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

Daniel Carvajal† Catalina Franco‡ Siri Isaksson§

August, 2024

Abstract

We conduct two survey experiments to examine gender differences in generative AI adoption and potential labor market consequences. First, we document a substantial gender gap 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 acquiring AI skills significantly enhances job prospects for top female students currently opting out. Finally, we provide causal evidence on policy tools to close this gap. Our findings show generative AI could widen existing gender gaps in the labor market, but appropriate encouragement and policies can prevent this outcome.

JEL Codes: I24; J16; J24; O33
Keywords: Generative AI, ChatGPT, gender, labor market, technology adoption

*We would like to thank Andrea Bocchino, Francesco Capozza, Finn Casey, Nisvan Erkal, Iver Finne, Dorothea Kübler, Danielle Li, Akhsay Moorthy, María Recalde, Erika Povea, Frode Skjeret, 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, MCV Workshop Madrid, GATE lab at Lyon University 2, NTU-GATE Workshop, ASFEE 2024, PET Lyon 2024, and the 2024 Asia Pacific ESA provided useful comments and feedback. This study was approved by the NNH IRB (NHH-IRB 42/22) and pre-registered in the AEA RCT registry (AEARCTR-0012452).

†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), spanning domains such as professional writing (Noy and Zhang, 2023), customer support tasks (Brynjolfsson et al., 2023), 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. However, only those who participate in this technological revolution may reap its benefits. This raises a key question: will existing inequalities in the labor market be further exacerbated if there are systematic disparities in generative AI adoption across different demographic groups?

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 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 study focuses on those who will soon be facing this rapidly evolving labor market—college students—and their potential future employers. We conduct two preregistered survey experiments in Norway to investigate gender differences in generative AI adoption and employer valuation of generative AI skills. We document significant gender differences in AI adoption alongside higher employer valuation of job candidates with AI skills. While the gender disparities are concerning, our findings also convey optimism. With the right encouragement and policies, female students can embrace AI, thereby mitigating potential future labor market inequalities.

[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).

We develop a comprehensive study targeting both the supply and demand sides of the labor market. On the supply side, we survey 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 assess 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. We then turn to the demand side by surveying managers to examine gender gaps in candidates' ratings based on their AI expertise: can female students enhance their job prospects by acquiring generative AI skills? We run a survey experiment on 1,134 managers in Norway who typically employ graduates from this business school and ask them to rate profiles of hypothetical candidates with and without generative AI skills. The managers are also asked to decide on the promotion of hypothetical workers who increase their productivity with or without the help of AI. The design of the manager survey experiment allows us to isolate the effect of acquiring AI skills for a male vs. female job candidate and of using AI in the workplace.

We make three important contributions to the literature. First, we provide the initial study focusing on understanding gender gaps in generative AI. We find a substantial and significant gender gap in AI adoption: female students are 25% less likely to report a high use of ChatGPT or similar AI tools. 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 to have a paid subscription.

Other studies have documented similar gender gaps in generative AI adoption. Our gap closely aligns with the one found by Humlum and Vestergaard (2024) in a large sample of Danish workers. This validates our findings and shows that the gap is present in different populations. While the main focus of these papers has been other topics such as AI aversion (Haslberger et al., 2024) or adoption among workers (Humlum and Vestergaard, 2024), our study is designed with the sole purpose of understanding gender gaps in AI allowing us to provide crucial insight into *why* the gender gap in AI emerges, and *how* to close it.

To gain a deeper understanding of the *why*, we ask a rich set of questions regarding students' preferences, perceptions and exposure to this new technology. For instance, we measure the extent to which they consider using AI as cheating and their persistence in using it, specifically whether they continue to use AI if it does not provide the desired answer on the first try. These questions yield novel insights crucial to understanding the gender gap in AI adoption. For example, male students are more likely than female students to disagree with statements that using AI as a learning aid (88%) and for course assignments (64%) constitutes cheating. Turning to our persistence measure, 71% of male students state that they keep trying when not obtaining the desired answer, as opposed to only 55% of female students.[2] These differences in perceptions on cheating and persistence 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.

Second, we make a policy contribution on *how* to close the gap: 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. They intend to use AI when it is allowed and refrain from using it when it is banned. 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. 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 are crucial insights given that universities and workplaces are currently in the process of formulating their rules and policies around AI use.

Third, we provide the first evidence that generative AI skills are valued in the labor mar-

[2]This persistence gap relates to previous work documenting gender differences in persistence in educational settings (Landaud and Maurin, 2020; Franco and Hawkins, 2023).

ket, in particular for high-skill female job candidates.[3] We find that female candidates with top grades who possess 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. Additionally, a hypothetical vignette experiment shows that a majority of managers (65%) would promote a worker who boosts their productivity using AI. This part of the study provides critical insights into how AI skills are rewarded in the labor market. The reward in the hiring process differs by gender and academic skill. Our findings imply that women, in particular, could significantly enhance their job prospects by acquiring generative AI skills. Furthermore, our results show that a gender gap in adoption on the job would translate into gender differences in promotion outcomes.

Throughout our analysis, we maintain a special focus on top female students, those who are at the top of the admission grade distribution.[4] Previous research has shown that gender differences are particularly pronounced 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 (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.

The gender gap in adoption of generative AI is particularly pronounced at the top: while female students with lower admission grades are just as likely as similar male students to use AI, top women are opting out and using it only about half as much. In the bottom two quintiles of the admission grade distribution, 88% of female students report high use, similar to 79-82% of the male students. In the top three quintiles, however, only between 44-53% of

[3]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 distinct from the generative AI focus discussed in our paper.

[4]Admissions to university programs in Norway rely on an admission grade based on high school grades and high school exit exams. Students self-reported this grade and 55% of our sample reported a valid grade.

female students report high use, in comparison to 75-88% of male students. Put differently, male students use AI frequently regardless of their academic skill whereas top women opt out from using it. Turning to the proficiency of AI use, we find that while male students are 34% more successful in writing prompts than female students on average, top women are just as good as men. In the top quintile, female students have a success rate of 46% vs. 39% for top male students. Finally, the most critical insight emerges from the policy experiment: top female students do not intend to use AI when it is forbidden in class while men across the admission grade distribution intend to use it at a high frequency regardless of whether it is forbidden or not. However, when AI is explicitly allowed, top female students intend to use it just as much as men.

Taken together, the supply and demand side 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. Our results suggest that the women opting out of AI are precisely those who would experience significant improvements in their labor market prospects by acquiring AI skills. Failing to adopt these skills could thus exacerbate existing gender gaps in career advancement. However, we also demonstrate that with the right encouragement, top women are willing and able to use AI tools. Importantly, in the prompting exercise they perform just as well as top men. Our findings provide a comprehensive understanding of how adopting and developing AI skills can 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 use and skill in generative AI, as well as the primary factors that drive the gap. Section 4 describes and shows the results of our policy experiment. In Section 5, we outline the experiments in the manager survey and present our results on the value of generative AI skills in two types of managerial decisions: hiring and promotions. 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), and (ii) whether these skills are valued by the demand side (employers). We use two complementary survey experiments: 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).

2.1 Survey Instruments and Administration

Survey of students. The first study aims to establish whether there is a gender gap in generative AI use among current students who will be facing a labor market that is rapidly changing due to this technology. The instrument was administered to 595 bachelor's and master's students at NHH Norwegian School of Economics in November 2023 and early 2024.[5] 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.[6] Questions regarding 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). Students were given the option of reporting their admission grade to university, with 328 students providing valid responses out of the 595 respondents (55% of the sample).[7] The full questionnaire of the student survey is in Appendix C.

[5] In November 2023, where most of our students sample comes from, we asked only about ChatGPT because other platforms were either not available or popular at the time. In 2024, we asked about ChatGPT and similar platforms and gave some examples on the alternative platforms.

[6] Descriptions of the policy experiment and the measure of prompting skills can be found in Sections 3 and 4, respectively.

[7] Higher education in Norway requires admissions to be based on an admission score determined through standardized testing and performance in high school. This provides us with a comprehensive measure of academic performance, which we exploit for heterogeneity analysis in our results.

Students were recruited during lecture hours of 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. The anonymous survey was implemented and supervised by the research team in the classroom using a QR code.

Survey of managers. The second study aims to evaluate whether generative AI skills are valued in the labor market by employers in two types of decisions: (i) hiring, and (ii) promotions. To achieve this, we conducted a survey experiment on a sample of 1,143 managers in Norway who work in areas where NHH graduates are commonly employed after graduation. 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. We also used a vignette experiment to determine whether managers would support for promotion workers who are more productive through the use of generative AI.[8] 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 its use by students. Finally, we collected information on background characteristics such as gender, age, level of education and tenure at the company.

Managers from Norway 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 industries/occupations. The survey was sent to 2,030 respondents in the Norstat panel. The full questionnaire of the manager survey is in Appendix D.

2.2 Sample and Participants

Students at NHH. The bachelor's program at NHH is the most popular program in Norway, listed as the first choice by most applicants to higher education.[9] The 2023 admission cutoffs

[8]Details on the conjoint-type and vignette experiments can be found in Section 5.

[9]In 2023, it was listed as a first choice by 2,170 applicants who competed for 500 slots. Almost 5,000 applicants listed the NHH program in any rank on their list. There were 62,757 higher education slots in Norway in 2023 (Direktoratet for høyere utdanning og kompetanse, 2023).

for first-time admission and regular admission were 55.6 and 59.5, respectively. For reference, grades in Norway range from 1 to 6, and GPAs are calculated from high school grades and the scores in five to six exams taken throughout high school (Landaud et al., 2023). The cutoffs, calculated by multiplying the GPA by 10, illustrate that successful applicants in both admission categories typically achieve scores close to a perfect 6 in every school and exam subject.

The school offers a 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. Education is free, and students who are admitted into the bachelor's program automatically secure a slot for the master's programs and typically continue with the master's, though they can leave after completing the three years of the bachelor's program only.[10]

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). Students took on average 8 minutes to respond to the survey.[11]

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).

Finally, we point out a few strengths of our sample that we think compensate for the apparent small sample size. First, the size of the typical cohort is 500, so considering that

[10]The bachelor's program is taught in Norwegian, while the master's programs are taught in English.
[11]On average, women spent 7.9 minutes and men 8.2 minutes. The difference is not statistically significant.

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 of 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.[12] Fourth, recent research focuses on students from the same university where we conduct our analysis, adding validity to our choice of sample (Cappelen et al., 2024).

Managers in Norway. We recruit managers from companies in the sectors that NHH graduates typically find jobs. Almost 90% of NHH 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 gradates from NHH will find jobs, we pre-selected managers working in the following industries/occupations: administration/personnel, banking/accounting/finance, consulting, and management services. In the study, these managers evaluate hypothetical first-time job candidates, recently graduated from NHH, making a direct link between the two samples.

Similar to the student survey, 60% of the sample is male. 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. Managers took a median of around 7 minutes to complete the survey.

2.3 Anonymity and Participant Incentives

In considering the best format to administer the survey, we weighed the prospect of linking student responses to their past and future academic performance against the potential for misrepresentation of generative AI use and experimenter-demand effects if students knew

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

that the survey was not anonymous. Since this is the first study documenting patterns in student use of generative AI, we opted for anonymity, valuing truthful responses above all.

Related to anonymity, incentivizing the truthful reporting of the measures collected and the prompting task would have required collecting some personal information to provide incentives. We also opted 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. Furthermore, incentivizing the hiring and promotion decisions would require a setting with a large and flexible sample of students, as well as a sample of managers actively seeking NHH graduates with generative AI skills, which was not feasible to collect (Kessler et al., 2019).

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, Hangartner, and Yamamoto, Hainmueller et al.; 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.

3 Gender gap in generative AI use

3.1 Main outcomes

We investigate two main outcomes related to use: adoption and skill. To generate our adoption measure we use 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 participant indicated "use it occasionally" or "use it regularly," which indicates continuous use. We also asked about a more 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."

To measure skill proficiency in the use of generative AI, 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.[13] 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, given by the proportion of ChatGPT answers that mention the official name out of over 50 queries made, for each prompt.

3.2 Econometric specification

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

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

We measure the gender gap through the coefficient α_1. 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 factors that may influence adoption such as AI-related preferences, perceptions, and experience in additional tables.

3.3 Main results

Generative AI adoption. Figure 1 shows the proportion of responses in each AI use category 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%

[13]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. We discuss some potential pitfalls of the task in Appendix E. Their prompting exercise was supposed to give the answer: Ebbinghaus-Titchener illusion. The Ebbinghaus-Titchener illusion (Titchener, 1901) is represented by two circles of the same size that are surrounded by a different context each: the first circle is surrounded by small circles and the second circle is surrounded by big circles. When most observers view these figures, the context affects perceptions of size. The image used is presented in Appendix C.

of female and 21.5% of male students have used it few times. Only 1 out of 595 students answered not having heard about it. 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.

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). This result is in line with Humlum and Vestergaard (2024) who find a 20 pp ChatGPT adoption raw gender gap in a sample of 100,000 survey respondents in Denmark. In terms of having a free or paid account to a generative AI chatbot,[14] 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 have a paid subscription, which we interpret as evidence that they have a higher willingness to pay for a more comprehensive generative AI toolkit.

The gender gap in adoption seems to be driven by students taking a first-year course, where adoption among female students is substantially lower (33.8% using occasionally or all the time) relative to female students taking higher-year courses (at least 85%). The gender gap in paid subscription, nevertheless, remains economically and statistically significant across students in different stages of the program (see Columns 1-3 of Table A1). These patterns may reflect differences in exposure as higher-year students have had more time to learn about and incorporate AI tools in a university setting. Whether these patterns suggest that female students can catch up to their male counterparts in terms of AI use with exposure 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.

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. The most popular task is "retrieving information" followed by "writing tasks." 65% of male students selected retrieving infor-

[14] For chatbots with paid subscriptions, e.g., ChatGPT and Claude, the monthly price is around US$20 as of June, 2024.

mation 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 on solving math questions and other tasks, which includes brainstorming. Importantly, all students are following the same study program and most of the subjects are mandatory as opposed to elective. Hence, the differences we find are not driven by self-selection into fields of study or specific subjects that lend themselves more or less to the use of AI.

Finally, we provide insights on heterogeneity by admission grade as previous research has shown that gender differences in other domains are particularly pronounced at the top of the skill distribution and understanding the effects of AI on people of different skill levels has been important in the emerging AI literature (Brynjolfsson et al., 2023; Dell'Acqua et al., 2023). We plot the high use variable by quintile of admission grade, a measure of relative academic ability.[15] Figures 3a and 3b show the fraction of female and male students reporting a high use by quintile of the admission score distribution.[16] 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 baseline 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).

The finding that women at the top of the skill distribution are less likely to use generative AI is particularly interesting in light of the work by Brynjolfsson et al. (2023), who find that using AI help reduces the quality of work for the most experienced workers at a technical-support firm. One may conjecture whether, for the best students, using generative AI might

[15]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 massification of generative AI.

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

reduce the quality of their output rather than improving it. If this is the case, top women would perform better in school an on the job because they do not use generative AI, and the gender gaps in the labor market could be reduced. While assessing the effects of using generative AI on human capital development is out of the scope of this paper, we report results from a survey of Norwegian managers in the sectors where NHH students are employed and find that using generative AI is a valued skill in both job applications and promotions (see details in Section 5).

Generative AI skills. As mentioned earlier, proficiency in AI tools like ChatGPT is becoming an increasingly important skill for labor market success (Amazon Web Services, 2024; Microsoft & LinkedIn, 2024). We note that while lower use rates can directly impact skill development since acquiring proficiency probably results from continued use of a tool, generative AI is a technology with low-entry costs and there are plenty of online resources providing guidance on how to interact with AI chatbots. We show that male students are more skilled at writing successful AI prompts than female students on average, and that this is driven by characteristics of the prompt such as the length and use of keywords.

Table 1, Panel B, Column 1 quantifies the raw gap in prompt success rates. The average success rate recording the fraction of times that the prompt obtains the desired answer (Ebbinghaus or Titchener illusion) for female students is 27.8%; in other words, their prompt gives the correct answer about 14 times out of 50 ChatGPT runs. The gender difference is estimated at 9.4 pp, which means that male students have success rates 34% higher than female students. In Column 2, we show that, on average, everyone spends about 129 seconds writing their prompt. Lastly, male students write about 31.6 more characters in their prompt relative to a mean of 145 characters among female students. These results seem to hold when looking at students at different stages in the program although the gender gap coefficient is no longer significant due to smaller sample sizes (see Columns 4-6 in Table A1). Appendix E provides more details on possible gender differences in recognizing the Ebbinghaus illusion and other prompting skills confounders.

As expected, in Figures 4a and 4b, students at the top of the admission grade distribution have higher success rates with their prompts regardless of gender. In the top two quintiles of the distribution, students have success rates of about 39-46%. 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 among all and half of the success rate for men in the same quintile (34%), who have similar levels of use.

In sum, our results on adoption and skills are in line with previous findings suggesting a correlation between women's choices according to their position in the skill distribution and choices based on laboratory tasks (Niederle and Vesterlund, 2007; Coffman, 2014; Isaksson, 2019; Kinnl et al., 2023) and on the grade in a principle's class determining what college major students enroll in (Rask and Tiefenthaler, 2008; Ost, 2010; Avilova and Goldin, 2018; Kugler et al., 2021; Ugalde, 2022). As in these previous studies, we find that the results concentrate on 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 (e.g., Bertrand et al., 2019).

3.4 Potential factors influencing adoption

An important aspect to understand the gender gaps reported in the previous subsections is to assess the role of different factors that may potentially affect or correlate with adoption. We elicited students' attitudes regarding generative AI, which we preregistered and classified into three categories: (i) preferences, (ii) perceptions, and (iii) exposure/experience. Preferences aim to measure potential utilitarian costs or benefits associated with AI usage and the role of persistence in AI use. Perceptions reflect perceived usefulness, whether generative AI usage is considered cheating, trust in the accuracy of information provided by AI chatbots, and confidence in one's abilities to use AI. Lastly, we explore exposure/experience, analyzing how prior exposure to AI might influence its adoption. The results are in Figure 5.

Gender Differences in Preferences. In Figure 5b we plot the share of students agreeing with *"I think ChatGPT is enjoyable to use,"* and disagreeing with *"I think ChatGPT is difficult to* *use,"* representing a utilitarian benefit and cost from using AI, respectively.[17] Male students have stronger preferences for the use of ChatGPT, as they find it more enjoyable (higher utilitarian benefit), and less difficult (lower utilitarian cost) to use than women. To measure "persistence" we asked *"If ChatGPT does not provide the desired answer on your first attempt, how many additional attempts do you typically make?"* with four options ranging from "One more try" to "I keep until satisfied." We find that 55% of female students indicate that they attempt twice or more, compared to 71% of male students, which indicates that men are more persistent as they maintain longer "conversations" with ChatGPT, something that could generate differences in skill as men can learn more from the increased prompting experience.

Gender Differences in Perceptions. We consider in Figure 5a belief-based motives that can affect behavior in our setting, which we categorize as perceptions regarding: (i) AI use considered as cheating, (ii) confidence in one's own skills using AI, (iii) trust in accuracy in providing information, and (iv) usefulness.

First, students might not adopt the technology if they perceive its use is unethical/cheating. We plot whether students disagree or strongly disagree with the statements *"Using ChatGPT as an aid to solve assignments in a course is equivalent to cheating"* and *"Using ChatGPT as a learning aid in a course is equivalent to cheating"* to capture this concept. While the majority of participants disagree with considering the use of ChatGPT as equivalent to cheating, there are important gender differences, with around 12 pp more men disagreeing relative to women. The levels in these two questions are relevant too, with 88% of male students disagreeing that ChatGPT as a learning aid is cheating, relative to 64% disagreeing when the use is as an aid to solve assignments. Around 52% of male students disagree with the statement *"It is easy for professors to identify if a student has used ChatGPT,"* relative to 44% of female students.

Second, being confident in one's own skills in using the technology might affect students' willingness to engage with AI, especially if it is perceived as a male-dominated setting (Coffman et al., 2023). To measure confidence, we use the prompting task the students performed,

[17]The binary variables are equal to one if students select agree/strongly agree or disagree/strongly disagree in a 5-point scale. The questions mention ChatGPT specifically, but we said before that it could be ChatGPT or similar tools.

and asked them *"How confident do you feel that the query you just provided will make ChatGPT get the information you need?,"* with choices within a 4-point scale ranging from "Not confident at all" to "Extremely confident." We observe important differences in confidence by gender with 60% of women and 81% of men indicating some level of confidence in their prompt. Moreover, as depicted in Figure A2a and in line with the literature, over 40% of male students indicate feeling very or extremely confident in their own prompt being correct, relative to only 18% of female students. When comparing the self-reported confidence with their actual performance in the task (a measure of overconfidence), we find that male students are 7 pp more overconfident that their prompt was correct relative to 38% of female students (see Figure A2c).

Third, there could also be potential differences in trust in the accuracy of the information provided by ChatGPT. For example, hallucinations may affect the perceived benefits of using the technology. We presented students with a screen capture of a real prompt and answer submitted to and by ChatGPT, respectively, and asked them whether they trust that the information provided by ChatGPT was accurate, using a 4-point scale from "Completely trust" to "Completely distrust."[18] Figure 5a shows that there are no differences in trust, with 63% of both male and female students indicating either "Somewhat trust" or "Completely trust."

Fourth, as highlighted in previous work on the "gender digital divide," perceptions on the usefulness of technology in different tasks seemed to be a driving factor of the gender differences in the use of the internet (OECD, 2018). We capture perceptions of usefulness of ChatGPT by asking students to indicate *"What do you believe are the main advantages of using ChatGPT in coursework?."* Figure 5a shows the percentage of students that indicated each statement as an advantage of using ChatGPT. While almost no one sees no advantages of using ChatGPT, there are strong gender differences in perceptions of usefulness 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

[18]The query asked to ChatGPT in the example provided was the following: "What is the poverty rate in Denmark?." The participants were later asked, *"Based on this response from ChatGPT, how much do you trust that the poverty rate reported is accurate?"* (see Appendix C).

improves the learning of course methods (56% vs. 43%). However, in terms of saving time, there are no strong gender differences in perceptions, with around 74% of both male and female students believing it is a main advantage of generative AI.

Gender Differences in Experience or Exposure. A gender gap in AI use and skills might be driven by male and female students having different levels of experience or exposure to the technology, through peers or their own previous experience. To measure exposure through peers we asked participants to *"indicate the percentage of people you believe use ChatGPT"* for three different groups: their group of friends, students in their course, and professors at NHH.[19] Figure 5c shows the average percentage indicated by the students for each of the groups. There are no substantial gender differences in these beliefs with students stating that almost 75% of friends and students in their course, and that about 45% of professors at NHH use ChatGPT. To measure own experience we asked students whether they have *"ever received inaccurate or misleading information from ChatGPT?,"* with possible answers being "No, never," "Yes, few times" and "Yes, many times," as well as an option for those who have not used it. In Figure 5c, the percentage of students who have experienced inaccurate information is 15 pp higher for men than for women, the latter being only 28%. Altogether, this evidence shows that exposure from their surroundings may not influence gender differences in adoption and skill, but own previous experience might.

3.5 Revisiting the gender gap in adoption and skill

We now aim to understand the relationship between the gender differences in the influencing factors in the previous section and AI adoption and skills. To do this, we add baseline characteristics and the preferences, perceptions and experience/exposure measures discussed above as controls in the regression of the main adoption and prompting skills outcomes. While most of these controls are clearly not exogenous since they could both be consequences as well as causes of students' use and proficiency with generative AI, this exercise may help understand which factors have a stronger influence or correlation with the main outcomes.

[19] To avoid concerns of men and women having different anchors when estimating this percentage, we provided the following statement before the question: "A survey conducted among university students in the US in the Spring of 2023 reports that 30% of students use ChatGPT for their schoolwork."

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 of each column.

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 to 0.8 pp 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). These small gaps are statistically insignificant after adding the full set of controls. Columns 2-5, which add the groups of controls individually in each column, suggest that the perceptions measures (cheating, overconfidence, trust and usefulness) are the ones that help reducing the gender gap the most for both the high use and subscription outcomes.

In the success rate of the prompts (Panel C of Table 2) we do not see any sets of covariates substantially reducing the gender gap. We perform a text analysis in Appendix E using a Lasso cross-validation methodology to identify the top keywords that predict a successful prompt. Once controlling for keywords and number of characters, we can fully explain the gender gap in prompting skills (see details in the appendix).

4 The impact of policies on the gender gap in generative AI use

Given 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 survey a policy experiment to assess student responses to such policies.

4.1 Experiment design and main outcomes

We rely on a hypothetical vignette experiment as follows.[20] 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:

[20]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.

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."

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 allows us to causally study the effects of the allow/forbid policy on intended use. A second layer of randomization was the type of evaluation of the course, where the evaluation could be either an in-person exam or a home exam.[21]

4.2 Econometric specification

Our second econometric specification involves estimating the gender gap for 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 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.

[21]Respondents that were presented with the home exam scenario were asked a second question: "*Given this 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. We are not using this layer of randomization in this draft.

4.3 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, we note that intended use is higher for both men and women under the hypothetical scenario when ChatGPT is allowed in the course than the adoption measure in Section 3.3. Our take on this difference is that, up to December 2023, there was no AI policy at NHH and without such policy the default behavior is up to students' interpretation, and some of them may interpret no rule as not allowed/encouraged.[22] Second, the differential response by academic skill for female students diminishes the weight of arguments that our results are driven by social desirability bias. One would have to make complicated assumptions on how social desir-

[22]The policy released in December 2024 provided guidelines for more transparent rules on how generative AI should be used and graded in the courses.

ability bias interacts with relative academic skill and gender to explain the results. Third, 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 female and male students spent 32 and 31 seconds, respectively, in the vignette experiment, so it does not seem that inattention may be differential by gender. Fourth, 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.

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 in student's adoption of AI and potentially their prospects of success in a rapidly evolving labor market. We analyze whether this is likely to be the case next.

5 Value of Generative AI Skills in the Labor Market

5.1 Experiment design and main outcomes

We assess whether the use of generative AI is valued by managers in two main decisions: hiring and promotions. Two experiments were implemented in the survey of managers.

The first experiment evaluates hiring decisions. In a conjoint-type design, each manager is randomly matched with two hypothetical candidates represented by profile cards. The cards contain basic information about the candidates, including gender, signaled through their name, grade in a core course of the bachelor's program, skills, degree and age (see Figure A3 for an example of a card). All job candidates presented to managers were 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.

Three main dimensions in the profiles were manipulated. First, gender was represented by assigning either a male or a female name. Second, we vary whether the candidate has 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). Finally, the profile card 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] The managers were randomly presented two profiles out of five possible pre-specified types:

1. **Top Woman No AI**: a female candidate with high scores and no generative AI skills.
2. **Top Woman AI**: a female candidate with high scores and generative AI skills.
3. **Top Man No AI**: a male candidate with high scores and no generative AI skills.
4. **Top Man AI**: a male candidate with high scores and generative AI skills.
5. **Low Man AI**: a male candidate with low scores and generative AI skills.

[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%.

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.

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 cards were manipulated simultaneously, we do not worry about potential experimenter demand effects, as it is unclear for the managers which of the characteristics changed is the most meaningful for the experimenter.[24]

As companies expect that the use of generative AI will become widespread in the near future (Amazon Web Services, 2024), we also measured managers' scores for certain 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 generative AI skills could compensate for lower grades relative to female candidates with high grades but no AI skills.

The second experiment studies promotion decisions. Each manager was presented with one hypothetical scenario as follows:

> *Daniel and Martin 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.*

We use insights from recent research to represent in the experiment the productivity benefits of using generative AI in the workplace (for a review see Capraro et al., 2024). Our main treatment variation corresponded to the disclosure of workers' performance time on the task. In the first treatment arm, *Known*, participants were told the following: *"Both Daniel and Martin complete the task with the same level of quality. Daniel took 8 days to complete*

[24]For more details on the randomization, see Appendix F.

it without generative AI. Martin used generative AI and completed it in 6 days." In the second treatment arm, *Unknown*, participants were told: *"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."* Note that in both treatment arms, one worker is 25% faster in completing the task, and 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."

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 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.

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

$$y_i = \beta_0 + \beta_1 \text{Known}_i + \epsilon_i \quad (4)$$

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

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 which 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 and if this selection differs based on whether the use of generative AI is 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.[25]

5.3 Main Results

Hiring. Managers gave an average score of around 6.5 to the hypothetical candidates, with 6 representing a "Good candidate" in the scale. 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 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 (benchmark). For male candidates, the premium is close to zero and not statistically significant. When comparing male candidates with low grades and AI skills ("Low Man AI") relative to the benchmark, we observe that grades still play an important role in the evaluation, as the low-performing male candidates are graded 11.1% lower, in spite of their AI skills. Furthermore, we observe that, in expectation, this gap persists at the same level in three years' time (column 2). The results hold even when controlling for other characteristics of the profiles and the managers, such as the grade and course distribution of the hypothetical candidate, the gender of the manager, and candidate order fixed effects (see Table A2). In Appendix F we discuss additional evidence indicating that managers are more likely to call the candidate

[25]In our preregistration, we indicated an interest in conducting a heterogeneity analysis using a specification similar to equation (5), but with an indicator variable representing the gender of the fastest worker instead of the "Encouraged" variable. Our findings in the preregistered analysis were difficult to interpret without speculation. Therefore, we decided to discuss the preregistered analysis in the Online Appendix and focus on the heterogeneity by company policy, which we believe is more relevant to the purpose of the paper.

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.

A potential explanation for the premium benefiting only female candidates with generative AI skills is that the generative AI expertise signal might be differently informative. 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 informativeness of the signal. For example, if a manager believes that women are not as likely as men to use generative AI, the signal of expertise might be more informative for women than for men. This is also in line with Bohren et al. (2019) who find that discrimination reverses in favor of women at the top in male dominated areas. 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. Table A3 shows a breakdown of the estimates for equation (3) by the subsamples of managers with correct and incorrect perceptions, with the score given to the hypothetical candidate as the dependent variable. The positive premium in scores for women signaling AI skills comes from managers who have correct perceptions of the gap. 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.

We validate the findings 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 A4).[26]

Taken together, the evidence suggests that signaling generative AI skills is valuable in hiring decisions, specifically for top women. If top women do not experience or use generative AI skills, and thus do not signal them, they might be missing opportunities to increase their chances of success in the labor market.

Promotions. A rapidly growing body of work has shown that the use of AI in the workplace can lead to substantial productivity gains (see review in Capraro et al., 2024). Panel B of Table 4 summarizes our findings on whether these productivity gains 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 when it is unknown. The majority of managers select the fastest candidate in both cases, with the proportion (65%) being significantly higher than selecting at random (50%).[27] When it is known who used generative AI, 56% of managers select the fastest candidate to the promotion track relative to 74% when it is not known. We take these findings as evidence that using generative AI in the workplace when there are productivity gains would be rewarded in promotion decisions.

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 associated with the use of generative AI in the workplace. To explore this hypothesis, we analyze 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

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

[27]We consider the random selection of a candidate (50%) as a benchmark following the case when two workers have the same level of skills and would have performed equally in the absence of generative AI aid.

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.[28] Figure A5 shows that 42% of companies that hire NHH/BI graduates allows 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 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.

We provide further interpretation and additional results from the manager survey and a discussion of social desirability bias in Appendix F.

6 Discussion and Conclusion

We conducted two survey experiments with students at NHH Norwegian School of Economics and managers of companies in the sectors where NHH graduates are often employed.

[28]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.

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 survey of managers indicates that top female students would greatly benefit from acquiring generative AI skills as these are rewarded in hiring and promotions. Overall, even though AI skills are valued by employers, our results suggest that low levels of adoption among top women would not necessarily harm their labor market entry prospects, but would rather limit their possibility to reach their full potential and career advancement.

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 technologies 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., 2023). 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 indicates otherwise. 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. Instead of relying on organic changes, we advocate for deliberate policy interventions. With well-designed policies, potential learning losses could be minimized, and top female students with AI skills could gain an advantage in hiring and promotions.

Finally, we point out some limitations to our study. First, although our sample has desirable characteristics such as a high degree of homogeneity among students, it is a specific setting, and we do not know how our findings would generalize to other educational programs and institutions. Second, we rely on hypothetical experiments because conducting the type of experiment needed to reach similar conclusions is not feasible and we discuss potential pitfalls of the study design in the text. Despite these limitations, we believe our findings are indicative of emerging trends in the value of generative AI across education and labor markets.

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.

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. R. Raymond (2023). Generative AI at work. Technical report, National Bureau of Economic Research.

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.

Cappelen, A., S. Jensen, M. Recalde, K. Salvanes, B. Tungodden, and L. Vesterlund (2024). Early onset of gendered work assignments.

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).

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.

Coffman, K. B., M. R. Collis, and L. Kulkarni (2023). Whether to apply. *Management Science*.

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*.

Dell'Acqua, F., E. McFowland, E. R. Mollick, H. Lifshitz-Assaf, K. Kellogg, S. Rajendran, L. Krayer, F. Candelon, and K. R. Lakhani (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (24-013).

Direktoratet for høyere utdanning og kompetanse (2023). Hovedopptak til høyere utdanning ved universiteter og høgskoler. https://hkdir.no/rapporter-undersokelser-og-statistikk/hovedopptak-til-hoyere-utdanning-ved-universiteter-og-hogskoler-juli-2023. [Online; accessed 12-December-2023].

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.

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. and A. Hawkins (2023). Strategic Decisions have “Major” Consequences: Gender Differences in College Major Choices.

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

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

Haslberger, M., J. Gingrich, and J. Bhatia (2024). Rage against the machine? generative ai use, threat perceptions, and policy preferences. *Generative AI Use, Threat Perceptions, and Policy Preferences (June 25, 2024)*.

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*.

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.

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.

Landaud, F. and É. Maurin (2020). Aim High and Persevere! Competitive Pressure and Access Gaps in Top Science Graduate Programs.

Landaud, F., É. Maurin, B. Willage, and A. Willén ((accepted) 2023). The value of a high school gpa. *Review of Economics and Statistics*.

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].

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.

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*.

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.

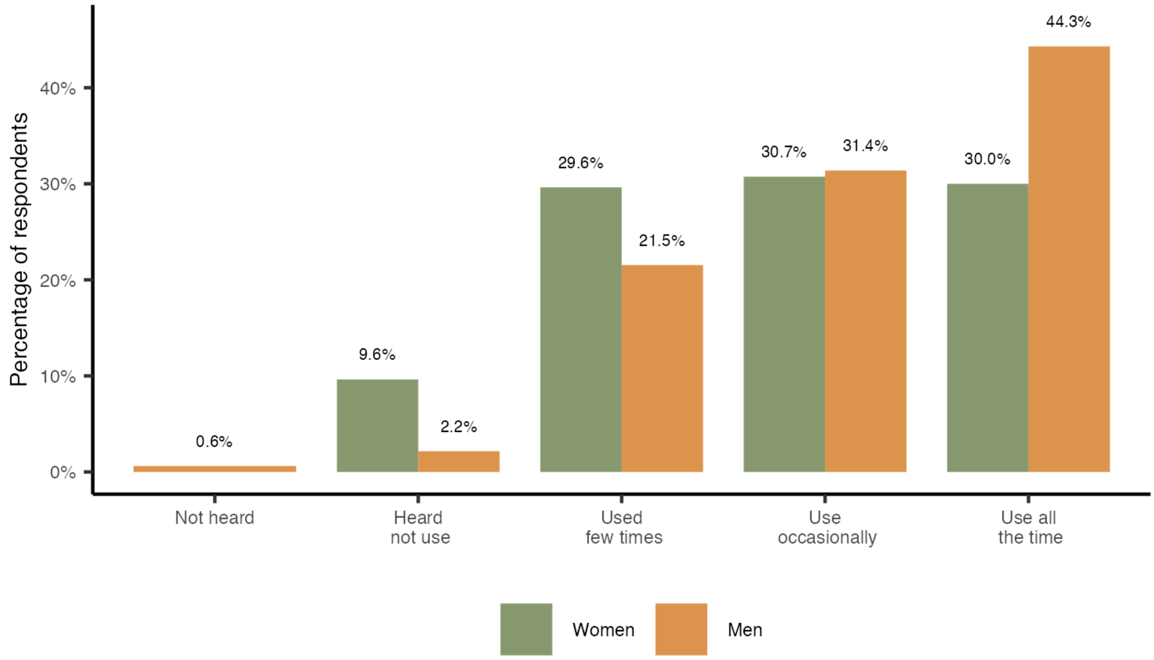
Stango, V. and J. Zinman (2023). We are all behavioural, more, or less: A taxonomy of consumer decision-making. *The Review of Economic Studies 90*(3), 1470–1498.

Titchener, E. B. (1901). *Experimental psychology: A manual of laboratory practice, vol. I: Qualitative experiments*. New York: The Macmillan Company.

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

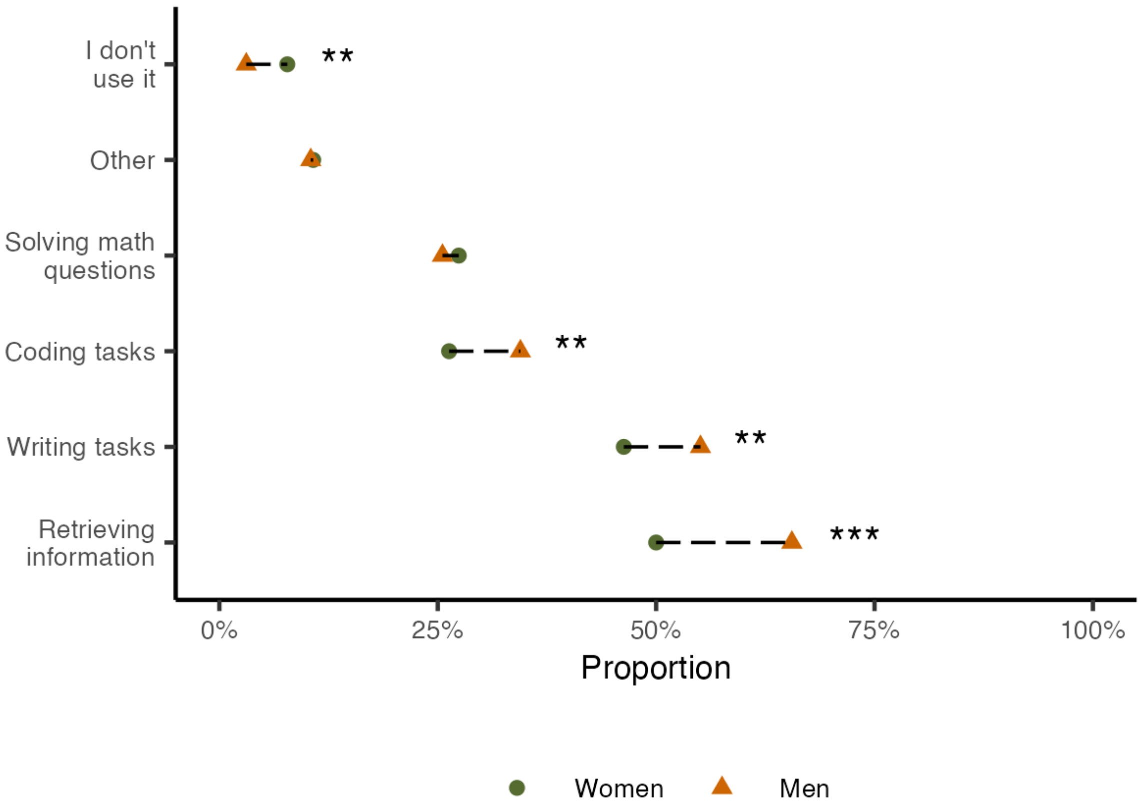
7 Figures

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%.

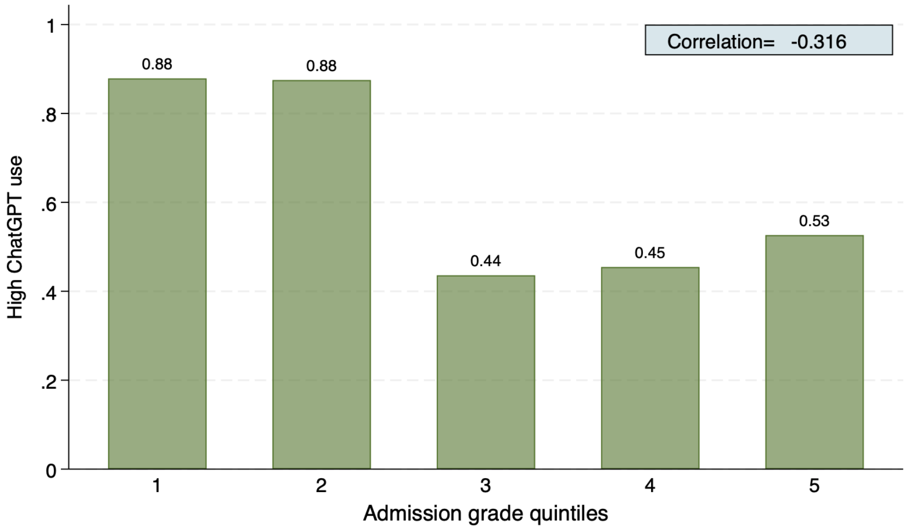
Figure 2: Tasks for which students typically get AI help



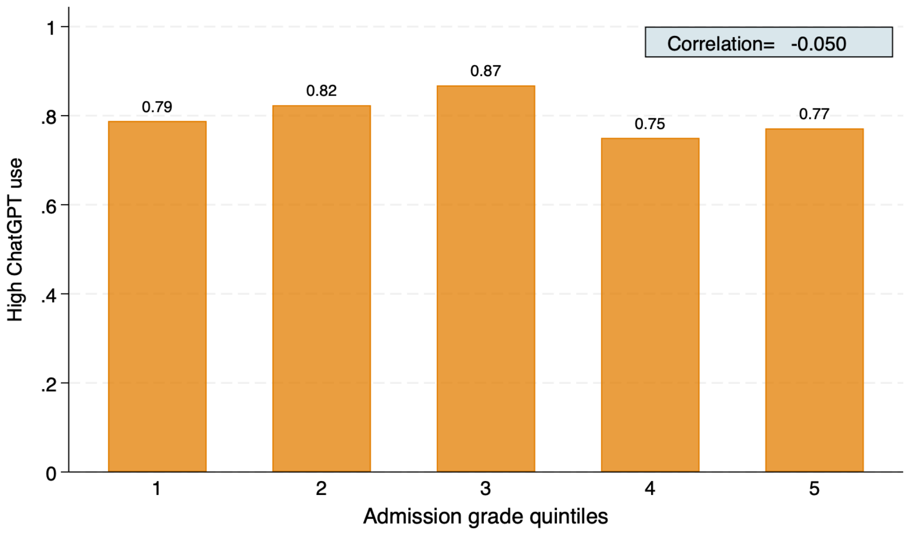
Notes: The figure plots the fraction of students who select the options on the vertical axis in the question: "What type of tasks do you typically ask ChatGPT to help with? (Please select up to the most common three)." The option "I don't use it was added so students who do not use AI could answer the 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



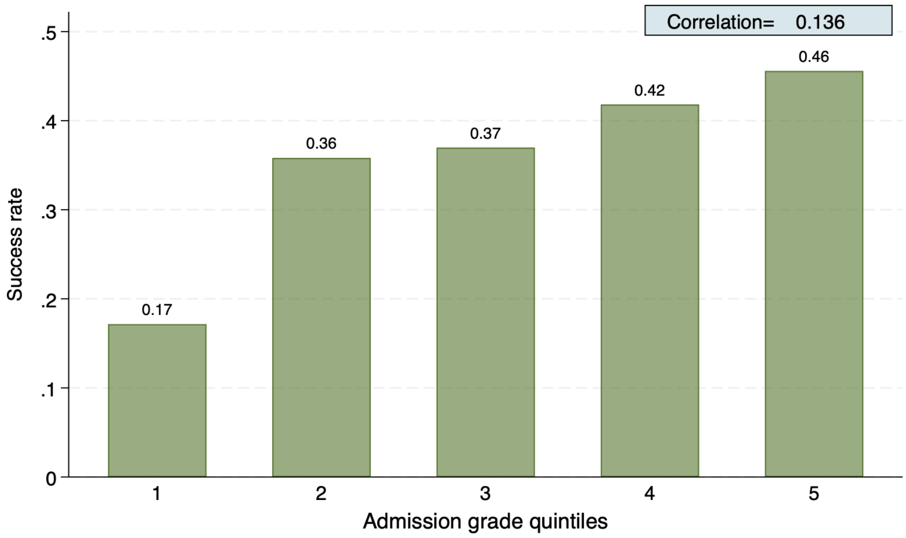
(b) Men



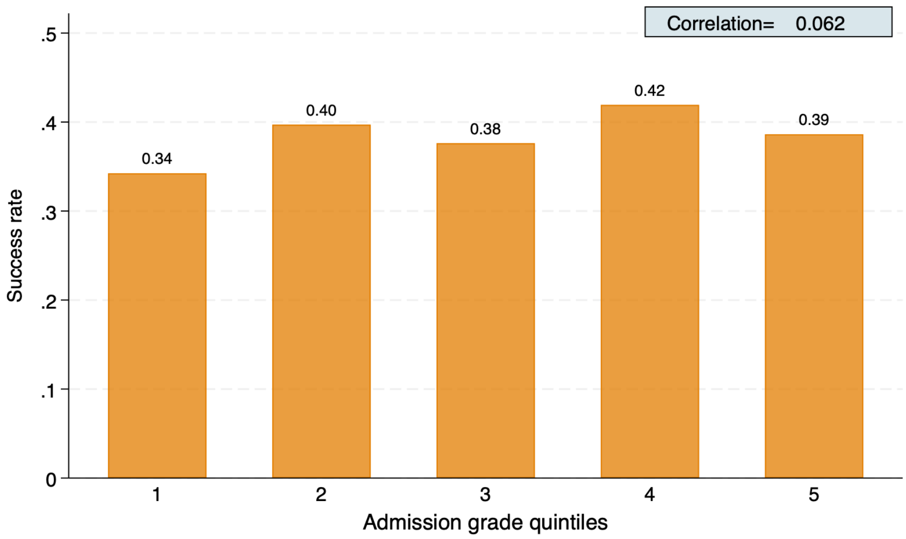
Notes: Panels (a) and (b) show the proportion of women and men, respectively, with high use of generative AI (use occasionally or all the time) across the self-reported admission grade quintiles (328/595 respondents).

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

(a) Women

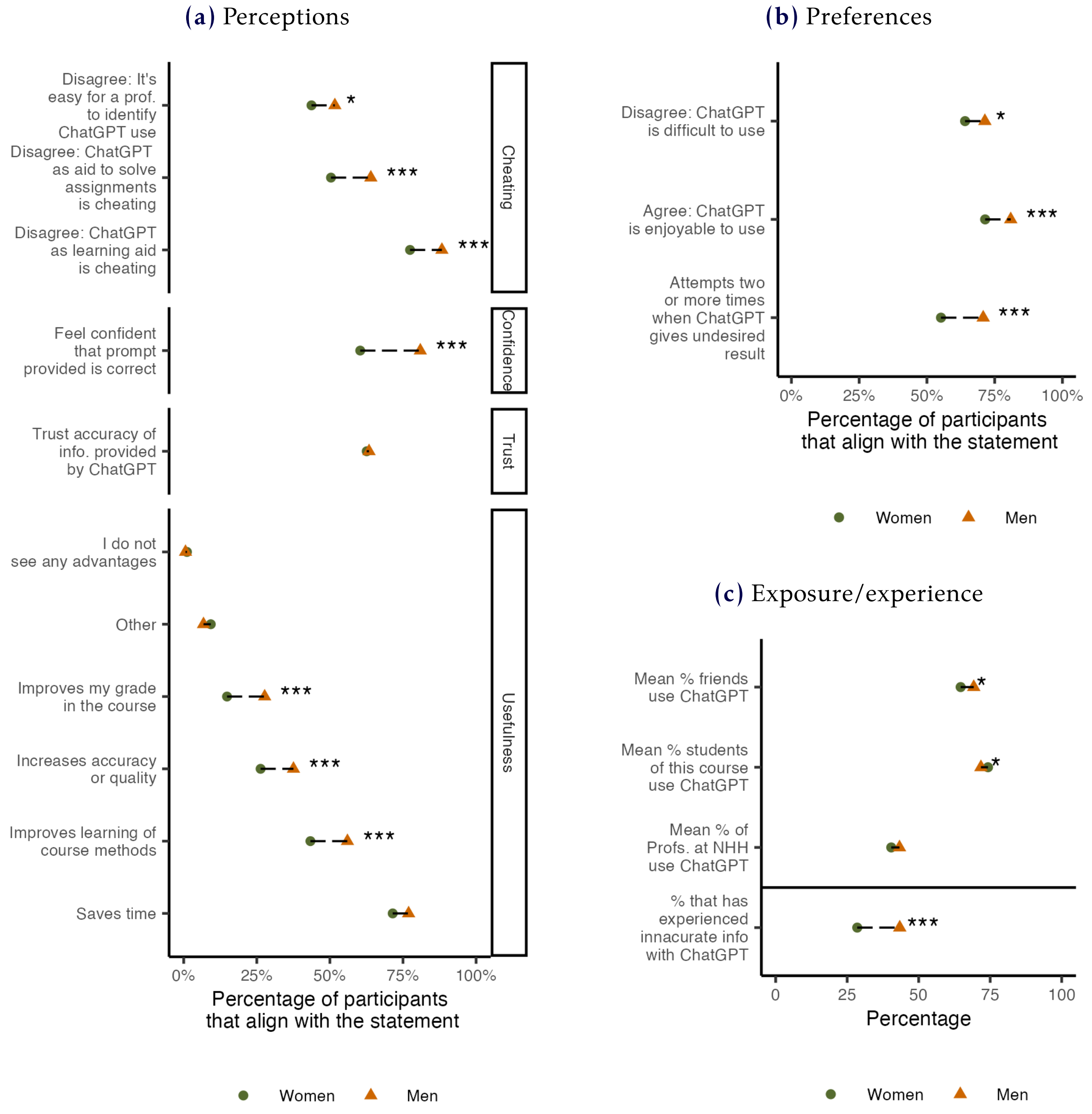


(b) Men



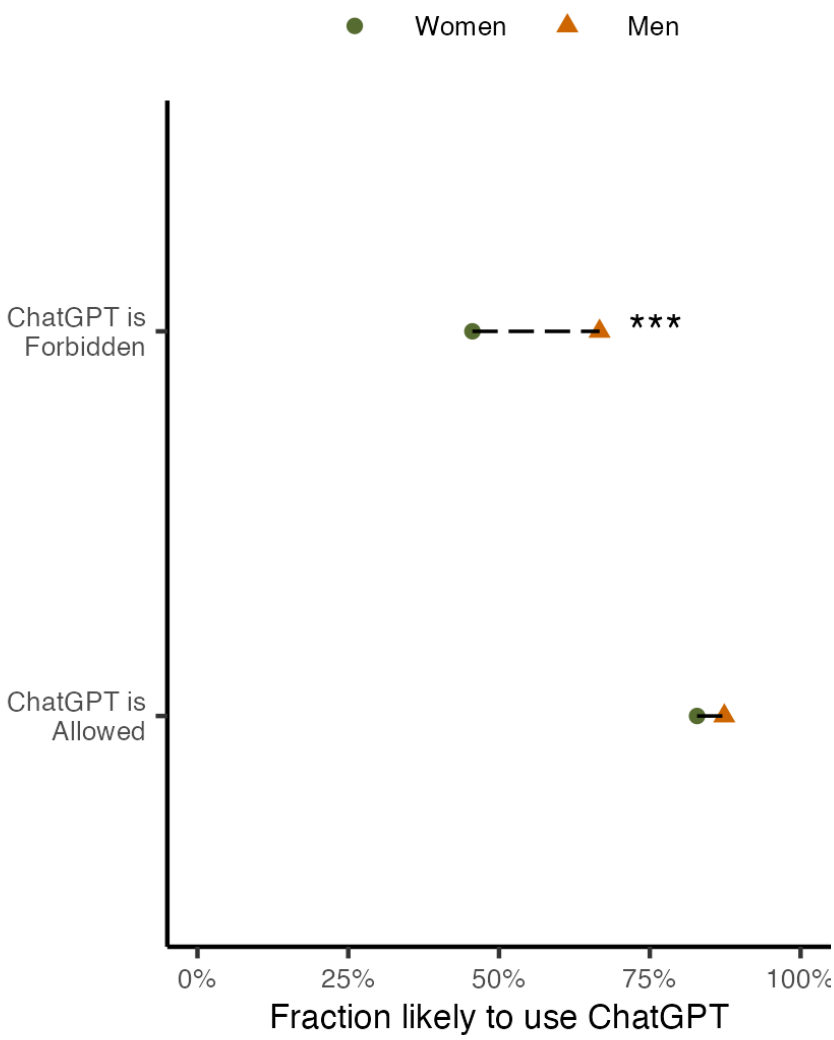
Notes: Panels (a) and (b) show the average success rate in the prompting task for women and men, respectively, across the self-reported admission grade quintiles (328/595 respondents). The success rate is calculated running each student's prompt 50 times on ChatGPT and recording how many times the prompt gets the correct answer.

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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

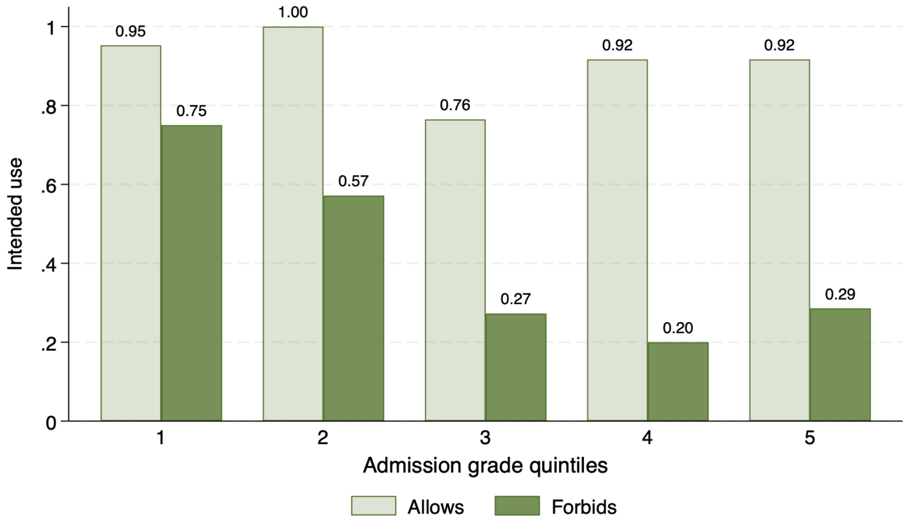
Figure 6: Policy responses



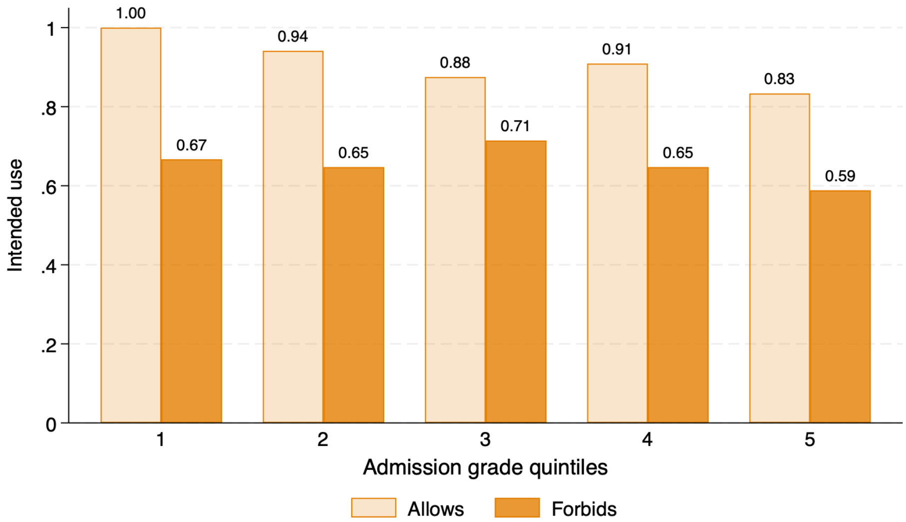
Notes: The figure shows, by gender, the fraction of students that indicated "Somewhat likely" or "Very likely" to the question of how likely would they use ChatGPT in the hypothetical course presented in the vignette experiment. We show the estimates for the two randomly assigned scenarios: professor "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). In brighter colors is the intended use in the professor "allows" treatment, whereas in darker colors is the intended use in the "forbids" treatment.

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 gneder 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)
Men AI premium (p-value)	0.646	-
Top Woman No AI (mean)	6.386	6.562
Fixed Effects	Manager	None
Observations	2,286	1,143

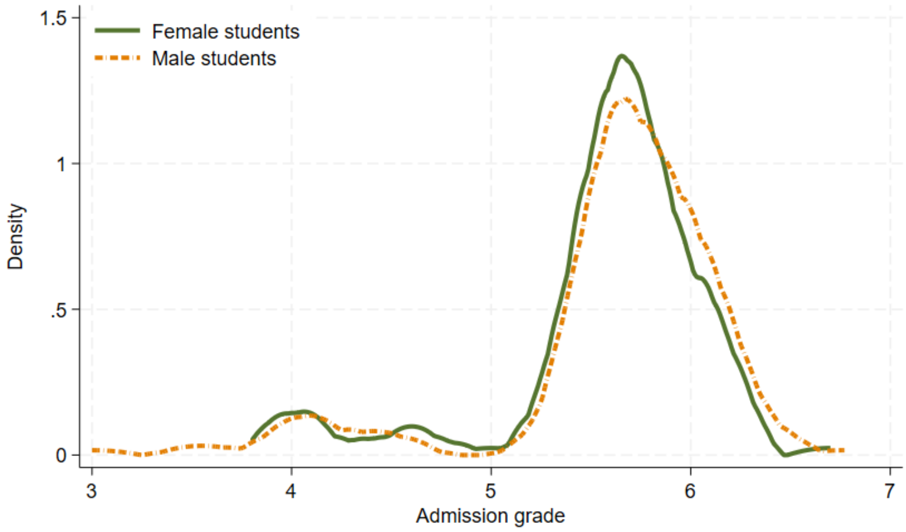
Panel B: Promotion decisions

	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%. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A Appendix Figures

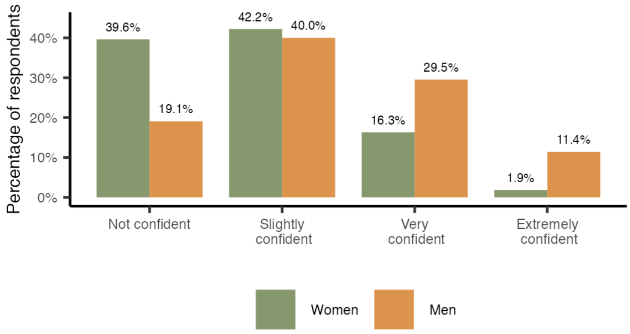
Figure A1: Distribution of admission grades by gender



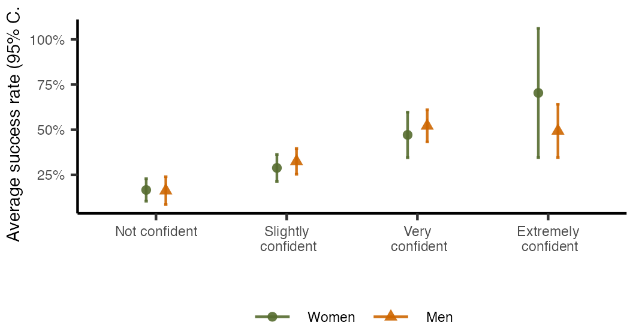
Notes: The plot shows the density of continuous admission grades separately for male and female students. Admissions take into account the high school GPA that is between 1 and 6. If not admitted to their intended program fresh from high school, students can take extra credits and increase their GPA to apply again. This is the reason why the admission grade can be above 6. NHH admits students through tow 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.

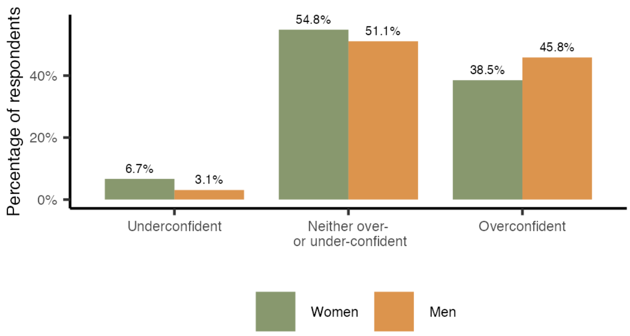
(a) Confidence in own prompt



(b) Success rate by level of confidence



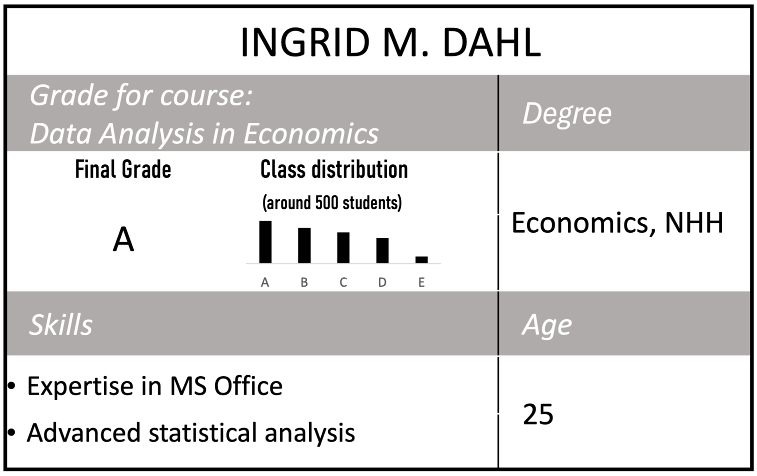
(c) Under or overconfidence in their prompt being successful.



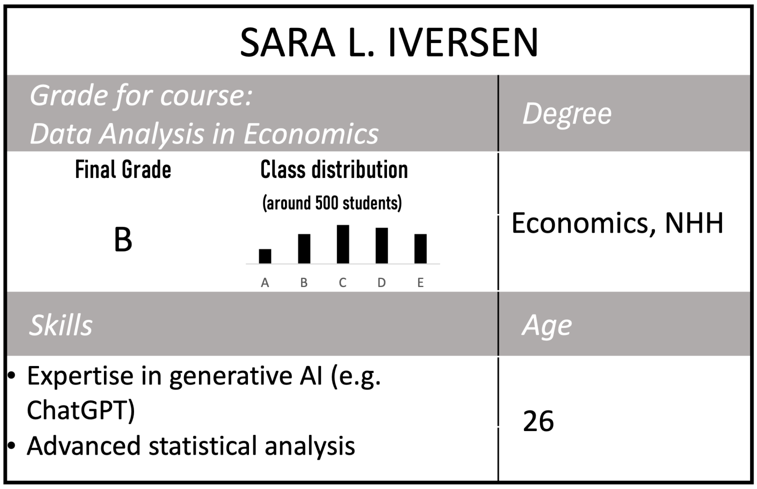
Notes: Panel (a) shows a bar plot with the percentage of women and men indicating each answer to the question "How confident do you feel that the query you just provided will make ChatGPT get the information you need?", which they answered after the prompting skills task. Panel (b) shows the average 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- or 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: Example of profile cards.

(a) Profile card for: Top Woman No AI



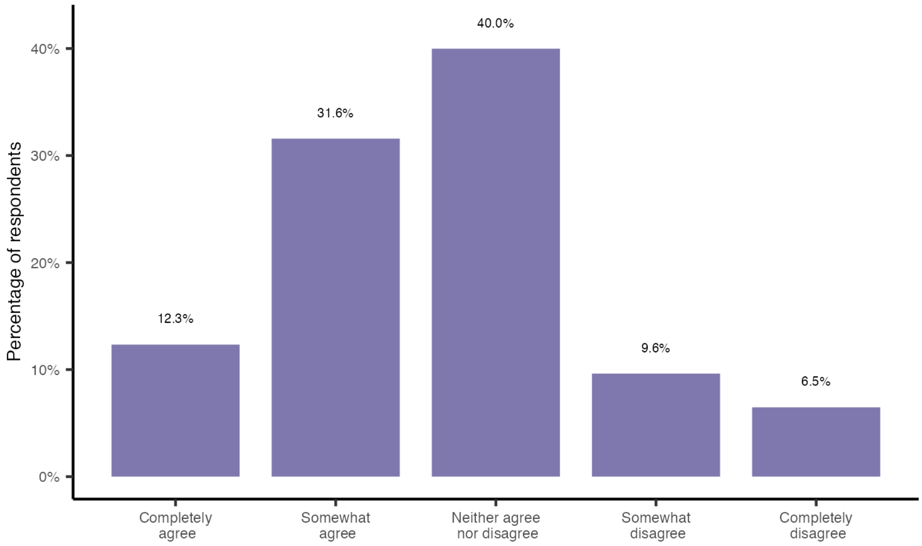
(b) Profile card for: Top Woman AI



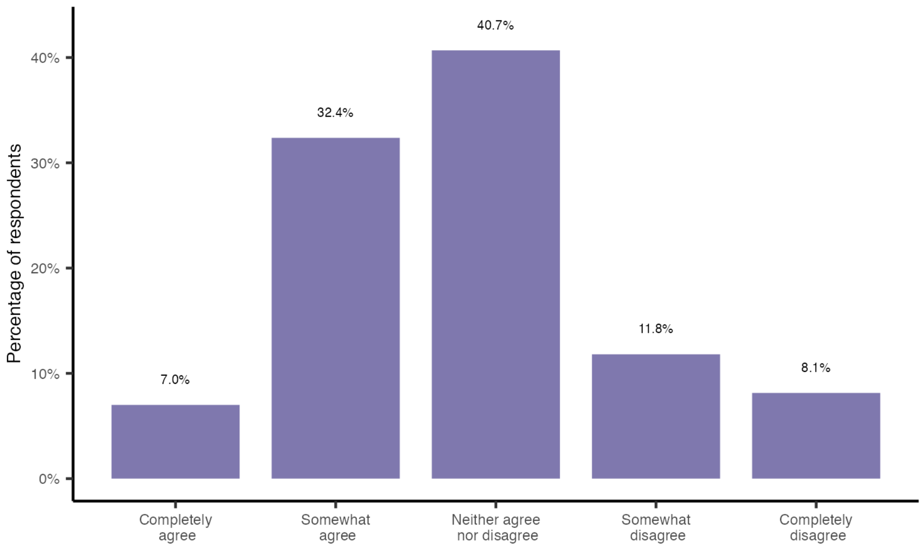
Notes: The figure shows two examples of profile cards of top-performing women—top 30% in class distribution of the course—presented to the managers. In the experiment, the name of the participant, the skills, the grade and grade distribution, and the age can change. In Figure (a), we show a high-skill female candidate without AI skills, and in Figure (b), we show a high-skill female candidate with AI skills.

Figure A4: 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.

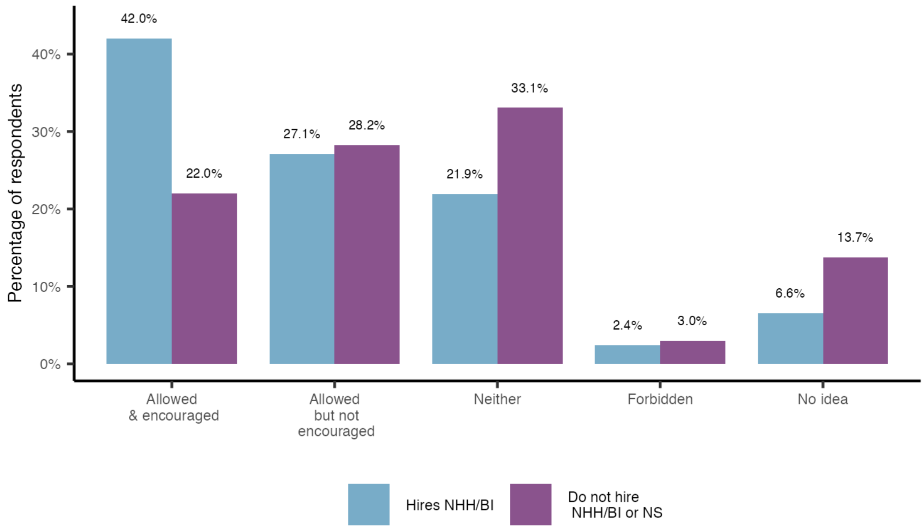


(b) Having generative AI skills can help a graduate earn a higher salary in their first job.



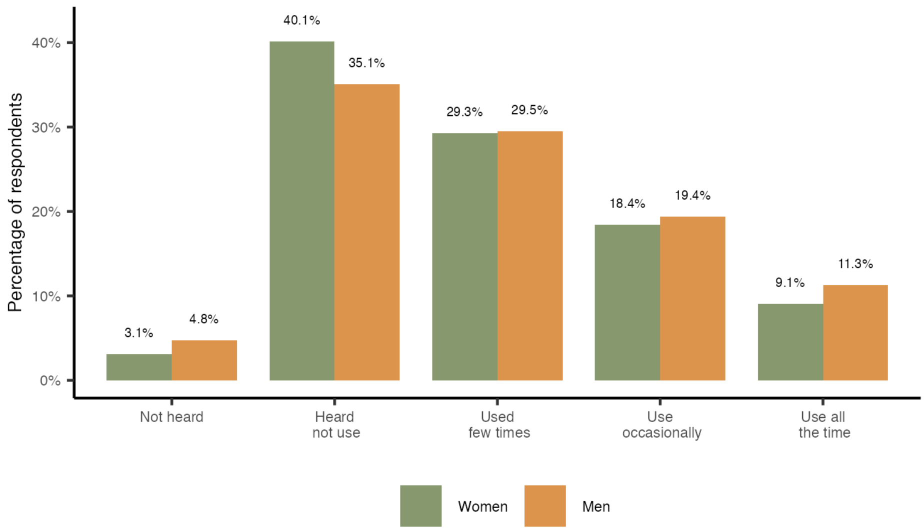
Notes: Panels (a) and (b) show the distribution of the answers of the extent to which participants agree/disagree the statements indicated in subcaptions. Within gender (color) the bars add up to 100%.

Figure A5: Company's policy by whether the company usually hires NHH/BI students



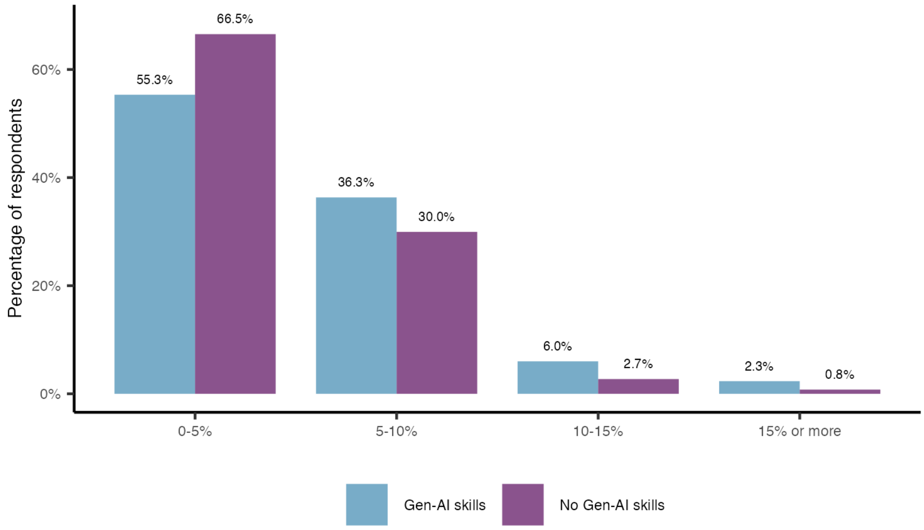
Notes: Managers are asked in the survey: "Does your company/your department employ newly graduated candidates with a master's degree in economics and administration (for example, candidates with a master's degree from NHH or BI)?" We split the sample into two, for participants who answered "Yes" and 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 A6: Familiarity of managers with generative AI, by gender of the manager.



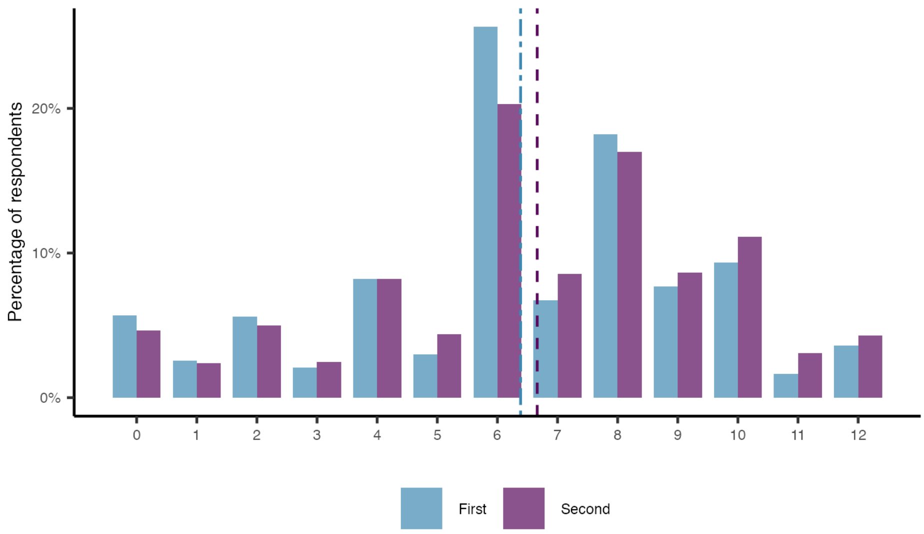
Notes: The figure shows a bar plot with the percentage of female and male managers indicating each answer to the question "How familiar are you with ChatGPT or similar tools?". Within gender the percentages across categories add up to 100%..

Figure A7: Salary negotiation potential of a candidate with and without generative AI skills



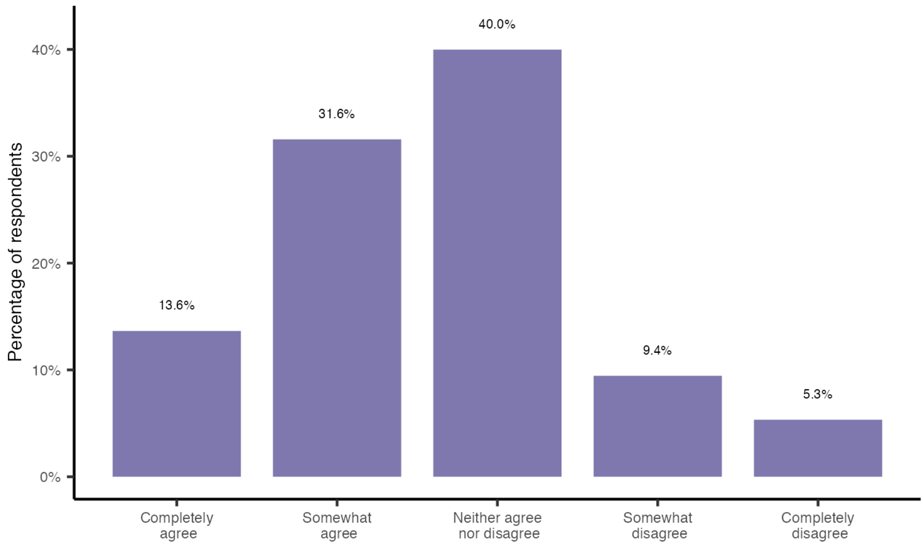
Notes: We consider the subsample of candidates whose evaluating manager faced two candidates with high grades where one candidate has generative AI skills and the other does not have generative AI skills. The bar plot shows the salary negotiation potential conditional on the candidate being selected 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?" 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 A8: Distribution of scores by whether the candidate was shown first or second



Notes: Each manager evaluates two hypothetical candidates. The barplot 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.

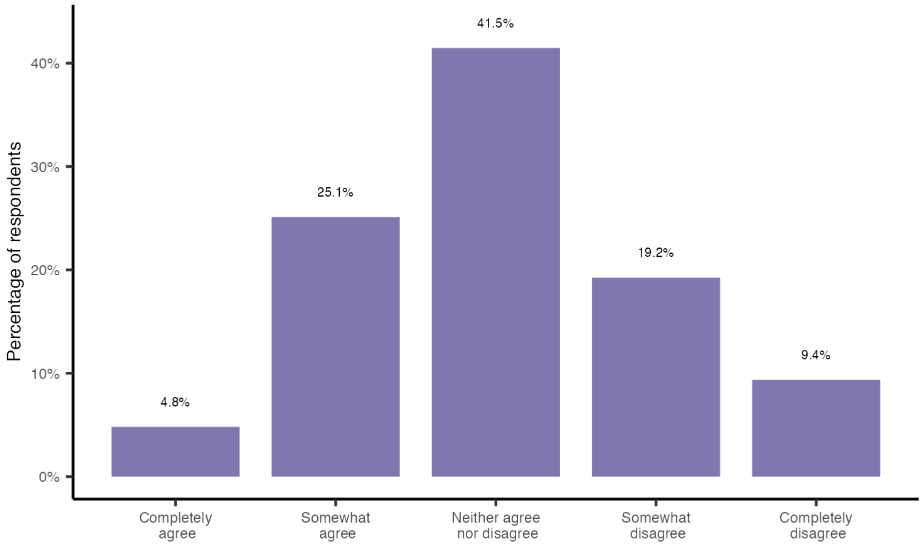
Figure A9: Level of agreement to the following statement: "If a student achieves higher grades by using generative AI, it is because the AI tools effectively improve learning, rather than replacing individual efforts".



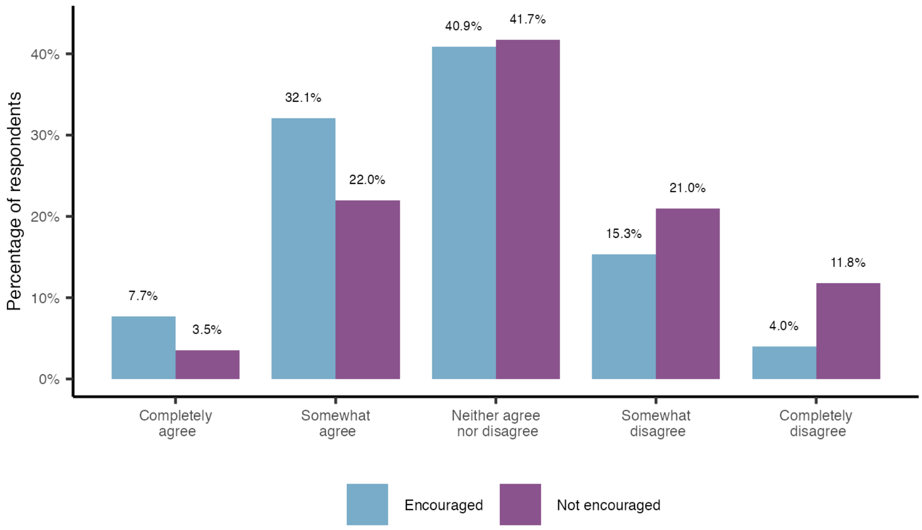
Notes: The plot shows the distributions of the answers to the statement for the full sample of managers. The bars add up to 100%.

Figure A10: 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 replacing individual efforts".

(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: AI adoption by course and year in which survey was administered

	Use	Has a subscription		Regular use for	
	(1)	(2)	(3)	(4)	(5)
	High use	Free	Paid	Coursework	Not coursework
Panel A: Students in first year course (2023)					
Male	0.273***	0.014	0.117***	0.112**	0.101***
	(0.058)	(0.052)	(0.032)	(0.044)	(0.037)
Constant	0.338***	0.243***	0.022*	0.110***	0.059***
	(0.041)	(0.037)	(0.013)	(0.027)	(0.020)
Controls	No	No	No	No	No
Observations	280	280	280	280	280
Panel B: Students in third year course (2023)					
Male	-0.031	-0.133**	0.095	0.016	0.097
	(0.045)	(0.065)	(0.063)	(0.071)	(0.065)
Constant	0.897***	0.368***	0.241***	0.471***	0.264***
	(0.033)	(0.052)	(0.046)	(0.054)	(0.048)
Controls	No	No	No	No	No
Observations	206	206	206	206	206
Panel C: Students in second year and master's courses (2024)					
Male	0.036	-0.199**	0.152**	0.056	0.019
	(0.066)	(0.094)	(0.072)	(0.155)	(0.141)
Constant	0.851***	0.489***	0.106**	0.500***	0.278**
	(0.052)	(0.074)	(0.045)	(0.121)	(0.108)
Controls	No	No	No	No	No
Observations	109	109	109	45	45

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. 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 A2: 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)
Top Woman AI	0.483***	0.415***	0.483***	0.490***	0.424***
	(0.143)	(0.141)	(0.143)	(0.142)	(0.140)
Top Man No AI	-0.010	-0.017	-0.010	0.002	-0.005
	(0.160)	(0.158)	(0.160)	(0.158)	(0.157)
Top Man AI	0.053	-0.015	0.053	0.035	-0.029
	(0.153)	(0.154)	(0.153)	(0.152)	(0.153)
Low Man AI	-0.712***	-0.729***	-0.712***	-0.714***	-0.730***
	(0.156)	(0.155)	(0.156)	(0.155)	(0.155)
Manager FEs	Yes	Yes	Yes	Yes	Yes
Grade Distribution FEs	No	Yes	No	No	Yes
Gender Manager FEs	No	No	Yes	No	Yes
Order of Candidate FEs	No	No	No	Yes	Yes
Men AI premium (p-value)	0.646	0.986	0.646	0.809	0.862
Observations	2,286	2,286	2,286	2,286	2,286
R^2	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), the gender of the manager who evaluated the candidate (column 3), and whether the candidate was presented first or second (column 4). Additionally, column 5 correspond to the regression including all aforementioned 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 premium of generative AI skills in score for male students. $* \; p < 0.10$, $** \; p < 0.05$, $*** \; p < 0.01$.

Table A3: 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
Men AI premium (p-value)	0.219	0.304	0.621	0.284
Top Woman No AI (mean)	6.227	6.227	6.462	6.462
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 misperceptions over the gender gap in use of AI 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 indicated that male students the technology 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 A2. 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 A4: Exploratory analysis of hiring decisions

	AI candidate selected	Able to Negotiate 5% or more	
	(1)	(2)	(3)
β_0: Constant	1***	0.304***	0.219***
	(0.034)	(0.037)	(0.052)
β_1: AI and No AI presented	-0.490***	0.077*	0.118***
	(0.038)	(0.041)	(0.045)
β_2: AI candidate selected			0.085**
			(0.036)
H_0: $\beta_0 + \beta_1 = 0.5$ (p-value)	0.554	-	-
H_0: $\beta_1 = \beta_2$ (p-value)	-	-	0.461
Observations	867	867	867
R^2	0.16	0.004	0.01

Notes: Column 1 shows a regression with an indicator variable as the dependent variable which takes value 1 if a candidate with AI skills was selected for an interview. The estimated coefficient "AI and No AI presented" is the coefficient for a explanatory variable that takes value 1 if the manager was presented with two candidates in which one has AI skills and the other one has not (contrasting skills) and 0 otherwise. We report the p-value a two sided test with H_0: $\beta_0+\beta_1 = 0.5$, which tests whether the proportion of candidates with AI skills is different than half selected when the manager faces candidates with contrasting AI skills. Columns 2 and 3 use as a dependent variable an indicator variable that takes value 1 if the candidate selected is able to negotiate 5% or more of the salary. In Column 2 we compare how salary negotiations are when the manager observes contrasting AI skills relative to when both candidates have AI skills. Column 3 assess whether candidates with AI skills are able to negotiate a higher salary, by adding as explanatory variable "AI candidate selected", which is an indicator variable that takes value 1 if the candidate selected has AI skills. The estimated coefficient β_2 provides the difference in salary negotiation potential between candidates with and without AI skills when the managers is presented with contrasting skills. We also test the hypothesis that the proportion that are able to negotiate 5% or more of their baseline salary is the same when there was manager observed contrast of skills, relative to when the candidate selected has no AI skills and the manager observed contrast of skills (H_0: $\beta_1 = \beta_2$). The analysis does not include all survey respondents (1143) as a coding error allowed participants not to answer the question while the survey was running, generating 25% of the sample missing. The error was corrected during data collection. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Survey of Students

Figure A11: Page 1. Consent

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

Figure A12: 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

Figure A13: 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

Figure A14: 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

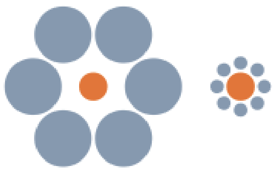
Given the scenario, how likely are you to use ChatGPT during the final exam?

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

Figure A15: 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.

Figure A16: 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

Figure A17: 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.

Figure A18: 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

Figure A19: 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

Figure A20: 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)

Figure A21: 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

Figure A22: 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

Figure A23: Page 11.3 Agree/Disagree

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

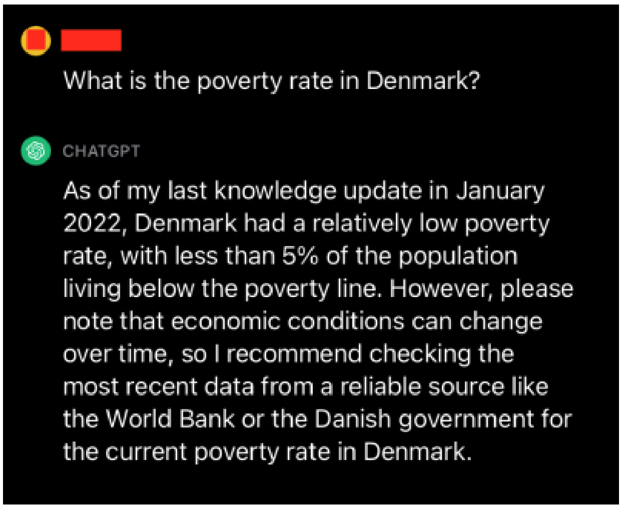
- Completely agree
- Somewhat agree
- Neither agree not disagree
- Somewhat disagree
- Completely disagree

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

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

Figure A24: 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?

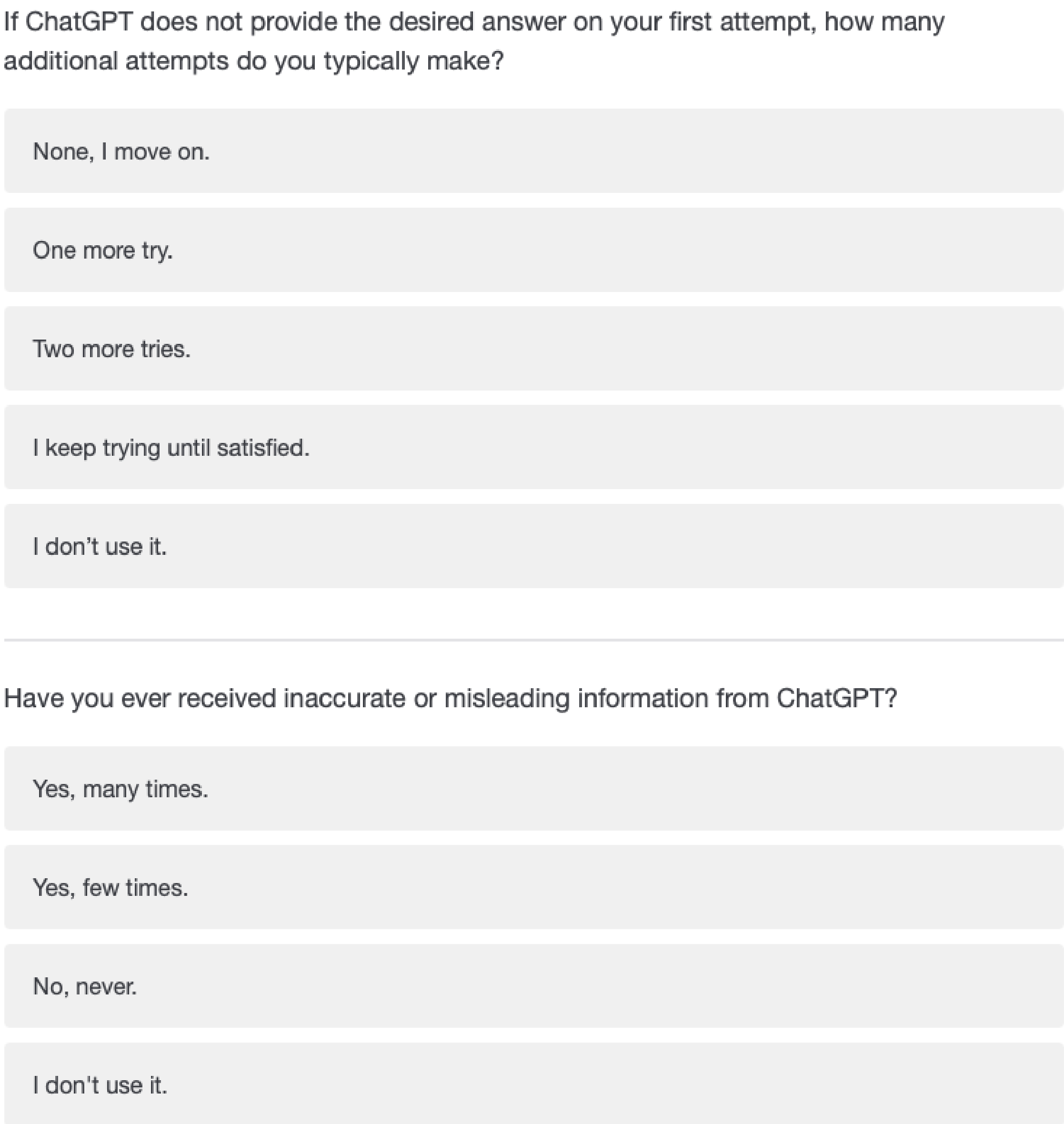
Completely trust

Somewhat trust

Somewhat distrust

Completely distrust

Figure A25: 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.

Figure A26: 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):

D Questionnaire for Survey of Managers

(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 CARD, 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 CARD, 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 card) [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. Martin 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 replacing individual efforts

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 CARD, 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 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)]

Q39 How likely do you think the following scenarios will be in three years: [Options: very likely, Somewhat likely, Neither likely nor unlikely, Somewhat unlikely, Very unlikely]

Q40 When your company hires, a candidate with generative AI skills will be preferred over a similar candidate without generative AI skills

Q41 In my company, an increase in productivity due to the use of generative AI tools will be rewarded with higher wages or bonuses

Q43 Do you think that male and female students use AI tools to the same extent? [Options: Yes, to the same extent (1), No, male students spend more (2), No, female students spend more (3), Don't know (4)]

Q44 Are you an economist from NHH? [Options: Yes, No]

Q45 How long have you been employed at the company you work for today [Options: Less than a year (1), 1-2 years (2), 3-5 years (3), More than five years (4)]

Q46 What is your education? [Options: Primary school level (1), High school level (2), Vocational school level (3), University and college level (4)]

E Prompting skills and text analysis

E.1 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.

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 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

F Manager survey

F.1 Randomization procedure for conjoint experiment

In the conjoint experiment, each manager is presented with 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 profile cards (see Figure A27), 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 profile cards.

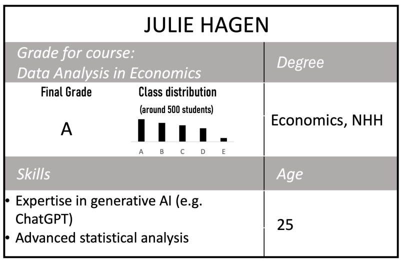
Second, in order to ensure that a manager is likely to face both a candidate with and without generative AI skills, we divide the set of 10 profile cards into two subsets:

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 mostly without AI skills (5 profiles):
    - Top Woman No AI: A and B distribution.
    - Top Man No AI: A and B distribution.
    - Low Man AI: B distribution.

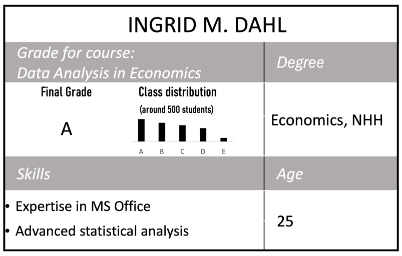
In the experiment, one of the two candidates is drawn from set 1, and the other candidate is drawn from set 2, with equal probability in both. In this way, we ensure that each profile has an equal probability of being presented. Furthermore, a majority of the managers will be presented with a candidate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

Figure A27: Possible profile cards.

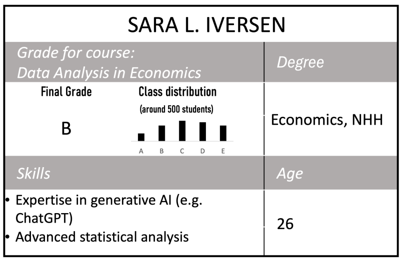
(a) Set 1: Top Woman AI (A)



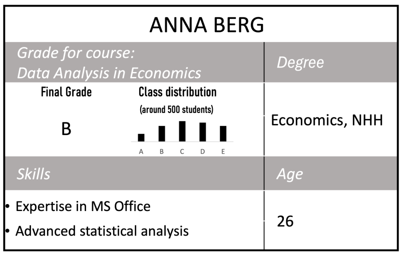
(b) Set 2: Top Woman No AI (A)



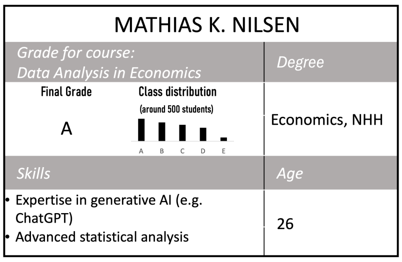
(c) Set 1: Top Woman AI (B)



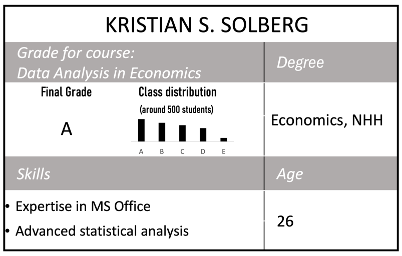
(d) Set 2: Top Woman No AI (B)



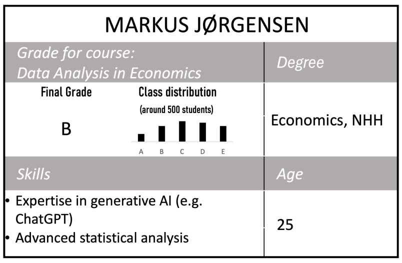
(e) Set 1: Top Man AI (A)



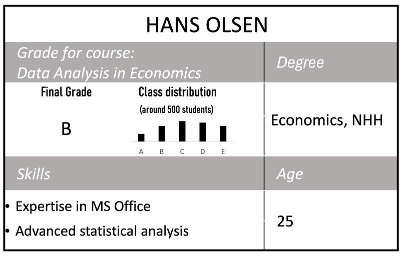
(f) Set 2: Top Man No AI (A)



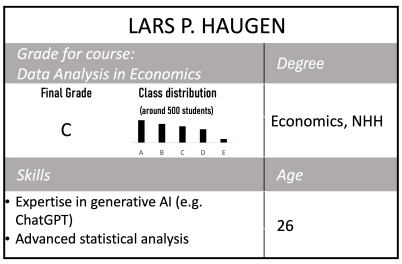
(g) Set 1: Top Man AI (B)



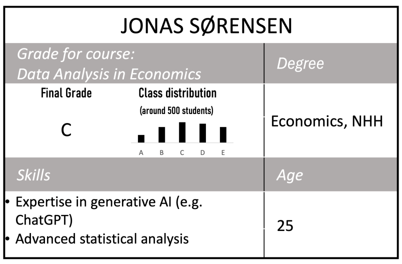
(h) Set 2: Top Man No AI (B)



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



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



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

F.2 Additional results

Managers were also asked who between the two candidates presented would be selected 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).[29] 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 baseline salary the selected candidate would be able to negotiate. We compare the salary negotiation ability of candidates who are in 9% higher negotiation in contrast relative to a base of 30%. Conditional on chosen candidate, 30% can negotiate 5% or more in AI-AI, 33% if no AI chosen (not significant), 42% is AI is chosen (see Table A4).

Figure A7, 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.

F.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 A6 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*

[29]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.

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 A9). Altogether, 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 replacing individual effort."* We found that while 30% agreed and 32% disagreed, most participants neither agreed nor disagreed (see Figure A10a). However, as shown in Figure A10b, 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.

F.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 A10a shows our proxy measure of attitudes towards generative AI, where we observe that 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.