male × ChatGPT forbidden	0.210**
	(0.092)
Male × Home Exam	0.102
	(0.085)
Forbids × Home Exam	0.098
	(0.110)
Male × ChatGPT forbidden × Home Exam	-0.098
	(0.143)
Observations	595
R^2	0.13

Notes: The table shows point estimates from specification 2 adding an interaction on all coefficients with an indicator variable named "Home exam" which takes value 1 if the final evaluation of the hypothetical course is given by a home exam. The regression has no additional control variables. The results from the main policy experiment showing that female students respond much more than male students to a forbidding policy hold, regardless of whether the final exam is in person or at home. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Student survey

C.1 Prompting skills measure

We developed a measure of proficiency in the use of generative AI, where we presented students with an image of the "Ebbinghaus illusion" and asked them to write in a text box the query/prompt they would provide to ChatGPT to arrive at the correct official name of the visual phenomenon represented by the image.

Although our novel measure provides insights into differences in skills and use of ChatGPT by gender, there are some limitations to our approach. We decided to use an image that could not be copied and pasted into a text box to ensure that students would engage with the task properly. However, we did not anticipate potential gender differences in visual recognition of images, as documented in some studies (Phillips et al., 2004). These studies are based on small samples, and therefore, we made an attempt to address this concern. In our last data collection (April 2024), we asked the following question to students after showing a picture of the Ebbinghaus Illusion: "What do you think the image presented is about?'' We provided three incorrect choices, a choice where they could indicate "I am not sure,'' and the correct answer, "Equal-size circles appear to be of different sizes.'' We used this question to assess whether men and women in our sample recognized the goal of the image better. However, the data collection resulted in a small sample (n=64), which does not allow us to perform statistical analyses with sufficient power. We found that men were more likely to recognize what the image was about (62% of men versus 45% of women); however, given the small sample size, the difference is not significant, and we cannot draw strong conclusions.

As a consequence, we do not emphasize gender differences in this measure, yet we highlight that women at the top of the distribution are as good at prompting as men, a result that would not be driven by the suggested bias.

C.2 Methodology to identify keywords

We analyze the text data of the prompt exercise performed by the students, where we are interested in determining what makes a prompt successful at achieving the desired result. In addition to the number of characters, we aim to identify what keywords used make a prompt more likely to be successful. We present how the gender gap in prompt success rates disappears when controlling for number of characters and keywords in Table A3.

Each student provided a paragraph in open-ended text as an answer to the prompt exercise question. From each paragraph, we first remove "stopwords," which correspond to common words in the English language which are not informative such as "the," "and," etc. We then process them into tokens, for which we use word stems of the remaining words. For each student, we remove duplicate stems, in order to analyze the impact of an individual word on success rate, and avoid bias emerging from certain words being used more often in the English language. Once we have the text data as unique tokens for each level of observation (student), we generate a matrix of binary variables that indicates the use by a student of each word present in the experiment. Using this matrix, we run a Lasso regression, with tuning parameter optimized using cross-validation, of the success rate as the dependent variable and indicator variables for each word used in the experiment as explanatory variables. This would give us an estimated coefficient for each word, relating its use with the success of the prompt. We estimate the regression 100 times, to get a distribution of coefficients for each word. To select the top 5 and top 10 keywords, we order them by their mean estimated coefficient:

Top Keywords	Mean coefficient (Success Rate)
appear	0.304
ident	0.297
compar	0.295
equal	0.266
effect	0.251
size	0.246
optic	0.244
make	0.244
illus	0.239
due	0.224

C.3 Change in school policy at NHH regarding generative AI

We discuss the introduction of a policy in December 2023 outlining clear guidelines on the use and evaluation of generative AI in its courses, providing greater transparency on how students should engage with this technology. As we collected the data by cohort at different points in time, this allows us to compare adoption before and after the policy was introduced. It is important to highlight that the analysis in this section is descriptive.

Table A2 we compare the adoption across students in different cohorts surveyed at different points in time. First, high use does not reach 100% even among the students surveyed in 2024. Second, the overall gender gap in adoption using the occasionally or all the time measure is mostly driven by students taking a first-year course, where adoption among female students is substantially lower (33.8%) relative to female students taking higher-year courses (at least 85%). Third, the effects of the policy on generative AI tools introduced by NHH in December 2023 may be reflected in the "all the time" measure in Column 2. While the gender gap exists among students surveyed in 2023, it disappears for those surveyed in 2024, suggesting that the policy may have influenced students, especially female students, who were using generative AI occasionally to have a more intensive type of use. We note that we have fewer observations in the cohort surveyed in 2024 so the estimates are noisier. The gender gap in paid subscription, nevertheless, remains economically and statistically significant across students in different stages of the program (see Column 4 of Table A2).

From the findings by cohort, one may wonder whether it is the case that women are not early adopters but quickly catch up. The catch up may happen naturally as higher-year students have had more time to learn about and incorporate generative AI tools in a university setting than first-year students. It can also happen as a result of policies as we see that the gap in "use all the time" exists even among higher year students in 2023 (before the policy), and completely vanishes in 2024 (after the policy). Whether these patterns suggest that gender gaps in generative AI use can close on their own is an important question for future research. However, even if there is catching up, the higher tendency of men to have a paid subscription remains, suggesting that the gaps may not fully close over time.

D Manager survey

D.1 Randomization procedure for conjoint experiment

In the conjoint experiment, each manager is presented with two randomly selected hypothetical candidate profiles. They must evaluate each candidate and select one of them for an interview. This section outlines the randomization procedure.

There are 10 possible profiles (see Figure A13 below), where all have the same probability of being presented to the candidate. Three features of the design allow us to compare candidates with and without AI skills, without concerns of experimenter demand effects.

First, to reduce concerns about experimenter demand effects, we represent the 5 types of candidates of interest through two different distributions of grades for the course: the A and B distributions. In both distributions, we keep fixed that a "Top" student corresponds to a student in the Top 30%. That gives us 10 profiles.

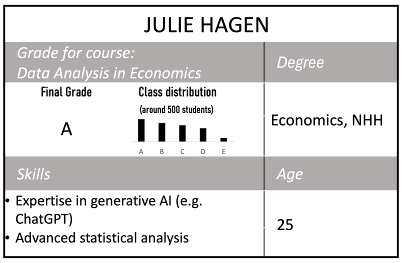
Second, in order to ensure that a manager is likely to face both a candidate with and without generative AI skills, we divide the set of 10 profiles into two subsets:

1. Candidates with AI skills (5 profiles):
   - Top Woman AI: A and B distribution.
   - Top Man AI: A and B distribution.
   - Low Man AI: A distribution.
2. Candidates without AI skills, except for the low man (5 profiles):
   - Top Woman No AI: A and B distribution.
   - Top Man No AI: A and B distribution.
   - Low Man AI: B distribution.

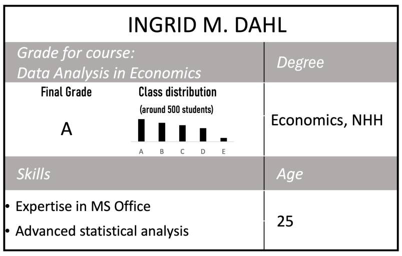
In the experiment, one of the two candidates is drawn from set 1, and the other candidate is drawn from set 2, with equal probability in both. In this way, we ensure that each profile has an equal probability of being presented. Furthermore, a majority of the managers will be presented with a candidate with AI skills and one without, which allows us to study the exploratory variable of which candidate was selected.

Figure A13: Possible profiles.

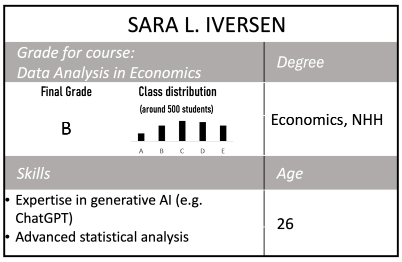
(a) Set 1: Top Woman AI (A)



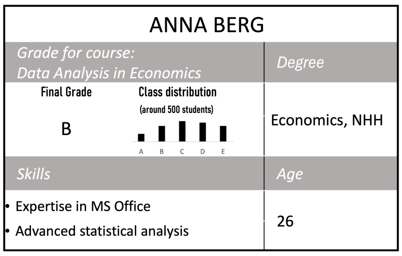
(b) Set 2: Top Woman No AI (A)



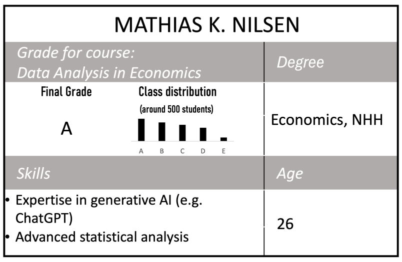
(c) Set 1: Top Woman AI (B)



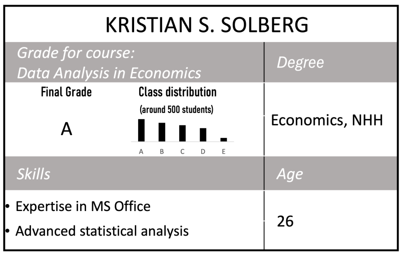
(d) Set 2: Top Woman No AI (B)



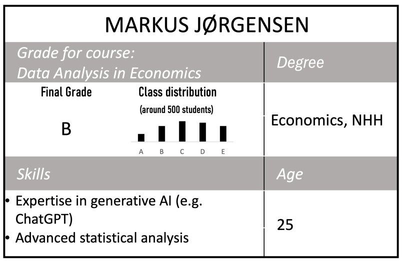
(e) Set 1: Top Man AI (A)



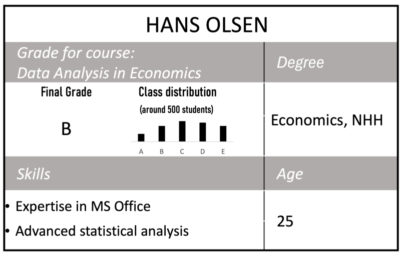
(f) Set 2: Top Man No AI (A)



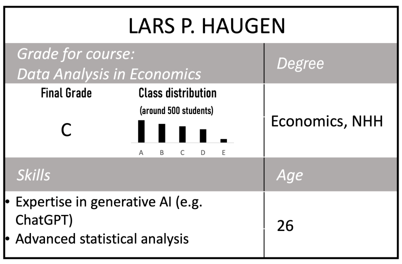
(g) Set 1: Top Man AI (B)



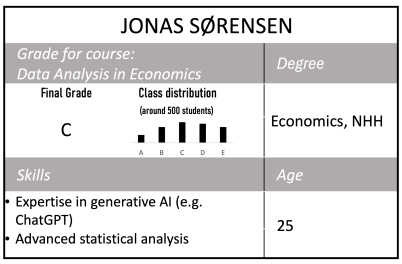
(h) Set 2: Top Man No AI (B)



(i) Set 1: Low Man AI (A)



(j) Set 2: Low Man AI (B)



Notes: The figure show all possible profiles. On the left are the profiles of Set 1: Candidates with AI skills, and on the right are the profiles of Set 2: Candidates mostly without AI skills.

D.2 Additional results

D.2.1 Hiring decisions

Managers were also asked to choose one of the two candidates presented for an interview. To estimate the advantages of expertise in generative AI skills in this scenario, we analyze decisions in the subset of managers who faced similar candidates (high grades) that differ in their generative AI skills, which correspond to a total of 557 managers (almost 50% of the sample).[28] Candidates with generative AI knowledge are 17% more likely to be invited for an interview (300 with AI knowledge against 257 without).

After selecting a candidate for an interview, we ask managers what percentage of the starting salary offer the selected candidate would be able to negotiate. We compare the salary negotiation possibilities for all candidates selected for an interview.[29] When managers face two candidates with generative AI skills (and hence choose a candidate with AI), 30% of them believe that the chose candidate can negotiate the initial salary offer by at least 5%. When managers face one candidate with generative AI skills and one without, if the manager selects the candidate with AI (no AI), 42% (33%) of the managers believe that the candidate can negotiate the initial salary offer by at least 5%.

After selecting a candidate for an interview, we ask managers what percentage of the starting salary offer they believe the selected candidate would be able to negotiate. We compare the perceived salary negotiation potential across all candidates selected for an interview.[30] When managers are faced with two candidates, both possessing generative AI skills, 30% of them believe that the chosen candidate would be able to negotiate the initial salary offer by at least 5%. When managers are faced with one candidate who has generative AI skills and another who does not, if the manager selects the candidate with AI skills, 42% believe the chosen candidate can negotiate at least 5% more in salary. Conversely, if the manager selects the candidate without AI skills, 33% believe that candidate can negotiate the initial

[28]Note that within the individuals that where presented to the candidates, there were male candidates with low grades and AI skills. As we observed in Section 5, grades played a very important role in evaluations. Therefore, we expect individuals with low grades (which are always having AI skills) to be discarded for an interview, and therefore we focus our analysis on candidates with similar profiles.

[29]Due to an implementation error, only 867 out of the 1143 managers responded this question.

[30]Due to an implementation error, only 867 out of the 1143 managers responded to this question.

offer by at least 5%. These findings suggest that candidates with generative AI skills may be perceived as having a stronger position in salary negotiations. However, we caution that the perceived negotiation potential is endogenous to the decision of which candidate to interview so the apparent positive relationship between AI skills and salary negotiation may be driven by omitted variable bias.

Figure A9, which shows the percentage that the selected candidate would be able to negotiate according to the manager, suggests that candidates with generative AI knowledge would be able to negotiate a higher salary than candidates that are also invited to the interview, but do not have generative AI knowledge.

D.3 Interpretation of managers' results as lower bounds

We believe our findings on the rewards of generative AI in the workplace represent a lower bound, as the managers in our sample seem to be underexposed to the technology and are not fully aware of its advantages. Figure A8 shows that around 41% of managers have not used the technology, indicating limited familiarity.

In the analysis of the value of generative AI in hiring decisions, we noted that around 35-40% of participants neither agreed nor disagreed with statements suggesting that AI has a positive effect on candidates facing hiring decisions. Additionally, we asked managers for their agreement with the statement: *"For my company, I think the advantages of generative AI outweigh the disadvantages."* Here, 47% of managers agreed with the statement compared to 16% who disagreed. However, as with other statements, 37% of managers neither agreed nor disagreed (see Figure A11). All together, these findings suggest uncertainty about the advantages and disadvantages of the technology, with a trend towards positive views.

We also obtained a measure that serves as a proxy for their attitudes towards the ethics of using generative AI. We asked them to agree with the statement: *"If a student achieves higher grades by using generative AI, it is because the AI tools effectively improve learning, rather than replace individual effort."* We found that while 30% agreed and 32% disagreed, most participants neither agreed nor disagreed (see Figure A12a). However, as shown in Figure A12b, managers working in companies where the use of generative AI is encouraged have substantially more positive attitudes towards the ethics of using generative AI, suggesting that exposure to the technology might lead to more favorable views.

These findings are consistent with recent research suggesting that current workers in firms misperceive the productivity benefits of using generative AI at work (Humlum and Vestergaard, 2024). We believe that as more companies encourage the use of the technology and it becomes more universally adopted, increased exposure will generate more favorable views towards the technology, making it more valued. Recent surveys in companies with increased exposure to the technology, such as Amazon Web Services (2024), show overwhelming support for the use of generative AI in the workplace.

D.3.1 Promotion decisions

Experimental design and empirical strategy. Each manager was presented with one hypothetical scenario in the workplace, where a manager makes a promotion decision after observing the output of two workers: Daniel or Ida, and Martin or Emma. The scenario is presented as follows:

> *Daniel (Ida) and Martin (Emma) started working at a company at the same time in the same type of job a few years ago. They are assigned a task that they must solve individually. They can use all appropriate resources, including generative AI. Their performance on this task will determine which of the two will be placed on the 'career development track' at the company.*

Recent research highlights productivity benefits of generative AI, including reduced task completion time and enhanced output quality (for a review, see Capraro et al., 2024). Building on this, we design a scenario where we fix output quality and focus on reduced completion time. We study two randomly assigned situations that differ in the ways the workers' performance time on the task is disclosed. First, we want to identify whether a worker known to have used generative AI to complete the task is rewarded when the AI use speeds completion. For this, in our treatment *Known*, participants were told the following: "*Both Daniel (Ida) and Martin (Emma) complete the task with the same level of quality. Daniel (Ida) took 8 days to complete it without generative AI. Martin (Emma) used generative AI and completed it* *in 6 days."* However, the most realistic scenario in a work setting corresponds to a situation in which it is not known whether a worker used generative AI or not. Thus, in our second treatment, *Unknown*, participants were told: *"Both Daniel (Ida) and Martin (Emma) complete the task with the same level of quality. Daniel (Ida) took 8 days to complete it. Martin (Emma) completed it in 6 days."*

Note that in both treatment arms, one worker is 25% faster in completing the task, reflecting the median productivity gain in the use of generative AI in work tasks according to five recent experimental studies (Bick et al., 2024). Moreover, the only difference is that in one scenario it is known who used generative AI, whereas in the other scenario it is not known, corresponding to a more realistic setting, as it is difficult to detect the use of generative AI. The gender of the workers is also randomized through the names of the hypothetical workers. Our main outcome is whether the fastest worker is selected for the "promotion track."

The following econometric specifications are used to estimate whether participants using generative AI in the workplace are rewarded in promotion decisions:

$$y_i = \beta_0 + \beta_1 \text{Known}_i + \epsilon_i \tag{4}$$

$$y_i = \beta_0 + \beta_1 \text{Known}_i + \beta_2 \text{Encouraged}_i + \beta_3 \text{Known}_i \times \text{Encouraged}_i + \epsilon_i \tag{5}$$

The outcome y_i is an indicator variable that takes the value 1 if the fastest worker was selected for the "promotion track" and 0 otherwise. The explanatory variables "Known" and "Encouraged" are indicator variables taking the value 1 if it was known to the manager whether the fastest worker used generative AI, and if the manager currently works at a company with a policy that allows and encourages the use of generative AI, respectively, and 0 otherwise. Equation (4) aims to measure whether the majority of managers select the fastest candidate in both cases, i.e., when generative AI use is known and the more realistic scenario (not known). Equation (5) allows us to perform an exploratory analysis to study whether the exposure of managers to policies at their companies can explain differences across known vs. unknown treatments.

Results. Panel B of Table 4 summarizes our findings on whether productivity gains, due to workers' use of generative AI in the workplace, are rewarded. Column 1 estimates equation (4) and shows the proportion of managers who selected the fastest candidate for the promotion track in each scenario, when it is known that the fastest candidate used generative AI, and the more realistic scenario in which it is not known. To have a point of reference, consider the case when the two workers have the same level of skills and would have therefore performed equally in the absence of generative AI aid. In this case a manager would select a worker randomly, with equal probability (50%). The majority of managers select the fastest candidate in our two treatments, with the mean proportion across the known/unknown treatments equal to 65%. When it is known who used generative AI, 56% of managers select the fastest candidate, a proportion higher than 50% at the 1% significance level. However, in the most realistic scenario where it is unknown whether a worker used it, 74% of managers select the fastest candidate for the promotion track. Thus, if two equally skilled workers are competing for the promotion track, a worker can increase their chances of being selected by using generative AI productively in the workplace.

We take these findings as evidence that using generative AI in the workplace when there are productivity gains would be rewarded in promotion decisions. However, the proportions choosing the fastest candidate are not the same in the known vs. unknown treatments. Even though in both cases a majority selects the fastest candidate, there is a substantial difference (18 pp) in the selection when it is known that the fastest candidate used generative AI relative to when it is not known. A potential explanation for this finding could be the presence of stigma by certain managers associated with the use of generative AI in the workplace. To examine this hypothesis, we perform an exploratory analysis that tests whether company policies affect managers' answers, as the attitudes/policy of the company towards the use of generative AI could formally determine the presence or not of stigma. Managers were asked "*What is your company's attitude towards the use of generative AI tools at work?*" with responses: "It is allowed and encouraged," "It is allowed but not actively encouraged," "It is neither explicitly allowed/encouraged nor prohibited/discouraged," and "It is forbidden." A share of 31% of managers work at companies that allow and encourage the use of generative

AI at work. Column 2 of Panel B shows the point estimates from equation (5). The "Known" coefficient indicates that stigma is absent among the subsample of managers in companies where the use of generative AI is encouraged and permitted. In these cases, managers select the fastest candidate at a similar rate whether the use of generative AI is known or unknown. Consistent with stigma, the difference in selecting the fastest worker is driven by managers who work in companies where generative AI is not encouraged, as highlighted by the negative and significant interaction coefficient "Known × Encouraged."

This finding is particularly informative for two reasons. First, companies that hire NHH graduates are overrepresented among those that allow and encourage generative AI use.[31] Figure A7 shows that 42% of companies that hire NHH/BI graduates allow and encourage the use of generative AI relative to 22% of companies that do not hire graduates from these institutions. We take this as evidence that prospective employees (current NHH students) in our setting face primarily the demand side of the labor market without stigma and that generative AI skills among NHH graduates are likely to be rewarded in the workplace. Second, the heterogeneity in managers' behavior due to the influence of their exposure to companies' policies resonates with our earlier findings on the effects of policies on the gender gap in AI use. As per our previous findings, not only the gender gap in use by current students would disappear with policies that allow/encourage the use of generative AI, but the productivity gains would be rewarded by employers.

D.4 Social desirability bias

We believe social desirability bias does not significantly drive our results. First, the direction of the bias is not clear in our setting, as the use of generative AI can be socially perceived as both beneficial and a strength signal, but also as a potential signal of cheating. Therefore, the presence of a bias would not necessarily skew the estimates in a specific direction. Figure A12a shows our proxy measure of attitudes towards generative AI, where we observe that

[31] We can know this because we asked mangers whether their company employs newly graduated candidates with a master's degree in economics and administration, such as NHH: "Does your company/your department employ newly graduated candidates with a master's degree in economics and administration (for example, candidates with a master's degree from NHH or BI)?" BI is the second biggest business school, after NHH, and their graduates posses a very similar profile to NHH graduates.

while an equal number of managers hold positive and negative attitudes, a majority remains uncertain. Consequently, the direction of social desirability bias is not a major concern in our analysis.

E Pre-registration and Pre-Analysis plan

We pre-registered and developed a pre-analysis plan for our three main results presented in this paper. In this section, we address deviations and indicate the procedures in our reporting of results in comparison to the pre-registrations. The nature of our research question is exploratory, as it is a very novel investigation in an understudied area.

We pre-registered our hypotheses and analyses regarding the gender gap in the use of generative AI, as well as the different primary factors that could potentially explain the gap. We also pre-registered and indicated a PAP for the three experiments of the study. The experiments correspond to (i) a vignette experiment on policies in the student survey, (ii) a conjoint experiment on the evaluation of hypothetical candidates, and (iii) a vignette experiment on a decision over a promotion task involving two hypothetical workers. The pre-registration and pre-analysis plan for the student and manager survey can be found in subsection E.4.

E.1 Gender gap in generative AI use

Use. For the gender gap in use, we explicitly indicated our research question and hypothesis, which correspond to higher use of AI by men relative to women. We also indicated our primary outcome to study the gender gap in current use as the answers to the question "How familiar are you with ChatGPT?" We did not specify how the variable would be constructed, and we focus on what we believe is the most intuitive partition of the answers: low use, referring to no or past use, and high use, referring to continuous use. However, we also use other variables such as the percentage of people who indicated "Use all the time," which reflects regular users, as well as the answers to the question "Do you have a subscription to ChatGPT?" As indicated in Section 3.3, the results are robust to the variables indicated. We did not specify the statistical test to be performed, but we use a standard OLS regression with an indicator variable "Male," which takes value 1 when the gender of the student is male.

Primary factors influencing adoption. For the analysis of the primary factors, we pre-specified three sets of primary factors that we expect to affect the gap in use: (i) preferences, (ii) perceptions, and (iii) experience/exposure. In accordance with this pre-registration, in the paper, we add controls in our main regression specification according to these pre-registered sets, and find that after controls, the gap is not significant. Additionally, we indicated that we would not pre-specify which of the factors would constitute the main driver of the gap. For this, we perform some exploratory analysis of model selection using Lasso.

Sample size. In our pre-registrations, we provided an expected sample size for collection. As we aimed for the most complete sample of our population of NHH graduates, we did not conduct a power analysis but instead attempted to collect as many responses as possible from current students in lectures. The realized sample size comes from all respondents with valid, complete answers to surveys that were possible to collect during our data collection. In the second-year data collection, there was a possibility of students in those lectures sharing courses with first- or third-year courses. The students were asked whether they had completed the survey before, and the responses of those who did were excluded.

E.2 The impact of policies on the gender gap in generative AI use

We pre-specified our main outcome of analysis in the policy experiment as a dummy variable that takes value 1 if an individual indicated likely to use ChatGPT or not. We also indicated our statistical test, an OLS regression of the dummy variable. The interaction with gender in equation (2) follows our motivation stated in the pre-registrations and pre-analysis plan of studying gender differences in adoption, which in this case would be "intended use."

Two adjustments from the pre-analysis plan were made. First, we focus our analysis of intended use on the between-subject variation, meaning the first scenario that students face. This is done to avoid experimenter demand effects and to use a cleaner identification strategy.[32] Second, as indicated in the pre-registration, we randomized along a second dimension: whether the evaluation of the hypothetical course is an in-person exam or a home exam.[33]

[32]In the last data collection on April 10th, participants only faced one single scenario, with an in-person exam.

[33]Respondents that were presented with the home exam scenario were asked a second question: "*Given this*

Table A8 shows the interaction of the explanatory variables as in equation (4), with an indicator variable that takes value 1 if the evaluation is a home exam and 0 otherwise. The additional effect of the final evaluation being a home exam is not significant at a 5% significance level for any of our coefficients of interest. Only the gap when ChatGPT is allowed in a course with an evaluation being a home exam, given by coefficient estimated "Home exam" is significant at the 10% level. The direction is consistent with our interpretation over differences in how women perceive rules, as when the evaluation is a home exam, there are stronger ethical concerns that might prevent women to use it under this type of evaluation. However, the effect is small. As there are no or weak additional effects of type of evaluation on intended use, we focus our analysis and report the results in the paper pooling both types of evaluations.

E.3 Value of generative AI skills in the labor market

Promotion decisions. For the vignette experiment that examines the value of using generative AI at work in promotion decisions, we pre-specified our interest in studying whether the gender of the worker who uses generative AI matters in a manager's decision. The analysis was pre-specified similar to equation (5), yet substituting the variable "Encourage" for an indicator variable for gender "Male."

The pre-registered hypothesis suggested that if stigma exists against the use of generative AI at work that would harm a worker being more productive through its use, the stigma might differ according to whether the worker is a man or a woman. Table A7 shows the estimated coefficients for the pre-registered equations. The coefficient "Known × Male" shows that the treatment effect of knowing whether the fastest worker used generative AI is stronger when the fastest worker is a woman than when it is a man. However, the difference in treatment effect seems to be coming from managers who do not know whether the fastest worker used generative AI, where 81% select the fastest worker for promotion when it is a woman versus 67% when it is a man. On the other hand, for managers who know that

scenario, how likely are you to use ChatGPT during the final exam?" This way, respondents would differentiate the use of ChatGPT throughout the course and during the exam in order to make the measures comparable across different evaluation scenarios.

the fastest worker used generative AI, around 55% of them select the fastest worker regardless of whether it is a man or a woman. The interpretation of this difference in treatment effects as differences in retaliation or stigma according to gender is not necessarily feasible, as the baseline proportion when the manager does not know who used generative AI is not the same, given by the coefficient "Male." As the interpretation of these results requires speculation and steps into discussions outside the main purpose of this paper, we leave it out of the main discussion of the results.

Sample size. In our pre-registrations, we indicated a sample size agreed upon with our survey provider Norstat, which was 1,000 managers in Norway who hire in industries that often employ NHH graduates. The implementation and recruitment were done by the company, and the realized sample constitutes the list of approved completions provided by Norstat that matched the completed answers in our Qualtrics survey. The matching corresponded to a total of 1,143 managers recruited.

E.4 Pre-registrations and PAPs of student and manager surveys

We uploaded the following documents into the AEA RCT registry (https://doi.org/10.1257/rct.12452-2.0) in two stages. For the student survey, we uploaded the documentation on November 6, 2023, ahead of the first data collection later that month. For the manager survey, we uploaded on June 10, 2024, before the end of the data collection by Norstat.

The documents below remove the appendices containing the survey questions to avoid repetition with the questionnaires. They also include the complete history of pre-registrations in AsPredicted.org (University of Pennsylvania, Wharton Credibility Lab) with their respective timestamps. The AsPredicted.org documents are ientified with the header: Confidential - for peer-review only.

As of August 20, 2024, these documents have been made public in the AEA RCT registry.

Pre-Registration

Project: Will Artificial Intelligence get in the way of gender equality?

Daniel Carvajal, Catalina Franco, Siri Isaksson

Date: 02.11.2023

Project Summary

This research project aims to investigate the existence of gender differences in the adoption and use of AI technologies, specifically ChatGPT. Previous studies have highlighted a "Digital Divide," showing disparities in internet usage between men and women (Bimber, 2000; OECD, 2018). Additionally, numerous studies in economics and social sciences have indicated gender-based differences in technology-related career choices (Buser et al., 2014), confidence in using technology and the prevalence of stereotypes (Bordalo et al., 2019). Taken together, this body of evidence suggests that women and men potentially use AI tools differently. This issue becomes particularly relevant as AI has become a widely used in work and school settings, with many potentially benefits including productivity enhancing (Noy & Zhang, 2023) and creativity (Doshi & Hauser, 2023).

This project seeks four purposes: (i) identify whether a gender gap exists in the adoption of ChatGPT, (ii) explore the underlying mechanisms driving any observed disparities, (iii) assess the impact of this gap on productivity, and (iv) evaluate the efficacy of interventions aimed at reducing the gender gap in AI adoption. We attempt to evaluate this in a comprehensive study involving students, faculty and administration staff in NHH.

The first study targets objectives (i) and (ii) with students. First, we collect a series of measures of **ChatGPT usage** to assess whether there is a gender gap in AI use, overall and on different margins of interest. In the event of a detected gap, the study will proceed to identify its underlying causes. For this purpose, we have pinpointed three primary factors influencing ChatGPT usage: *preferences*, *perceptions*, and *exposure/experience*.

In terms of ***preferences***, we aim to measure potential utilitarian costs or benefits associated with ChatGPT usage, examine the role of patience in the use of technology, and investigate any gender-based disparities in rule-following tendencies. Concerning ***perceptions***, our focus will be on four key areas: perceived usefulness of the technology, ethical considerations in ChatGPT usage, perceived risks associated with ChatGPT, and confidence in one's abilities to use the technology. Lastly, we will explore the ***exposure/experience*** factor, analyzing how familiarity and prior exposure to the technology might influence its adoption.

Our goal is to document any gender disparities in ChatGPT usage and to unravel the components that contribute to this gap.

References

Doshi, A. R., & Hauser, O. (2023). Generative artificial intelligence enhances creativity. *Available at SSRN 4535536.*

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*, 187-192.

Bimber, B. (2000). Measuring the gender gap on the Internet. *Social science quarterly*, 868-876.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, *129*(3), 1409-1447.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, *109*(3), 739-773.

OECD (2018). Bridging the digital gender divide: include, upskill, innovate. *Available at: www.oecd.org/digital/bridging-the-digital-gender-divide.pdf*

Participants

The study consists of a survey experiment involving around 650 master and bachelor students from the NHH Norwegian School of Economics. It will be run in November 2023. The survey will be administered through Qualtrics, and participants will be recruited during class hours. All survey answers are anonymous, and the data collected will be used exclusively by the research team.

Survey Design

The survey is structured to take around 7 minutes and includes a series of hypothetical vignettes, a prompt elicitation task, survey questions, and information on demographics and past grades. The primary outcomes are participants' willingness to use ChatGPT under a series of hypothetical scenarios, and their responses to a prompt elicitation task. Secondary outcomes include demographic information and responses to a series of questions about ChatGPT usage, exposure/experience, perceptions, and preferences.

Main outcomes

ChatGPT usage measures:
- Willingness to use ChatGPT. Vignette
- How familiar you are with ChatGPT. Q: 10
- How do you use ChatGPT. Q: 11

Secondary outcomes

Preferences
- Direct utility benefit of using ChatGPT: enjoyable to use. Q: 15
- Direct utility cost of using ChatGPT: difficult to use. Q: 15
- Patience: number of attempts of using ChatGPT. Q: 16
- Rule-following: vignette.

Exposure/Experience
- Prompt elicitation
- Usage of ChatGPT in the surroundings. Q: 9
- Experience inaccurate information. Q: 8
- Reason to start using ChatGPT. Q: 7

Perceptions
- Usefulness/relevance: main advantages. Q: 13, 15
- Ethics: is it cheating. Q: 15
- Risk: professor identifying usage of ChatGPT. Q: 15
- Confidence. Q: Prompt
- Trust accuracy: fixed prompt. Q: 14

Vignette experiment

To assess whether there are gender differences in rule-following, participants will observe two hypothetical scenarios, in a within-subject experiment, where they must indicate their willingness to use ChatGPT in each scenario. Both hypothetical scenarios correspond to the attendance to a specific course. Each scenario differs in the following way:

- Scenario 1: The professor explicitly allows the use of ChatGPT during coursework.
- Scenario 2: The professor explicitly forbids the use of ChatGPT during coursework.

The order of the scenarios will be randomized, allowing for a between subject analysis of the difference in behavior.

Hypotheses

Drawing on insights from a pilot survey in Prolific and existing literature on internet technology usage, we anticipate observing gender-based disparities in ChatGPT adoption and usage. While each of the proposed underlying factors has the potential to explain the existence of a gender gap, we will not pre-specify which factor(s) will emerge as the primary driver(s) of this phenomenon. Our analysis aims to contribute with valuable insights to the ongoing discussions on gender, AI technology, and the digital divide.

Appendix

The survey questions and vignettes are attached for review.

Gender and AI adoption - Study 1: Students (#150413)

Created: 11/09/2023 05:57 AM (PT)

1) Have any data been collected for this study already?

It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?

This first study is part of a bigger project "Will Artificial Intelligence get in the way of gender equality?". The first study seeks two purposes: (i) identify whether a gender gap exists in the adoption of ChatGPT and (ii) explore the underlying mechanisms driving any observed disparities. First, we collect a series of measures of ChatGPT usage to assess whether there is a gender gap in AI use, overall and on different margins of interest. In the event of a detected gap, the study will proceed to identify its underlying causes. For this purpose, we have pinpointed three primary factors influencing ChatGPT usage: preferences, perceptions, and exposure/experience.

Drawing on insights from a pilot survey in Prolific and existing literature on internet technology usage, we anticipate observing gender-based disparities in ChatGPT adoption and usage. While each of the proposed underlying factors has the potential to explain the existence of a gender gap, we will not pre-specify which factor(s) will emerge as the primary driver(s) of this phenomenon.

3) Describe the key dependent variable(s) specifying how they will be measured.

1. Willingness to use ChatGPT: participants will be presented two hypothetical scenarios. In each the main outcome is to indicate how likely are you to use ChatGPT throughout this course, in a scale from 1: Very Unlikely to 5: Very Likely. One treatment has an additional questions that corresponds to indicate how likely are you to use ChatGPT during the exam, in a scale from 1: Very Unlikely to 5: Very Likely.
2. Usage of ChatGPT: we obtain a measure of how often they use ChatGPT and for which tasks.

4) How many and which conditions will participants be assigned to?

This is a vignette study where participants will face a scenario where they participate in a course. There are two levels of randomization:
1. Whether the final assessment is an in-person exam or a home exam. This randomization is between subject.
2. Whether the professor allows or forbids the use of ChatGPT. This dimension is studied within subject, with randomization on the order of the scenarios.

In total, each participant faces two scenarios. The randomization is done by the gender indicated by the participant.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

Linear regression over the Willigness to use ChatGPT across treatments. We will also pool the Likely and Unlikely levels of the scale, and perform analysis over two dummies, one with value 1 if answer is Likely and another with value 1 if answer is Unlikely.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We will not analyze data of participants who started the survey outside of 20 minutes from the time at which the survey was implemented in classroom.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

The study is made in three lectures of three different courses at NHH. A master's course: with around 30 participants, and two bachelor's courses, where we expect around 300 participants in each course to attend the lecture and participate.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

As we were not able to edit our preregistration at AEA RCT Registry (November 6, 2023), which did not include the between-subject randomization, we make this pre-registration to indicate this addition. Relative to the test data we collected on November 7, 2023 (N=28), we will delete the option Neutral/Unsure from the trust question. The data from the bachelor's courses has not been collected.

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Gender and AI adoption - Study 1: Master Students (#157598)

Created: 01/11/2024 03:47 AM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
We continue collecting data for the project "Will Artificial Intelligence get in the way of gender equality?". The first study seeks two purposes: (i) identify whether a gender gap exists in the adoption of ChatGPT and (ii) explore the underlying mechanisms driving any observed disparities. First, we collect a series of measures of ChatGPT usage to assess whether there is a gender gap in AI use, overall and on different margins of interest. In the event of a detected gap, the study will proceed to identify its underlying causes. For this purpose, we have pinpointed three primary factors influencing ChatGPT usage: preferences, perceptions, and exposure/experience.

Drawing on insights from a pilot survey in Prolific and existing literature on internet technology usage, we anticipate observing gender-based disparities in ChatGPT adoption and usage. While each of the proposed underlying factors has the potential to explain the existence of a gender gap, we will not pre-specify which factor(s) will emerge as the primary driver(s) of this phenomenon.

3) Describe the key dependent variable(s) specifying how they will be measured.
1. Willingness to use ChatGPT: participants will be presented two hypothetical scenarios. In each the main outcome is to indicate how likely are you to use ChatGPT throughout this course, in a scale from 1: Very Unlikely to 5: Very Likely. One treatment has an additional questions that corresponds to indicate how likely are you to use ChatGPT during the exam, in a scale from 1: Very Unlikely to 5: Very Likely.
2. Usage of ChatGPT: we obtain a measure of how often they use ChatGPT and for which tasks.

4) How many and which conditions will participants be assigned to?
This is a vignette study where participants will face a scenario where they participate in a course. There are two levels of randomization:
1. Whether the final assessment is an in-person exam or a home exam. This randomization is between subject.
2. Whether the professor allows or forbids the use of ChatGPT. This dimension is studied within subject, with randomization on the order of the scenarios.

In total, each participant faces two scenarios. The randomization is done by the gender indicated by the participant.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Linear regression over the Willigness to use ChatGPT across treatments. We will also pool the Likely and Unlikely levels of the scale, and perform analysis over two dummies, one with value 1 if answer is Likely and another with value 1 if answer is Unlikely.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will not analyze data of participants who started the survey outside of 20 minutes from the time at which the survey was implemented in classroom and participants who answered the exact same survey in the past.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
The study is made during lecture times of a master course, where we expect around 100 participant course to attend the lecture and participate.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
As we were not able to edit our preregistration at AEA RCT Registry (November 6, 2023), we make this pre-registration to indicate the additional data collection.

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Gender and AI adoption - Study 1: Students Robustness (#169979)

Created: 04/09/2024 10:59 AM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
We continue collecting data for the project "Will Artificial Intelligence get in the way of gender equality?". The first study seeks two purposes: (i) identify whether a gender gap exists in the adoption of ChatGPT and (ii) explore the underlying mechanisms driving any observed disparities. First, we collect a series of measures of ChatGPT usage to assess whether there is a gender gap in AI use, overall and on different margins of interest. In the event of a detected gap, the study will proceed to identify its underlying causes. For this purpose, we have pinpointed three primary factors influencing ChatGPT usage: preferences, perceptions, and exposure/experience.
As with the previous data collections, we anticipate observing gender-based disparities in ChatGPT adoption and usage. This time, we incorporate a series of questions for the following purposes:
1. Use a more objective measure of use of ChatGPT.
2. Evaluate whether men and women identify the image in the prompt elicitation task as an optical illusion.
3. Evaluate expectations of students about how ChatGPT will be beneficial for the labor market.

3) Describe the key dependent variable(s) specifying how they will be measured.
1. Willingness to use ChatGPT: participants will be presented a hypothetical scenario, where the main outcome is to indicate how likely are you to use ChatGPT throughout this course, in a scale from 1: Very Unlikely to 5: Very Likely. One treatment has an additional questions that corresponds to indicate how likely are you to use ChatGPT during the exam, in a scale from 1: Very Unlikely to 5: Very Likely.
2. Usage of ChatGPT: we obtain a measure of how often they use ChatGPT and for which tasks.

4) How many and which conditions will participants be assigned to?
This is a vignette study where participants will face a scenario where they participate in a course. There is one level of randomization:
1. Whether the professor allows or forbids the use of ChatGPT. This randomization is between subject.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Linear regression over the Willigness to use ChatGPT across treatments. We will also pool the Likely and Unlikely levels of the scale, and perform analysis over two dummies, one with value 1 if answer is Likely and another with value 1 if answer is Unlikely.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
This data has two purposes:
1. To evaluate the robustness of measurements and to collect new variables.
2. To increase the sample size on the main analysis, adding to observations from previous collections.

In purpose 1, we will only exclude participants who started the survey outside of 20 minutes from the time at which the survey was implemented in classroom.
In purpose 2, we will exclude participants who started the survey outside of 20 minutes from the time at which the survey was implemented in classroom and who answered a similar survey in November 2023.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
The study is made during lecture in one bachelor's course, where we expect around 300 participants to attend the lecture and participate.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Nothing else to pre-register.

Pre-Analysis-Plan: Managers survey

Project: Will Artificial Intelligence Get In The Way of Achieving Gender Equality?

Daniel Carvajal, Catalina Franco, Siri Isaksson

Date: 10.06.2024

Project Summary

This research project aims to investigate the existence of gender differences in the adoption and use of generative AI technologies, specifically ChatGPT. This project seeks four purposes: (i) identify whether a gender gap exists in the adoption of ChatGPT, (ii) explore the underlying mechanisms driving any observed disparities, (iii) evaluate the efficacy of interventions aimed at reducing the gender gap in AI adoption, and (iv) assess the impact of this gap on labor market outcomes. Previous data collections for this project have assessed purposes (i)-(iii), using a sample of university students from NHH Norwegian School of Economics. We find that there is a significant gender gap in use and skills of ChatGPT in students. Moreover, we observe that explicitly allowing the use of ChatGPT in the classroom closes the gap, while on the other hand explicitly forbidding the use of ChatGPT increases the gap.

The current study targets objective (iv) by surveying managers in industries that employ graduates from NHH. First, we use a conjoint-type study to examine whether managers value job candidates that showcase generative AI expertise in hiring decisions, currently and prospectively. Second, we use a hypothetical vignette experiment to assess whether improvements in productivity due to the use of generative AI are rewarded in the workplace. Finally, we collect a series of measures of perceptions and attitudes towards the use of ChatGPT from employers.

Our goal is to document whether the gender disparities in ChatGPT usage that we document with the student sample can contribute to a gender gap in labor market success when students graduate and transition to the labor market.

Participants

The study consists of a survey experiment involving around 1000 individuals employed in companies who hold managerial positions and work in one of the top industries where NHH graduates work: Consulting, Finance, Administration, Accounting. It will be run in June 2024. The survey will be administered through Qualtrics, and to the panel of respondents available to the survey provider company Norstat. All survey answers are anonymous, and the data collected will be used exclusively by the research team.

Survey Design

The survey is structured to take around 7 minutes and includes a conjoint study, a hypothetical vignette, and survey questions on perceptions and attitudes towards the use of generative AI in the workplace. From now on, we will refer to each participant in the study as a manager.

Main outcomes

Value of generative AI in hiring (conjoint study):

Y1. Score given by managers to a hypothetical **current** candidate represented by a profile card. Each manager gives scores to two randomly selected profiles.

Y2. Score given by managers to a hypothetical **prospective** candidate in three years, represented by a profile card. Each manager gives scores to one selected profile.

Value of generative AI in workplace (vignette study)

Y3. Each manager is presented with a hypothetical situation in which two employees worked on a task and their performance on that task defines whether they go into a "promotion track" or not. Workers were allowed to use generative AI and one worker finished the task 25% faster than the other worker. The outcome is a binary variable that takes value 1 if the worker that finished the task faster is selected for the promotion track.

Secondary outcomes

Value of generative AI in hiring (conjoint study):

- Managers are also asked to select between the two current candidates presented to them to be called for an interview.
- Moreover, for the individual selected for the interview, managers are asked how much the participant can negotiate the starting salary if given the position.

ChatGPT Usage and Workplace

- Usage of ChatGPT
- Policies in place in companies
- Influence of ChatGPT in grades
- Value in hiring
- Value in the workplace

Attitudes towards ChatGPT usage

- Advantages
- Disadvantages
- Net benefit

Expectations about the future value ChatGPT (three years)

- Valued in hiring
- Valued in the workplace

Conjoint study

Each manager is presented with two profiles, randomly selected, where the manager must give a score to each candidate, and then select which one will be called for an interview, as well as how much will the candidate be able to negotiate the starting salary. The profiles vary in several dimensions:

- ChatGPT expertise
- Grade and grade distribution
- Gender

The two selected profiles come from a set of 10 hypothetical profiles that represent the following 5 types of workers:

- WHC: Woman with High grades and ChatGPT expertise
- WHN: Woman with High grades and No ChatGPT expertise
- MHC: Man with High grades and ChatGPT expertise
- MHN: Man with High grades and No ChatGPT expertise
- MLC: Man with Low grades and ChatGPT expertise

Vignette experiment

Each manager is presented with a situation in which two employees must work in the same task, and their performance in the task defines whether they are selected for the promotion track at the company. The managers are explicitly told that the use of generative AI is allowed. Both workers have the same output quality, but one worker does the job 25% faster than the other one. Managers are randomly assigned to one out of two main conditions:

- Explicit: managers are given the number of hours that each worker took to finish the task and explicitly told that the faster used ChatGPT and the slower did not.
- Not explicit: managers are given the number of hours that each worker took to finish the task, but it is not explicitly mentioned who used ChatGPT.

The gender of the employees is also randomized, which will be exploited for heterogeneity analysis of the main treatment effect.

Main analyses:

Value of generative AI in hiring (conjoint study)

Current scenario (Y1)

$$y_1 = \beta_{10} + \beta_{11}WHC + \beta_{12}MHN + \beta_{13}MHC + \beta_{14}MLC + X\gamma_1 + \varepsilon$$

Where y_1 is the score given to the candidate, and WHC, MHN, MHC, MLC are dummy variables that take value 1 if the candidate is of each of the respective types, and 0 otherwise. Note that the baseline is type $WHN.$ The comparisons of interest in our analysis correspond to:

1.1. WHC-WHN: given by β_{11}
1.2. MHC-MHN: given *by* $\beta_{13} - \beta_{12}$
1.3. MLC-WHN: *given by* β_{14}

Future scenario (Y2)

$$y_2 = \beta_{20} + \beta_{21}MLC + X\gamma_2 + \varepsilon$$

Where y_2 is the score given to the candidate, and MLC is a dummy variable that takes value 1 if the candidate is of the respective type, and 0 otherwise. Note that the baseline is type $WHN.$ The comparisons of interest in our analysis correspond to:

2.1. MLC-WHN: given by β_{21}

Value of generative AI in workplace (vignette study)

Promotion track to fastest worker (Y3)

$$y_3 = \beta_{30} + \beta_{31}Explicit + X\gamma_3 + \varepsilon$$

$$y_3 = \beta'_{30} + \beta'_{31}Explicit + \ \beta'_{32}Female + \ \beta'_{33}Explicit \times Female + X\gamma'_3 + \varepsilon$$

Where y_3 is a dummy variable that takes value 1 if the fastest worker was selected for the promotion track and 0 otherwise, $Explicit$ is a dummy variable that take value 1 if the manager was explicitly informed about who used ChatGPT and 0 otherwise. We intend to first estimate the effect of making the use of ChatGPT explicit on y_3, given by coefficient β_{31} (first equation). We also want to study heterogeneity of this effect when the fastest worker using ChatGPT is a man or a woman, given by the estimated coefficient of the interaction term β'_{33} (second equation).

Hypotheses

Drawing on insights from the growing recent literature on the effects of access to generative AI on productivity across fields (Noy and Zhang, 2023; Brynjolfsson et al., 2023; Peng et al., 2023), as well as recent survey data from employees in Amazon (2024), we expect managers to value the use and signaling of generative AI skills in workers. Thus, we expect positive estimates for comparisons 1.1 and 1.2.

We intend to study whether knowledge of ChatGPT makes a low-grade male candidate comparable to a high-grade female candidate without knowledge of ChatGPT. Therefore, we intend to test whether the estimate for 1.3 and 2.1 is zero.

Finally, we do not specify a direction of the hypothesis in whether making explicit the use of ChatGPT affects positively, negatively or has no impact on the decision to choose the fastest candidate. However, we prespecify our interest in looking for heterogeneous effects according to the gender of the fastest candidate. For example, if the use of ChatGPT is perceived negatively, previous work suggests there might differences in retaliation on the use of ChatGPT by gender of the user (Dehdari et al, 2019).

References

Amazon (2024). Accelerating AI skills: Preparing the workforce for jobs of the future. AWS Reports.

Bimber, B. (2000). Measuring the gender gap on the Internet. *Social science quarterly*, 868-876.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, *129*(3), 1409-1447.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, *109*(3), 739-773.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work (No. w31161). National Bureau of Economic Research.

Dehdari, S. H., Heikensten, E., & Isaksson, S. (2019). What goes around (sometimes) comes around: gender differences in retaliation. Available at SSRN 3378279.

Doshi, A. R., & Hauser, O. (2023). Generative artificial intelligence enhances creativity. *Available at SSRN 4535536.*

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*, 187-192.

OECD (2018). Bridging the digital gender divide: include, upskill, innovate. *Available at: www.oecd.org/digital/bridging-the-digital-gender-divide.pdf*

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of ai on developer productivity: Evidence from github copilot. arXiv preprint arXiv:2302.06590.

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Managers Survey May-June 2024 (#176722)

Created: 05/27/2024 06:01 AM (PT)

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

We continue collecting data for the project "Will Artificial Intelligence get in the way of gender equality?". This study examines the following hypothesis: whether generative AI skills (e.g., ChatGPT) are valued in the labor market, with the aim of examining the labor market consequences of a gender gap in use of generative AI. We conduct a survey on managers in the industries that hire students from our earlier sample (NHH students) and assess whether they value: (i) applicants to jobs showcasing generative AI skills, (ii) workers with increased productivity using generative AI skills. We use the following survey methods to study this question:

1. Conjoint study with managers evaluating hypothetical applicant profiles in the present and their expected answers in three years.
2. Vignette experiment where managers make a career advancing decision between two candidates.
3. Survey questions on usage and perceptions over the use of ChatGPT in the workplace.

3) Describe the key dependent variable(s) specifying how they will be measured.

We have three main dependent variables.

Y_1: Score given by managers to a current hypothetical candidate represented by a profile card (conjoint study). This is collected by presenting two randomly selected profiles to the manager and asking to give a score to each. 5 types of profiles are being evaluated: (1) Woman - High grades - ChatGPT expertise (WHC), (2) Woman - High grades - No ChatGPT expertise (WHN), (3) Man - High grades - ChatGPT expertise (MHC), (4) Man - High grades - No ChatGPT expertise (MHN), (5) Man - Average grades - ChatGPT expertise (MLC).

Y_2: Score given by managers to a future hypothetical candidate, applying in three years, represented by a profile card. This is collected by presenting to the manager one out of two possible profiles, and asking to give a score (possible profiles: WHN-MLC).

Y_3: Binary variable with value 1 if a manager selected for a "promotion track" the fastest candidate out of two to finish a task (hypothetical scenario).

4) How many and which conditions will participants be assigned to?

For dependent variables:

Y_1: participants will be presented two profiles randomly assigned from a set of 10 possible profiles, representing the five types.

Y_2: participants will be presented one out of two possible types of profiles: WHN and MLC.

Y_3: there are 2 main conditions, which are (1) the managers are explicitly told which candidate used ChatGPT for performing the task (2) the managers are not told who used ChatGPT. A second layer of randomization corresponds to variations in the gender of the two possible candidates.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

Y_1: using linear regression, we will estimate the differences between average scores across the following comparison of types: (1) WHC-WHN, (2) MHC-MHN, (3) WHN-MLC.

Y_2: using linear regression, we will estimate the differences between average scores across the types: WHN and MLC.

Y_3: using linear regression we estimate differences between selecting the fastest candidate across the two main specifications: whether the use of ChatGPT was explicit or not (Explicit). We are also interested in the interaction term of Explicit and Female, which is a binary variable that takes value 1 if the fastest candidate was a woman.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Any incomplete survey or surveys answered in less than 30 seconds will be discarded.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

Survey will be implemented until 1000 managers have completed the survey.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

As secondary outcomes in the manager's hiring tasks (Y_1 and Y_2) we also obtain decisions of whether a candidate is called for an interview or not, and whether they are able to negotiate salary. In addition, we collect answers for a series of questions over perceptions and attitudes towards generative AI and other questions capturing the value of generative AI in the workplace.

F Questionnaires

F.1 Questionnaire for student survey

Page 1. Consent

NHH

Welcome to this research project!

We very much appreciate your participation in this 5-minute survey. All data obtained is anonymous. Please make sure to always read the instructions carefully, **answer truthfully**, and do not leave the survey until reaching the end. Participation in this research study is completely voluntary. If you have questions regarding this study, you may contact: thechoicelab@nhh.no

Please click **Accept** below if you have understood the above and wish to participate in this study.

Accept

Page 2. Background characteristics

Are you from Norway?

Yes

No

To which gender identity do you most identify:

Male

Female

Non-binary / third gender

Prefer not to say

How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?

Completely unwilling to do so										Very willing to do so
0	1	2	3	4	5	6	7	8	9	10

In general, how willing are you to take risks?

Completely unwilling to take risks										Very willing to take risks
0	1	2	3	4	5	6	7	8	9	10

Page 3. "Allows" treatment

Imagine you are enrolled in a course on Environmental Policy and Economic Impact. This course explores the intersection of environmental regulations, economic incentives, and their effects on industry practices and sustainability. The professor explicitly allows the use of ChatGPT during coursework. It is an 8-week course with final evaluation given by a final home exam.

Given this scenario, how likely are you to use ChatGPT throughout the course?

- Very unlikely
- Somewhat unlikely
- Neither likely nor unlikely
- Somewhat likely
- Very likely

Given the scenario, how likely are you to use ChatGPT during the final exam?

- Very unlikely
- Somewhat unlikely
- Neither likely nor unlikely
- Somewhat likely
- Very likely

Page 4. “Forbids” treatment

Imagine you are enrolled in a course on Climate Change Economics. This course delves into the economic analysis of climate change, including the evaluation of mitigation strategies, adaptation costs, and international climate policy agreements. The professor explicitly forbids the use of ChatGPT during coursework. It is an 9-week course with final evaluation given by a final home exam.

Given this scenario, how likely are you to use ChatGPT throughout the course?

- Very unlikely
- Somewhat unlikely
- Neither likely nor unlikely
- Somewhat likely
- Very likely

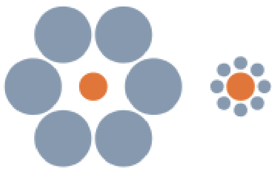
Given the scenario, how likely are you to use ChatGPT during the final exam?

- Very unlikely
- Somewhat unlikely
- Neither likely nor unlikely
- Somewhat likely
- Very likely

Page 5. Prompting skills task

Do you know how to use ChatGPT?

Please take a moment to carefully check the image presented below.



Using the space provided, please write down the question that **you would ask to ChatGPT** to learn about the official name of this visual phenomenon. Remember ChatGPT cannot observe the image.

Page 6. Confidence question

How confident do you feel that the query you just provided will make ChatGPT get the information you need?

Not confident at all

Slightly confident

Very confident

Extremely confident

Page 7. ChatGPT use

How familiar are you with ChatGPT?

- I have not heard of it.
- I have heard of it but have not used it myself.
- I used it a few times.
- I use it occasionally.
- I use it regularly.

Page 8. Exposure and typical tasks

A survey conducted among university students in the US in the Spring of 2023 reports that 30% of students use ChatGPT for their schoolwork.

Now, for each of the groups below, please indicate the percentage of people you believe use ChatGPT:

0 20 40 60 80 100

Your group of friends

Students in this course

Professors at NHH

What type of tasks do you typically ask ChatGPT to help with? (Please select up to the most common three)

- Coding tasks
- Writing tasks
- Retrieving information
- Solving Math questions
- Other (Please specify)
- I don't use it

Page 9. Frequency by task

How frequently do you use ChatGPT for the following purposes:

Preparing for exams in a course:

- Never
- Occasionally
- Regularly

Solving home assignments for a course:

- Never
- Occasionally
- Regularly

Tasks unrelated to coursework:

- Never
- Occasionally
- Regularly

Tasks related to coursework:

- Never
- Occasionally
- Regularly

Page 10. Advantages (Usefulness)

What do you believe are the main advantages of using ChatGPT in coursework? (Please select all that apply.)

- Saves time.
- Increases accuracy or work quality.
- I do not see any advantages.
- Improves learning of course methods.
- Improves my grade in the course.
- Other (Please Specify)

Page 11.1 Agree/Disagree

How much do you agree with the following statements?

I think ChatGPT is enjoyable to use:

- Completely agree
- Somewhat agree
- Neither agree not disagree
- Somewhat disagree
- Completely disagree

Using ChatGPT as an aid to solve assignments in a course is equivalent to cheating:

- Completely agree
- Somewhat agree
- Neither agree not disagree
- Somewhat disagree
- Completely disagree

Page 11.2 Agree/Disagree

Using ChatGPT as a learning aid in a course is equivalent to cheating:

- Completely agree
- Somewhat agree
- Neither agree not disagree
- Somewhat disagree
- Completely disagree

I think ChatGPT is difficult to use:

- Completely agree
- Somewhat agree
- Neither agree not disagree
- Somewhat disagree
- Completely disagree

Page 11.3 Agree/Disagree

It is easy for professors to identify if a student has used ChatGPT:

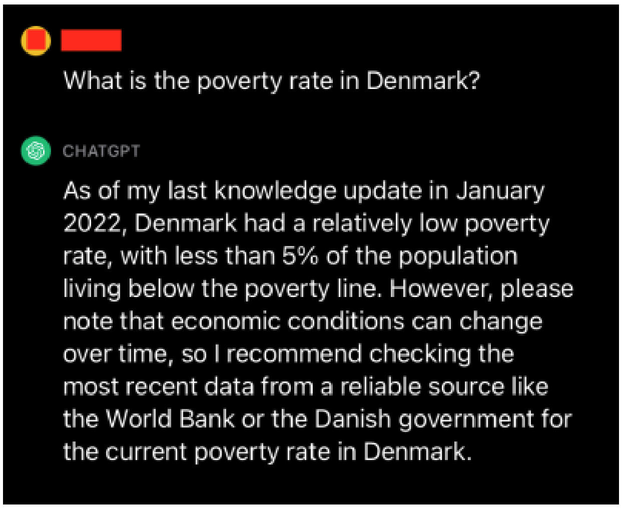
- Completely agree
- Somewhat agree
- Neither agree not disagree
- Somewhat disagree
- Completely disagree

ChatGPT is mostly a tool complementing skills rather than substituting effort:

- Completely agree
- Somewhat agree
- Neither agree not disagree
- Somewhat disagree
- Completely disagree

Page 12. Trust accuracy

Below is a screen capture of a query made to ChatGPT, along with the response it provided.



Based on this response from ChatGPT, how much do you trust that the **poverty rate reported** is accurate?

- Completely trust
- Somewhat trust
- Somewhat distrust
- Completely distrust

Page 13. Persistence and inaccuracy

If ChatGPT does not provide the desired answer on your first attempt, how many additional attempts do you typically make?

- None, I move on.
- One more try.
- Two more tries.
- I keep trying until satisfied.
- I don’t use it.

Have you ever received inaccurate or misleading information from ChatGPT?

- Yes, many times.
- Yes, few times.
- No, never.
- I don't use it.

Page 14. Subscription and admission grade

Do you have a subscription for using ChatGPT or other similar AI platforms?

- No.
- Yes, I have the free subscription.
- Yes, I have the paid subscription.

What was your admission grade at NHH? Please provide an estimate if you don't remember the exact grade (or NA if you don't have):

F.2 Questionnaire for manager survey

(Consent)

Q1 Welcome to this research project!
We appreciate your participation in this 7-minute survey. Participation in this research study is voluntary. All data collected is anonymous. Please read the instructions carefully, answer the questions honestly and do not end the survey until you have answered all the questions. If you have any questions regarding this study, you can contact: thechoicelab@nhh.no.
Click on ”Start the survey” if you have understood the text above and wish to participate in this study.

(Screening)

Q3 How many employees do you have direct personnel responsibility for? [Options: None (0) , 1-5 employees, 6-10 employees, 11-20 employees, Over 20 employees]

Display This Question: If How many employees do you have direct personnel responsibility for? != None (0)

Q4 How long have you held a position with personnel responsibility? [Options: Less than a year, 1-2 years, 3-5 years, More than 5 years]

Display This Question: If How many employees do you have direct personnel responsibility for? = None (0)

Q5 Do you have a position that gives you influence over decisions about employment and promotions, or that involves you assigning tasks, giving guidance, evaluating or giving feedback to other employees? [Options: Yes, No]

Display This Question: If Do you have a position that gives you influence on decisions about employment and promotions, or... = Yes Or How many employees do you have direct personnel responsibility for? != None (0)

Q6 Does your company/your department employ newly graduated candidates with a master's degree in economics and administration (for example, candidates with a master's degree from NHH or BI)? [Options: Yes, No, I don't know]

Display This Question: If Does your company/your department employ newly graduated candidates with a master's degree in economics and administration... = Yes

Q7 Are you regularly involved in these hiring processes? [Options: Yes, No , Prefer not to answer]

(Job tasks – Not used in this study)

Q8 The list below indicates a number of tasks that may be relevant for a newly graduated candidate with a master's in economics and administration. Please mark all tasks that may be relevant for such a newly hired candidate in your company/department: [Options: Project work (1), Advice/guidance (2), Finance/accounting (3), Administration/personnel tasks (4), Customer care/front line (5), Management (6), Case management (7), Sales/marketing/advertising (8), Teaching/training/pedagogical work (9), Research (10), Technical development/project planning (11), Information/communication/journalism (12)]

Carry Forward Selected Choices from "The list below indicates a number of tasks that may be relevant for a newly graduated candidate with a master's in economics and administration. Please mark all the tasks that may be relevant for such a newly employed candidate in your company/department:"

Q10 The list below indicates the tasks you marked as relevant for a newly graduated candidate with a master's in economics and administration in your company/department. Please rank the tasks according to how important they are for promotion in your company/department. Mark the task most important for promotion as 1, and tasks less important for promotion with progressively higher numbers.

Carry Forward All Choices - Displayed & Hidden from "The list below indicates the tasks you marked as relevant for a newly graduated candidate with a master's in economics and administration in your company/department. Please rank the tasks according to how important they are for promotion in your company/ department. Mark the task most important for promotion as 1, and tasks less important for promotion with progressively higher numbers."

Q12 Now think of a task that is not on the list from the previous question, "Practical administrative tasks", e.g. finding meeting times that suit everyone, writing meeting minutes, summaries or notes from meetings, planning social work events, etc.
Think through how important practical administrative tasks are for promotion in your company/department compared to the tasks you ranked in the previous question. When "Practical Administrative Tasks" is added, there will be N tasks in the list, as shown below.
Please enter a number between 1 and N to rate "Practical administrative tasks", where 1 indicates it is most important and N that it is the least important for promotion in your company/department. Practical administrative tasks [Option: Fill number]

(Conjoint experiment)

Q14 We would now like to show you two profiles of newly graduated candidates with an economic-administrative master's degree, who are applying for a job at your company/department.
Please give each candidate a score between 0 and 12 based on how well qualified you think they are for a typical job for recent graduates in your department/company.

Q15 [PROFILE, one out of Options 1, 2, 3, 4 or 5]
Give this candidate a score between 0 and 12. 0 means average candidate (Avg.). 12 means outstanding candidate (Forward): [Option: slide]

Q16 [PROFILE, one out of Options 6, 7, 8, 9, 10]
Give this candidate a score between 0 and 12. 0 means average candidate (Avg.). 12 means outstanding candidate (Forward): [Option: slide]

Q17 Choose one of the two candidates you would invite to an interview in your company? (Click on the profile) [Options: each of the candidates presented]

Q18 Imagine that the selected candidate is offered the position and receives an offer of starting salary. The candidate can negotiate the starting salary. What do you think is the maximum starting salary this candidate will be able to get in this job? [Options: 0-5% above the offered starting salary (1), 5-10% above the offered starting salary (2), 10-15% above the offered starting salary (3), 15% or more above the offered starting salary (4)]

(Vignette experiment)

Q20 Consider the following hypothetical situation. [Only two randomly selected names are presented]
Daniel/Ida and Martin/Emma started working at a company at the same time in the same type of job a few years ago. They are assigned a task that they must solve individually. They can use all appropriate resources, including generative AI. Their performance on this task will determine which of the two will be placed on the 'career development track' in the company.
[One of the two following paragraphs presented: Both Daniel and Martin complete the task with the same level of quality. Daniel took 8 days to complete it without generative AI. Mar-

tin used generative AI and completed it in 6 days. / Both Daniel and Martin complete the task with the same level of quality. Daniel took 8 days to complete it. Martin completed it in 6 days.]

Q21 Who do you think should be placed on a career development track? [Options: Daniel/Ida (1), Martin/Emma (2)]

(Questions about use and attitudes to generative AI use)

Q23 The next questions will deal with tools based on generative artificial intelligence (hereafter: generative AI tools). When replying, consider ChatGPT or other similar tools (eg Claude, BingChat, etc.).

Q24 Are you familiar with generative AI tools? [Options: I haven't heard of them (1), I've heard of them but haven't used them myself (2), I have used them a few times (3), I use them occasionally (4), I use them regularly (5)]

Q25 In which areas do you think generative AI tools can increase productivity in your company? (Select all that apply) [Options: Automate repeating/repetitive tasks (1), Improve communication (2), Improve workflow and results (for example, data analysis) (3), Support for learning new skills (4), Increased innovation and creativity (5), Save time (6); It does not improve productivity (7), Other (please write): (8)]

Q26 What do you think are the biggest challenges regarding the use of generative AI tools for your company? (Select all that apply) [Options: Confidentiality and data protection (1), Risk of plagiarism (2), Prevents the learning of new skills (3), Risk of inaccurate information (4), Discrimination from AI technologies (5), Lower effort from employees (6), I see no challenges for our company (7), Other (please write): (8)]

Q27 What is your company's attitude towards the use of generative AI tools at work? [Options: It is allowed and encouraged (1), It is allowed but not actively encouraged (2), It is neither explicitly allowed/encouraged nor prohibited/advised (3), It is forbidden (4), I don't know (5)]

Q29 State how much you agree with the following statements: [Options: Strongly agree, Somewhat agree, Neither agree nor disagree, Somewhat disagree]

Q30 For my company, I think the advantages of generative AI outweigh the disadvantages

Q31 I would prefer to hire a graduate with generative AI skills rather than a similar candidate without generative AI skills

Q32 Having generative AI skills can help a graduate earn a higher salary in their first job

Q33 If a student achieves higher grades by using generative AI, it is because the AI tools effectively improve learning, rather than replace individual effort

Q35 For the following questions, try to imagine what the situation will be like in your company in the years to come (three years from now). Please give a score between 0 and 12 to the candidate below based on how well qualified you think she/he will be for a typical job for recent graduates in your department/company in three years.

Q36 [PROFILE, either a Top Woman No AI or a Low Man AI]
Give this candidate a score between 0 and 12. 0 means average candidate (Avg.). 12 means outstanding candidate (Forward): [Option: slide]

Q37 Imagine that this candidate is offered the position and receives an offer of starting salary. The candidate can negotiate the starting salary. What do you think is the maximum starting