

Hausman, Naomi; Rigbi, Oren; Weisburd, Sarit

Working Paper

Generative AI's Impact on Student Achievement and Implications for Worker Productivity

CESifo Working Paper, No. 11843

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Hausman, Naomi; Rigbi, Oren; Weisburd, Sarit (2025) : Generative AI's Impact on Student Achievement and Implications for Worker Productivity, CESifo Working Paper, No. 11843, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/319211>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Generative AI's Impact on Student Achievement and Implications for Worker Productivity

Naomi Hausman, Oren Rigbi, Sarit Weisburd

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Generative AI's Impact on Student Achievement and Implications for Worker Productivity*

Naomi Hausman[†], Oren Rigbi[‡], and Sarit Weisburd[§]

March 2025

Abstract

Student use of Artificial Intelligence (AI) in higher education is reshaping learning and redefining the skills of future workers. Using student-course data from a top Israeli university, we examine the impact of generative AI tools on academic performance. Comparisons across more and less AI-compatible courses before and after ChatGPT's introduction show that AI availability raises grades, especially for lower-performing students, and compresses the grade distribution, eroding the signal value of grades for employers. Evidence suggests gains in AI-specific human capital but possible losses in traditional human capital, highlighting benefits and costs AI may impose on future workforce productivity.

*We gratefully acknowledge outstanding research assistance from Nataliya Kris and funding from the Asper Center for Entrepreneurship.

[†]The Hebrew University Business School, Jerusalem, Israel, 9190501, naomi.hausman@mail.huji.ac.il

[‡]The Hebrew University Business School, Jerusalem, Israel, 9190501, oren.rigbi@mail.huji.ac.il

[§]The Hebrew University Business School, Jerusalem, Israel 9190501, sarit.weisburd@mail.huji.ac.il

1. Introduction

Artificial intelligence (AI) is changing not only how we work, but also how we learn (Bastani et al., 2024; Nie et al., 2024). Student use of AI in the course of their higher education is redefining the skills and knowledge of future workers. AI tools in education—such as chat bots, personalized learning platforms, automated tutoring, assisted programming, and content development—offer students new ways to both build technical skills and produce assignments. Early exposure to these tools could equip future workers with the skills to use AI efficiently, enhancing their workplace performance, their firms’ productivity, and their own entrepreneurship rates (Babina et al. (2024); Gofman and Jin (2024)). But there is a growing concern that students may become overly dependent on AI, using it as a substitute for essential learning processes, such as problem-solving and critical thinking. This over-reliance could lead to knowledge gaps, ultimately lowering productivity and leaving future workers less prepared.¹ In addition, AI use on take-home exams and assignments could obfuscate underlying ability, reducing the signal value of grades and complicating job matching. The net effect on ultimate workforce productivity of these potential benefits and costs of AI use in universities is unknown.²

We study the effect of AI availability on human capital development and signaling in higher education via early adoption of generative AI tools by university students. We measure the extent to which student performance changes when students are presented with opportunities to employ AI in their graded work. Courses in which assignments and the final exam are completed at home are especially conducive to AI use (“AI-compatible courses”), while those with only in-person work compel students to rely on their human capital alone, without the help of AI (“AI-incompatible courses”). The difference in student performance (and grade distribution) between these two types of courses, after relative to before the introduction of AI tools, estimates the AI effect.

Detailed data on course syllabi and grades for all BA, MA, and MBA courses at one of Israel’s largest universities facilitate tight identification with student and semester fixed effects and careful comparisons between similar courses. While baseline estimates control for unobserved heterogeneity across students and time, the many observable course characteristics in the data enable a propensity score matching strategy in which AI courses are compared to

¹Bastani et al. (2024); Lehmann, Cornelius and Sting (2024); Darvishi et al. (2024) provide evidence on knowledge gaps resulting from AI use.

²Acemoglu (2024) estimates only modest TFP gains due to the complexities of AI use in certain contexts.

similar-on-observables courses that nevertheless used assessments not conducive to AI.³ The data follow approximately 36,000 students in 6,000 courses from September 2018 to June 2024, spanning the introduction of ChatGPT for widespread use in November 2022. Because the data begin three years prior to ChatGPT’s introduction, we can assess the parallel trends assumption underlying our estimation.

Since performance in AI-compatible courses reflects the combination of human capital and AI, a relative increase in performance may be measured even if AI partially or completely substitutes for non-AI related human capital development. To the extent that workforce tasks and university assignments are similarly amenable to AI use, a performance increase foretells corresponding increases in labor market productivity. However, university education tends to be grounded in critical thinking precisely to train students to perform in new situations; students who miss this training due to over-reliance on AI during their studies may be less able to translate their university performance into labor market productivity. We take two approaches to assessing the extent to which AI substitutes for human capital development. First, we measure student rank inconsistencies across more and less AI-compatible courses. Students whose within-course rank is much higher in AI-compatible courses than in others may be unprepared for occupations in which some tasks still require non-AI-related, basic human capital. These rank mismatches provide an estimate of knowledge gaps. Second, we estimate the effects of experience with AI in introductory courses on subsequent performance in both AI-compatible and AI-incompatible advanced courses. While the former is informative on development of AI-specific human capital, in which students learn to use the tool more productively, the latter is informative on the substitution of AI for basic human capital accumulation. Lower performance in AI-incompatible advanced courses, in which students must rely on their own knowledge and skills, suggests AI use in introductory courses likely crowds out some core learning.

We find a significant increase in student grades in both baseline and propensity score matching estimates of about 0.6-1 points in AI-compatible courses relative to others. This change is meaningful relative to the average grade of 86, potentially pushing students over the margin between a B and a B+, for example. Consistent with [Brynjolfsson, Li and Raymond \(2023\)](#), [Noy and Zhang \(2023\)](#), and [Cui et al. \(2024\)](#), the increase is largest for students toward the bottom of the distribution, giving 25th percentile students more than 2.5 additional points in our most controlled specification. Similarly, we find that the introduction of ChatGPT

³Observables include features of faculty (group of departments), degree program (BA/MA), credits, language of instruction, whether the course is mandatory, ex-ante enrollment, and type of course (e.g. lecture/seminar/lab/workshop).

was associated with a 33% drop in the share of students failing AI-compatible courses in the first year, and a 50% reduction in the share receiving a D grade in AI-compatible courses the second year. These effects are driven by non-STEM, non-business, advanced-level courses, in which grades may be especially flexible when AI is used to improve assignment content, and in which the range of potential assignment quality when using AI may be larger.

We provide evidence that the grade distribution is not only shifted right but also compressed, pooling students of a variety of abilities under similar grades and reducing grades' signal value. This compression of the grade distribution has broader implications for the labor market, where the traditional role of grades in differentiating talent and predicting productivity may be diminished, potentially complicating hiring, forcing employers to find other means of assessment, and reducing the efficiency of worker-job matching. This heightened difficulty in assessing student ability coincides with many firms' desire to shift towards labor with AI-related skills without sacrificing on the quality of human capital performing AI and non-AI-related tasks ([Acemoglu et al. \(2022\)](#); [Babina et al. \(2023\)](#)).

Meanwhile, we show that experience with AI can have both positive and negative effects on future productivity, depending on the extent to which future tasks permit the use of AI. While AI training unsurprisingly makes students better AI users, outperforming in advanced AI-compatible courses, it does seem to replace some basic human capital accumulation. Despite their higher performance in AI-compatible advanced courses, students who can use AI in introductory courses perform no better and perhaps worse in advanced courses that are not compatible with AI use in graded work. Further supporting human capital replacement, we illustrate increasing rank inconsistencies for a given student across AI-compatible and AI-incompatible courses. When ChatGPT becomes available, a student's performance in AI-compatible courses becomes less predictive of his performance in AI-incompatible courses, in which he has to rely on his own skills.

Measuring the effect of generative AI tools on human capital development is a fundamentally difficult task for at least two reasons. First, students who choose to use AI for assignments are different than others in ways we can't observe or account for; they may use AI to compensate for lower ability, or they may be the most talented subset who have the skills and foresight to adopt AI most quickly and effectively. Naive estimates of the AI effect that don't account for selection into use may thus be upward or downward biased. Second, researchers generally don't observe how AI tools are used and thus to what extent they may complement or substitute for human capital in production of an output, in practice. When only the output

is measured, it's then difficult to discern the extent to which it reflects human capital, AI, or the interaction of the two, and so inferring underlying human capital accumulation requires some empirical strategy. This paper uses granular data on student performance, along with naturally occurring variation in courses' compatibility with AI use, to identify the effect of AI. It further isolates the effect on components of performance that reflect human capital alone. Since we still don't observe individual students' AI use—but rather the availability of AI and its compatibility with tasks at hand—our estimates can be viewed as analogous to intent-to-treat (ITT) effects. Course compatibility with AI (post ChatGPT roll-out) is a sort of instrument for use, but, since use itself is unobserved, we measure the direct effect of AI availability on performance. As a result, our estimates are *availability* effects and should be viewed as a lower bound on the effect of AI *use* per se among students.⁴ This approach eliminates selection bias, improving upon the confounded comparison of AI users to non-AI users, who are fundamentally different in many ways that also affect their performance.

The question of how generative AI tools affect university student performance and human capital acquisition is highly relevant to the design of universities' educational programming and assessment. Universities must balance the need to train students in modern technological tools that they'll use on the job and the desire to harvest potential human capital-AI complementarities, on one hand, with the risk that under-monitored students facing outdated assignments may knowingly or unknowingly over-rely on these tools to substitute for important learning processes. While avoiding training students in AI augmentation would limit their AI-related human capital accumulation and ultimate firm productivity (Babina et al. (2024)), allowing unfettered use would limit their traditional human capital development. Understanding the extent of the tradeoff and the location of the status quo on this spectrum is a critical starting point for forming modified curricula and education policy.

AI adoption among students in university today has implications for the productivity of the workforce tomorrow. To the extent grades become less useful signals of human capital because AI facilitates pooling, employers will have to develop alternative means of screening for both basic and AI-related human capital. As employers investing in AI shift towards STEM and IT-skilled workers (Acemoglu et al. (2022); Babina et al. (2023)), they may face increasing difficulty in identifying the graduates with the desired combination of skills, potentially overestimating the basic human capital of workers with significant AI experience. Even if the assessment challenge were to be solved, the underlying human capital—both

⁴The latter effect is analogous to a treatment-on-the-treated (TOT) effect, which, under some assumptions, is equal to the ITT scaled up by the probability of AI use given being in an AI-compatible course post ChatGPT roll-out.

AI and non-AI related—is clear to have real effects on students’ subsequent labor market productivity. AI experience of workers affects firm productivity (Babina et al. (2024)), and reduced AI knowledge acquisition of students shrinks their entrepreneurship rates (Gofman and Jin (2024)). Encouraging AI use, including among students, could spur innovation by allowing for faster knowledge accumulation (Cockburn, Henderson and Stern (2018)). Meanwhile, excessive focus on AI-related skill development at the expense of basic, non-AI-related human capital acquisition could generate a productivity crisis given that most job tasks still cannot be completed solely using AI (Eloundou et al. (2023); Georgieva (2024); Lin and Parker (2025)). AI use and investment require human expertise to complement big data and computing technology.

The paper is organized as follows. Section 2 discusses relevant literature. Section 3 delineates the empirical strategy, while Section 4 describes the data used to implement it. Section 5 presents results. Section 6 discusses implications and concludes.

2. AI Tools in Education and the Labor Force

The growing literature on the adoption of generative AI tools has focused primarily on how they enhance the productivity of workers across industries, with a particular emphasis on tasks that align closely with the technology’s strengths (e.g. Agrawal, Gans and Goldfarb (2019); Brynjolfsson, Li and Raymond (2023); Dell’Acqua et al. (2023); Demirci, Hannane and Zhu (2025); Acemoglu (2024)). Research on generative AI highlights its relatively fast adoption rate and heterogeneous effects across types of workers. Already in 2024, 39% of the US population between the ages of 18-64 used generative AI tools (Bick, Blandin and Deming (2024)). Higher quality firms and workers tend to benefit disproportionately from new technologies, in general (Acemoglu et al., 2023; Acemoglu, Lelarge and Restrepo, 2020; Koch, Manuylov and Smolka, 2021). Evidence on AI adoption, specifically, points in both directions, depending on context. (Otis et al., 2023) find that GPT-4 improves outcomes for high performing entrepreneurs in Kenya but worsens outcomes for low performers. But substantial research finds that AI tools have an equalizing effect, boosting worker productivity most among less skilled or experienced employees (Brynjolfsson, Li and Raymond, 2023; Noy and Zhang, 2023; Cui et al., 2024).

Not all tasks exhibit a positive impact of generative AI on productivity. Dell’Acqua et al. (2023) find that AI tools improve both productivity and output quality for workers when

applied to tasks well-suited to AI’s capabilities. However, they also report worse outcomes on tasks requiring nuanced judgment, such as synthesizing insights from both structured and unstructured inputs. AI tools lead to more homogenized ideas, they show, reducing variability in participants’ outputs.

These complexities align with broader macroeconomic projections. [Acemoglu \(2024\)](#) estimates modest total factor productivity gains from AI—less than 0.53% over the next decade—due to diminishing returns on complex, context-dependent tasks. These findings suggest a nuanced role for AI in productivity enhancement and highlight the importance of understanding heterogeneity in adoption effects.

Even more complex is understanding the productivity effects of AI tools in learning environments. Productivity in an educational task, of course, is not just about the quality of the output but also, especially, about human capital gained in the process of producing the output. Studies of specialized AI interfaces that encourage learning through critical thinking rather than allowing solution-seeking report an improvement in exam performance, suggesting human capital gains ([Bastani et al., 2024](#); [Nie et al., 2024](#)). These findings contrast with others on more general use of AI by students, which indicate performance improvements on tasks for which AI tools are available but worse outcomes when the tool is no longer available, even relative to students who never had AI access ([Bastani et al., 2024](#); [Lehmann, Cornelius and Sting, 2024](#); [Darvishi et al., 2024](#)). Such studies support the concern that students substitute typical AI tools for key learning processes.

This paper contributes to this literature by leveraging a novel empirical strategy and large, comprehensive dataset that tracks student outcomes both before and after the introduction of ChatGPT across a range of courses with varying levels of suitability to generative AI tools. While some course grades are determined mostly by in-class exams and are thus difficult to affect with AI, others are determined in large part by assignments that can be improved substantially with AI. By analyzing grades in these more versus less AI-compatible courses, we can disentangle the productivity benefit of ChatGPT from concerns that it can serve as a crutch, leading to knowledge gaps and worse performance when the tool is absent. We are further able to examine heterogeneity in effects across students and course types, contributing to the evidence on context. Measuring distributional effects extends our understanding of whether ChatGPT exacerbates existing disparities in skill acquisition and performance or serves as an equalizing force in the educational context. Because we follow students’ paths through university courses, we can estimate AI experience effects, which

may be especially relevant in considering the effects of educational AI policies on likely labor market performance. Together, these analyses contribute to our understanding of how generative AI will alter the workers entering the labor force and their productivity on tasks more and less amenable to AI use.

3. Estimation of AI Availability and Experience Effects

3.1. Effect of AI Availability on Grades and the Grade Distribution

An ideal experiment to measure the effect of student AI adoption on performance in university would be to randomize its availability across students or courses and measure differences across groups in grade outcomes. Randomizing across courses has the advantages of facilitating (1) controlled, within-student performance comparisons across courses, and (2) analysis of course-level outcomes like grade distribution.

While we are unable to randomize AI availability, we use two forms of naturally occurring variation in AI availability to approximate the ideal experiment. First, we use variation over time in AI availability stemming from the roll-out of ChatGPT. ChatGPT was publicly launched by OpenAI on November 30, 2022, allowing free use of the tool and leading to a surge in popularity and public interest. This release serves as an “event” that fundamentally changed the availability and salience of generative AI tools in the general population, including among students.⁵

Second, we use variation across university courses in the extent to which graded work was amenable to AI use. Courses whose final grade was largely or fully determined by in-person work, like in-class exams or lab work, are defined as AI-incompatible, or “control” courses. Courses whose final grade was largely or fully determined by home assignments, such as problem sets, essays, take-home exams, or projects, are defined as AI-compatible, or “treatment” courses.⁶ Any baseline difference in grades between these two types of courses is captured by comparing them before ChatGPT was released. A change in that gap after the roll-out of ChatGPT will reflect the AI effect—the extent to which students were able to alter their grades differentially in AI-compatible relative to AI-incompatible courses because

⁵Our survey of students indicates that, indeed, students did not use AI for their schoolwork before this date. See appendix [Figure A1](#) for the distribution of adoption rates we observe in this survey.

⁶Specifically, we define a continuous variable for AI-incompatibility of a course from 0-100%. Control courses are defined to be those above 90% AI-incompatible. Treatment courses are defined as those below 60% AI-incompatible. Courses between 60% and 90% AI-incompatible are omitted from the analysis.

of the new availability of AI.

We implement this difference-in-differences (DID) strategy in a dynamic regression, estimating treatment effects by year relative to the year before the ChatGPT roll-out. Formally, we regress course grades g_{ict} for student i in course c completed in year t on the interaction of an indicator for being an AI-compatible course in year t , $AI-comp_{ct}$ and a series of year indicators, D_t , excluding the 2021-2022 academic year prior to the event.⁷

$$g_{ict} = \sum_{\substack{t=2019 \\ t \neq 2022}}^{2024} \delta_t D_t \times AI-Comp_{ct} + \sum_{\substack{t=2019 \\ t \neq 2022}}^{2024} \gamma_t D_t + \beta AI-Comp_{ct} + \theta' X_{ct} + \rho_i + \epsilon_{ict} \quad (1)$$

We control for time effects, $\sum_{\substack{t=2019 \\ t \neq 2022}}^{2024} \gamma_t D_t$, whether the course is AI-compatible as defined by its components, $AI-comp_{ct}$, and a vector of course-year characteristics, X_{ct} , including indicators for class size quartile, department, number of credits, class type (e.g. lecture/lecture with sections/lab course), mandatory course, language of instruction, and degree program (B.A./M.A.).⁸ ϵ_{ict} is an error term. Student fixed effects, ρ_i , ensure that treatment effects δ_t are estimated from within-student variation, comparing a student's performance in AI-compatible courses to *her own* performance in AI-incompatible courses, after relative to before ChatGPT's roll-out.

The coefficients of interest, δ_t , are expected to be positive in 2023 and 2024 if AI improves student performance relative to the excluded year, 2021-2022, which was the year before ChatGPT became widely available. Meanwhile, the δ_t coefficients for the years before the event should be indistinguishable from zero if the treatment and control groups followed parallel trends in grades prior to ChatGPT. Such evidence would lend confidence to the parallel trends assumption required for our DID estimate to be interpreted as causal.⁹ Having three years of data prior to the event to enable to assess pre-treatment trends is thus a significant advantage.

Because we measure AI availability rather than AI use, the estimated treatment effects, $\hat{\delta}_{2023}$ and $\hat{\delta}_{2024}$ should be interpreted as intent-to-treat (ITT) effects. These would be identical

⁷We refer to academic years by the year in the spring semester. For example, the 2021-2022 academic year is referred to as 2022 for brevity.

⁸In fact, there are 19 class types in the data and for which we include fixed effects.

⁹For the DID effect to be interpreted as causal, the assumption that the treatment and control groups would have followed parallel trends in the absence of treatment must hold. While this assumption is fundamentally untestable, a finding of parallel trends prior to treatment is strong evidence in its favor.

to AI *use* if everyone for whom AI was available in fact used it for their coursework. To the extent that availability and use differ, these estimates provide a lower bound on the use effect.¹⁰

While the basic DID strategy estimates a causal effect if the parallel trends assumption holds, there may still be differences between AI-compatible and AI-incompatible courses that affect outcomes and are not fully accounted for. We can potentially improve the estimates by comparing a subset of treated and control courses that are more similar on pre-treatment observables. We use propensity score matching (PSM) to implement this more controlled comparison. For each course, we estimate the propensity to be AI-compatible (treated) as a function of ex-ante course observables: class size, course type (lecture/lab/seminar), mandatory, language of instruction, degree program (B.A./M.A.) and department. We use a nearest neighbor 1 (NN-1) matching algorithm to match AI-compatible courses to AI-incompatible courses in the same faculty (e.g. social sciences) with similar propensity scores, generating a new, “PSM sample” containing only matched courses in the common support. Each matched pair gets a unique match group ID. To generate PSM estimates, we estimate a version of Equation 1 including a propensity score control and match group fixed effects to ensure that treatment effects are estimated from comparisons of each treated course to its own matched control. The regression is still estimated at the student-course-year level, which allows us to control for unobserved student-level heterogeneity using student fixed effects.

If AI affects student grades heterogeneously—as one might expect if students vary in their speed of adoption, extent of use, or proficiency—then AI may also affect the grade distribution. We first estimate distributional effects at the student-course-year level using Equation 1 with the dependent variable being an indicator for the student’s grade being within a certain range, e.g. $I_{g \in [0,59]}$. The equation is estimated multiple times for the grade ranges 0-59, 60-69, 70-79, 80-89, and 90-100. A negative $\hat{\delta}_{2023}$ or $\hat{\delta}_{2024}$ when the dependent variable is $I_{g \in [0,59]}$, for example, would indicate AI causes a reduction in failure rate.

The second way we estimate distributional effects is to aggregate the data to the course-year level and estimate regressions predicting the 25th, 50th, and 75th percentile grades in an analogous DID framework to estimate causal effects. These regressions allow us to measure how the bottom, middle and top of the distribution are affected by AI.

¹⁰Roughly 30% of students in our survey reported adopting generative AI tools in the 2022-2023 academic year. This percentage increases to almost 80% during the 2023-2024 academic year (see Appendix Table A1).

We estimate the following regression for grades g at percentiles $p = 25$, $p = 50$, and $p = 75$ of the grade distribution within each course-year:

$$g_{ct}^p = \sum_{\substack{t=2019 \\ t \neq 2022}}^{2024} \alpha_t D_t \times \text{AI-Comp}_{ct} + \sum_{\substack{t=2019 \\ t \neq 2022}}^{2024} \rho_t D_t + \omega \text{AI-Comp}_{ct} + \Omega' X_{ct} + \epsilon_{ct} \quad (2)$$

α_t will be positive in 2023 and 2024 for p25 if AI raises grades at the bottom of the distribution. A larger grade increase at the bottom of the distribution than at the top would suggest a compression in which students receive more similar grades as a result of AI.

3.2. Basic versus AI-specific Human Capital Development

One concern with AI use among university students is that it may replace important learning processes, leading to under-development of key analytical skills, including problem solving and critical thinking. On the other hand, practice using AI may be necessary for productive use of the tool in the labor market. In the new AI regime, in short, students need to develop both basic and AI-specific human capital to equip them for a variety of future tasks for which AI tools may or may not be available.

We conduct two analyses to assess the extent to which AI use among students may affect basic versus AI-specific human capital development. The first measures the consistency of student within-course rank across AI-compatible and AI-incompatible courses within a given year. If AI differentially affects performance in AI-compatible and AI-incompatible courses, then rank in AI-compatible courses should be more predictive of rank in AI-incompatible courses before the roll-out of ChatGPT. A drop in rank consistency across these two types of courses drops once AI is widely available would be consistent with the development of AI-specific human capital at the expense of the basic human capital that drives performance in AI-incompatible courses.

To test this hypothesis, we run the following regression year by year, on data collapsed to the student-year level, using a student's average performance percentile in AI-incompatible courses to predict his average performance percentile in AI-compatible courses in the same year:

$$pctile_{it}^{AI-comp} = \beta_0 + \beta_1 pctile_{it}^{AI-incomp} + \epsilon_{it} \quad (3)$$

This specification implicitly includes student fixed effects by comparing performance percentile across course types within student. A steeper β_1 (closer to 1) indicates that performance in AI-incompatible courses is more predictive of performance in AI-compatible courses than if the slope is flatter. A flattening of the slope after the roll-out of ChatGPT would support the notion that students are using AI at the expense of basic human capital accumulation. Because these are percentile ranks, not grades, a simple relative increase in grades in AI-compatible versus AI-incompatible courses would not produce a flattening of the percentile-percentile slope after the event.

To get more directly at the effect of experience using AI on both AI-specific and basic human capital, we conduct a second analysis that leverages the randomness in the timing with which students began their degrees. While some students took their first year introductory courses in the academic year before ChatGPT was available, others did so in the following year, 2022-2023, when they could already use AI for assignments in their AI-compatible introductory courses. The two sets of students are plausibly comparable, as they began university in adjacent years and could not have anticipated the roll-out of ChatGPT.

We thus define a new set of treatment and control groups. Students identified to be first year students in 2022-2023 and who took at least one AI-compatible introductory course are in the treatment group. Students identified to be first year students in 2021-2022 and who took at least one AI-compatible introductory course are in the control group. We then compare these students' performance outcomes in their second year along two dimensions. First, by comparing their grades in their second year AI-compatible courses, we test the AI experience effect on AI-specific human capital: do students with AI experience from their introductory courses do better in their AI-compatible advanced courses? Second, by comparing their grades in their second year AI-incompatible courses, we test the AI experience effect on basic human capital accumulation: do students who could rely on AI in their introductory courses miss out on basic human capital accumulation and as a result do *worse* in advanced courses in which AI cannot help them in their graded work?

We estimate the following two DID regressions, which are in principle identical except for the dependent variable and the sample of second-year courses included:

$$g_{ict}^{AI-comp} = \alpha_1 I_{ct}^{2ndYear} * I_i^{AIexper} + \theta'_1 X_{ct} + \rho_{1i} + \varepsilon_{ict} \quad (4a)$$

$$g_{ict}^{AI-incomp} = \alpha_2 I_{ct}^{2ndYear} * I_i^{AIexper} + \theta'_2 X_{ct} + \rho_{2i} + \varepsilon_{ict} \quad (4b)$$

$I_{ct}^{2ndYear}$ is equal to 1 if the course is an advanced course being taken in the second year and 0 otherwise, and $I_i^{AIexper}$ is equal to 1 if the student was treated (took at least one AI-compatible introductory course during his first year) and 0 otherwise. X_{ct} is defined as before and includes a control for whether the course is an advanced course. Similarly, the main effect of being treated is absorbed in the individual fixed effects, ρ_{1i} and ρ_{2i} . Intuitively, these fixed effects control for any underlying differences between students who began their studies in 2021-2022 versus students who began their studies in 2022-2023.

If AI substitutes for basic human capital accumulation, then we should see a negative effect of AI experience on performance in AI-incompatible advanced courses, or $\alpha_2 < 0$ in which students must rely on their own knowledge and skills. Meanwhile, if experience using AI increases AI-specific human capital and productivity in AI-compatible tasks, we should see a positive effect of experience on grades in advanced AI-compatible courses ($\alpha_1 > 0$).

We estimate these equations using data on the 2021-2022, 2022-2023, and 2023-2024 academic years. Control students are observed in the first two years, while treated students are observed in the last two years. Equation 4a is estimated using data on students' AI-compatible course grades in their first year and their AI-compatible course grades in their second year. Equation 4b is estimated using data on students' AI-compatible course grades in their first year and their AI-incompatible course grades in their second year.

4. Data

4.1. Administrative Grade and Course Data

The primary data source for this paper is administrative student record and course information from a large research university in Israel over the 6 academic years from 2018-2019 to 2023-2024. The Israeli academic year begins in late October, such that students had access to ChatGPT for all but the first month of the 2022-2023 school year—the first treatment year. The data come in two main parts, one at the student-course-semester level and one at

the course-semester level.

Student records include all courses a student took with dates, grades from each attempt at the final exam, and the final course grade (which may include non-exam components).¹¹ Student demographics are included— including gender, birth year, primary department of study, immigration year, and ethnic/religious group membership—but the immigration and ethnicity variables are sparsely populated.¹² All of our estimates in student-level data contain student fixed effects or differencing such that time-invariant demographic would be absorbed in any case.

Course records include the course format (lecture/ lecture and section/ lab/ seminar), whether the course is mandatory, professor’s name, department, number of credits, language of instruction, and degree program (B.A./M.A.). Features of courses can vary over time, and we observe a course’s features each time it is taught. Course-semester data is easily matched to student-course-semester data by course number and semester.

4.2. Syllabus Data

Data from course syllabi are collected from the university’s online course catalog. While substantial course information is available in principle, we use this source for its detailed decomposition of the final course grade into components. The syllabi detail the percent of each grade that is determined by home assignments, in-class exams, home exams, presentations, class participation, lab work, and so on. Syllabi are binding; professors are not allowed to change the published grade allocation or nature of assignments.

We use these breakdowns to assign course instances to the treatment or control group based on the fraction of the grade that is determined by AI-incompatible work: in-class exams and lab work. The histogram in Figure 1 presents the distribution of courses by AI-incompatible grade share and indicates the definitions of treatment and control groups. One mass of courses, at 0, has grades not determined at all by AI-incompatible work. Another mass sits at 100% AI-incompatible work, with additional density just below 100%. Guided by this histogram and the logical implications of having much of a course’s grade determined by AI-compatible assignments, we define treated courses as those for which 0-60% of the grade (inclusive) is determined by AI-incompatible work. Control courses are defined as those

¹¹Students are typically entitled to take the final exam twice and use the highest grade.

¹²Ethnic and religious groups include Arab, Bedouin, Druze, Ethiopian, immigrant, and ultra-Orthodox Jew.

for which $\geq 90\%$ of the grade is determined by AI-incompatible work, like in-class exams. Courses between 60% and 90% AI-incompatible are excluded.¹³

4.3. Survey Data

To support the implications of our findings for the the labor market, and to provide evidence on student AI adoption, we conduct a survey of business school students at the same large, Israeli research university for which we have administrative data. The survey asks about timing and extent of students' adoption of AI for use on their graded assignments. It asks for their assessment of AI availability in their future labor market work, and the extent to which they think their current AI use for their studies will make them more productive in the labor market or replace basic learning they otherwise would have done (in non-mutually-exclusive questions). And it collects demographic information on the students that we had only sparsely in the administrative data—including immigration year (if applicable), ethnicity, and English/Hebrew preference—as well as basic demographics such as age, gender, and fields of study. We use these data separately from the administrative data to support the main analysis and offer insight into its implications.

The survey questionnaire is presented in Appendix A. 91 student responses were received and analyzed in the first wave.

4.4. Main Data for Analysis

Our two main data sets for analysis are at the student-course-semester level and the course-semester level for analyzing grade effects and grade distribution effects, respectively. Table 1 summarizes the data along several dimensions, comparing treatment and control groups before treatment. While course characteristics differ somewhat across groups, student profiles are similar, suggesting that confounds due to pre-existing differences in students taking these courses are less likely.

¹³As one can see from the histogram, very few courses exist in this range in any case. In the figure, the exclusion range is that between the two vertical dotted lines.

5. Results

5.1. AI Effects on Student Performance

Estimates from Equation 1 are presented in Table 2 and as a visual event study in Figure 2. Points in the figure are coefficient estimates $\hat{\delta}_t$ for each academic year from 2018-2019 to 2023-2024, excluding 2021-2022, which is the base year. Error bars are 95% confidence intervals. Effects are indistinguishable from zero in the years leading up to the roll-out of ChatGPT, indicating that students in AI-compatible and AI-incompatible courses were on parallel grade trends before treatment. This finding importantly supports the parallel trends assumption required for a causal interpretation of the DID estimates. After the roll-out of ChatGPT, indicated by the vertical dotted line, grades increase differentially in AI-compatible courses relative to AI-incompatible courses. The increase is significantly different from zero in both 2022-2023 and 2023-2024, increasing slightly over time.

Table 2 presents the magnitude of these effects in several alternative specifications. Column 1 estimates the baseline DID from Equation 1 on the full sample, while column 2 estimates the same specification on the reduced, matched sample for comparability to the PSM estimates in columns 3-5. The estimates in the baseline DID in the full sample indicate a 0.97 point grade increase in 2022-2023, the first year in which AI was available, and an even larger 1.48 point increase in 2023-2024. In both years, the increase is significant at the 1% level. Estimates of the same specification are slightly smaller but still significant at the 5% level in the reduced, matched sample (column 2). Figure 3 graphs the propensity score densities of treated and control courses in the matched sample, indicating their strong similarity and supporting the comparability of AI-compatible and AI-incompatible courses in the reduced sample. The comparability of courses in the reduced sample, along with evidence of parallel pre-treatment trends, supports the validity of the estimates in column 2.

The most restrictive PSM estimates in column 5, in which each treated course is compared to only one matched control, still show positive and significant effects in 2022-2023 of 0.63 points and similarly sized, positive but insignificant effects in the second year after treatment.

Grades are measured on the standard 0-100 scale, such that a 1-1.5 point increase is meaningful; AI availability can push a B+ student over the threshold of the A range. Recall that, because the estimates measure the effect of AI availability, rather than AI use per se, they can be interpreted as intent-to-treat effects and provide a lower bound on the effect of use. AI use effects are thus even larger.

5.1.1. *Heterogeneity*

One might imagine that there is substantial heterogeneity across students, task types, and contexts in students’ ability to leverage AI to improve performance. While our data don’t include complete information on race, ethnicity, or religion, they contain rich detail on courses. We estimate heterogeneity in effects along several dimensions to understand whether certain types of students seem to benefit the most and which contexts facilitate grade improvement via AI. Estimates are presented in Table 3.

Gender: Figure 4 presents the event study estimates separately for men and women, for whom effects are nearly identical. This result leads us to update our prior that men, who tend to be earlier technology adopters ((Nie et al., 2024; Bick, Blandin and Deming, 2024)), would experience larger grade effects.

Student Age: The estimates suggest that grade effects are driven more by the young, which is consistent with the notion that younger student are quicker technology adopters and perhaps more effective users. However, it’s also entirely possible that older students simply exercise more restraint in using AI tools for graded work due to an understanding that over-reliance on AI could replace important basic human capital accumulation.

Course Level: Students appear more able to increase their grades using AI in advanced courses, perhaps because the grade structure in these courses is more flexible, because there is more potential variance in assignment quality, or because students simply have had more experience using AI by the time they get to advanced courses. We explicitly test the effect of experience with AI and discuss the results below, in Section 5.3.

Field: The grade effect is driven by non-STEM, non-business fields of study, in which coefficients are positive and significant in both post-treatment years. Coefficients in STEM fields and business are indistinguishable from zero. It may be that STEM and business have more objective criteria for evaluation (perhaps similarly to introductory courses in many fields) such that there is less room for AI to push grades up.

Instructor Seniority: Junior professors may be more likely to update their assignments for the AI regime—making it harder for students to improve grades with AI—if they are heavier AI users themselves or if they are simply more likely than Senior Professors to update their courses at all. However, effects are very similar for courses taught by junior and senior professors.

Course Size: The AI effect is slightly larger in smaller courses than large ones, though it is positive and significant in both. One possible reason for this difference is that smaller courses that are AI-compatible are more likely to have *all* of the course’s graded work be AI-compatible. Figure 5 indicates a larger density at 0% AI-incompatible work in below median course sizes. Such a phenomenon could occur if small courses are more likely to assign essays, for example, which are difficult to grade for the many students in a large course.

5.2. Grade Distribution Effects and Signal Value

Heterogeneous effects of AI across students imply that not only the mean but also the distribution of grades may be affected. Figure 6 depicts estimates of a series of linear probability models predicting the probability of receiving a grade in the given range.¹⁴ We graph treatment effects separately for the two years after treatment. The figure indicates that AI availability reduces the probability of failing a course (grade of 0-59) by about 1 percentage point off a 3% base failure rate, which is approximately a 30% reduction in failure. The probability of scoring from 60-69 is also significantly reduced, while the probability of scoring from 90-100 is significantly increased.

Table 4 alternatively measures distributional effects with course-semester level estimates of Equation 2, in which the dependent variable is the grade at a given percentile of the distribution. We present estimates for the 25th, 50th, and 75th percentiles, all using the reduced, matched sample to improve comparability of AI-compatible and AI-incompatible courses. Estimates for both the baseline DID and PSM DID suggest that grade improvements are not uniform across the grade distribution. Rather, AI disproportionately helps students toward the bottom of the distribution (at p25) to improve their grades. Estimates from the baseline DID suggest significant increases of 2-3 points at the bottom of the distribution (p25) and increases of just over 1 point at the median (p50), though this effect is significant only in 2022-2023 (Table 4, columns 1 and 2).¹⁵ This heterogeneity across students becomes perhaps more pronounced in the more restrictive PSM specifications in columns 4-6. While we estimate a 2.62 point and significant increase at p25 in 2022-2023, we measure no significant effects at p50 or p75.

These results suggest that AI may disproportionately help the weaker students in terms of

¹⁴We estimate Equation 1 but with a binary dependent variable equalling 1 if the grade is in the given range.

¹⁵The point estimate in 2023-2024 is very similar, but the standard error is larger.

the size of their grade increase. At the same time, it moves the distribution to the right, increasing the probability of the highest grades (Figure 6).

This compression of the distribution is likely to have negative effects on employers’ ability to infer student ability from their grades, as grades’ signal value is reduced. Employers may have to develop alternative means of screening, which could be costly. In the meantime, job match quality and on-the-job productivity may be reduced to the extent that employers infer ability imperfectly in the transition.

5.3. Effects on Basic and AI-Specific Human Capital

A key issue in understanding the AI effect on student performance is whether the grade increase reflects a general increase in productivity that will be applicable across contexts, or whether over-reliance on AI during their studies leads students to miss out on key learning processes that will leave them less productive in certain situations. We thus distinguish between two types of human capital that students can develop in university in the AI regime. AI-specific human capital refers to skills that complement AI use and can increase productivity in situations in which AI is available. Basic, or traditional human capital refers to skills that can produce output when AI is not available. The analysis in this section aims to distinguish between the AI availability effects on each of these types of skill.

To do so, we estimate Equations 4a and 4b, which measure the effect of AI experience in introductory courses on grades in advanced AI-compatible and AI-incompatible courses, respectively. Students that enter the university and take their introductory classes in 2022-2023, the first year in which ChatGPT is widely available, have the opportunity to gain experience with AI for their introductory coursework, while students who entered a year earlier did not. Under the assumption that the timing with which students enter university is plausibly random for students in adjacent years, these estimates can be interpreted as causal. We find consistent positive and significant effects of AI experience on advanced, AI-compatible course grades (Table 5, columns 1-3). The magnitudes range from 0.93-1.2 points, depending on the specification. Meanwhile, AI experience effects on grades in advanced, AI-incompatible courses are consistently negative, though standard errors are large and the estimates are significant only in one of three specifications (Table 5, columns 4-6).

These results suggest first that experience using AI matters. It appears to facilitate AI-specific human capital accumulation that leads to better student performance in future

contexts in which AI is available. In other words, practice helps students learn how to use AI well. Second, however, while the finding isn't statistically significant in all specifications, it appears that traditional human capital development may suffer when students have the opportunity to use AI in their introductory courses. When AI is not available later on, they do no better and probably worse than similar peers who completed their introductory courses without AI.

These findings support the concern that student use of AI does not come without costs. While students experience an average increase in grades as a result of AI, as discussed in Section 5.1, the average masks a bifurcated effect across future contexts, with positive effects only for contexts in which students can continue to apply AI tools. Underdevelopment of basic human capital appears to leave them less prepared for what are still many tasks and situations that are not conducive to AI.

6. Conclusion

The rapid emergence of generative AI tools is transforming how we work. Yet access to these tools during university studies—a critical period for skill development—raises important questions about how AI will shape the skill set and productivity of the future workforce. This paper provides early causal evidence on how AI availability in higher education affects student performance. Distinguishing between basic and AI-specific human capital development, it shows that the average increase in grades as a result of AI masks a bifurcated effect depending on whether AI is available for future tasks. Relying more on AI for studies unsurprisingly improves performance on future tasks with AI, but it leaves students less prepared with the type of basic human capital that drives performance when AI is absent.

We document three main findings. First, the introduction of generative AI tools results in a significant 0.6-1 point increase in student grades in AI-compatible courses, with the largest effect among lower-performing students (those at the 25th percentile of the grade distribution). Failure rates in these courses decrease by 37% following the roll-out of AI tools, suggesting substantial benefits for the weakest students. However, these effects may not reflect underlying learning gains.

Second, this upward shift in grades comes alongside a compression of the grade distribution, reducing the accuracy of grades in distinguishing between students of different abilities. This erosion of the signaling value of grades risks decreasing the quality of worker-job matches.

Third, experience with generative AI tools has both benefits and costs for skill development. Students with early AI exposure in introductory courses perform better in subsequent AI-compatible courses, reflecting gains in AI-specific human capital. However, in advanced courses where AI cannot be used, these same students (if anything) perform worse, suggesting that AI may substitute for the development of foundational skills. These results demonstrate both the potential benefits and pitfalls of students using AI during the learning process. On one hand, they gain experience using the tool for tasks where the tool is relevant. However, over-reliance on the tool during the learning process, may undermine student preparedness in contexts where AI is unavailable or insufficient for the task they need to complete.

These findings carry important implications for both universities and employers. Universities must rethink how they design coursework and assessments to ensure students develop both AI-specific and traditional human capital. Assessments should strike a balance between encouraging AI literacy by nudging students to use it intelligently while preserving opportunities to measure basic knowledge and skills in the absence of AI.

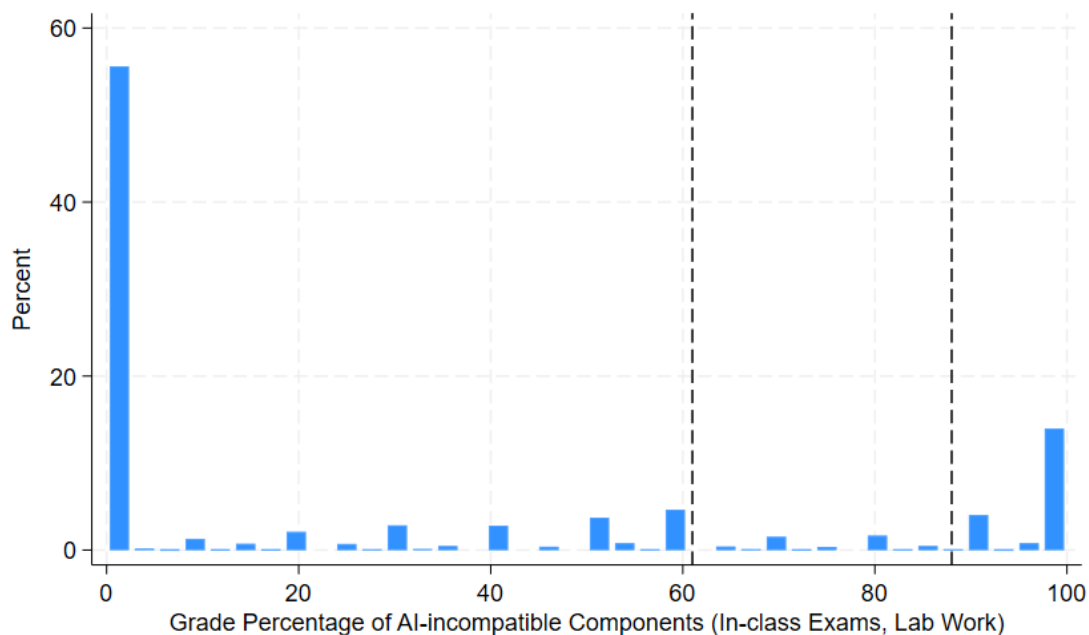
For employers, the diminishing signal value of grades may require a shift towards alternative evaluation tools and/or an attuned and informed reading of grades across courses with varying levels of AI compatibility. As students rapidly adopt these tools, educators and firms must adapt both teaching and evaluation of student skill. At least in the near term, when most job tasks still cannot be completed with AI alone ([Eloundou et al. \(2023\)](#); [Georgieva \(2024\)](#); [Lin and Parker \(2025\)](#)), future workers must be equipped with both AI-related and traditional skills in order to maximize productivity.

References

- Acemoglu, Daron.** 2024. “The simple macroeconomics of AI*.” *Economic Policy*, eiae042.
- Acemoglu, Daron, Claire Lelarge, and Pascual Restrepo.** 2020. “Competing with robots: Firm-level evidence from France.” Vol. 110, 383–388, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.
- Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo.** 2022. “Artificial Intelligence and Jobs: Evidence from Online Vacancies.” *Journal of Labor Economics*, 40(S1): S293–S340.
- Acemoglu, Daron, Gary Anderson, David Beede, Catherine Buffington, Eric Childress, Emin Dinlersoz, Lucia Foster, Nathan Goldschlag, John Haltiwanger, Zachary Kroff, et al.** 2023. “Advanced Technology Adoption: Selection or Causal Effects?” Vol. 113, 210–214, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.
- Agrawal, Ajay, Joshua S Gans, and Avi Goldfarb.** 2019. “Artificial intelligence: the ambiguous labor market impact of automating prediction.” *Journal of Economic Perspectives*, 33(2): 31–50.
- Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson.** 2024. “Artificial intelligence, firm growth, and product innovation.” *Journal of Financial Economics*, 151: 103745.
- Babina, Tania, Anastassia Fedyk, Alex X He, and James Hodson.** 2023. “Firm investments in artificial intelligence technologies and changes in workforce composition.” National Bureau of Economic Research Working Paper no. 31325, National Bureau of Economic Research.
- Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Ozge Kabakcı, and Rei Mariman.** 2024. “Generative AI can harm learning.” *Available at SSRN*, 4895486.
- Bick, Alexander, Adam Blandin, and David J Deming.** 2024. “The rapid adoption of generative AI.” National Bureau of Economic Research.
- Brynjolfsson, Erik, Danielle Li, and Lindsey R Raymond.** 2023. “Generative AI at work.” National Bureau of Economic Research.
- Cockburn, Iain M., Rebecca Henderson, and Scott Stern.** 2018. “The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis.” *The Economics of Artificial Intelligence: An Agenda*, 115–146. University of Chicago Press.
- Cui, Zheyuan Kevin, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz.** 2024. “The effects of generative AI on high skilled work: Evidence from three field experiments with software developers.” *Available at SSRN* 4945566.

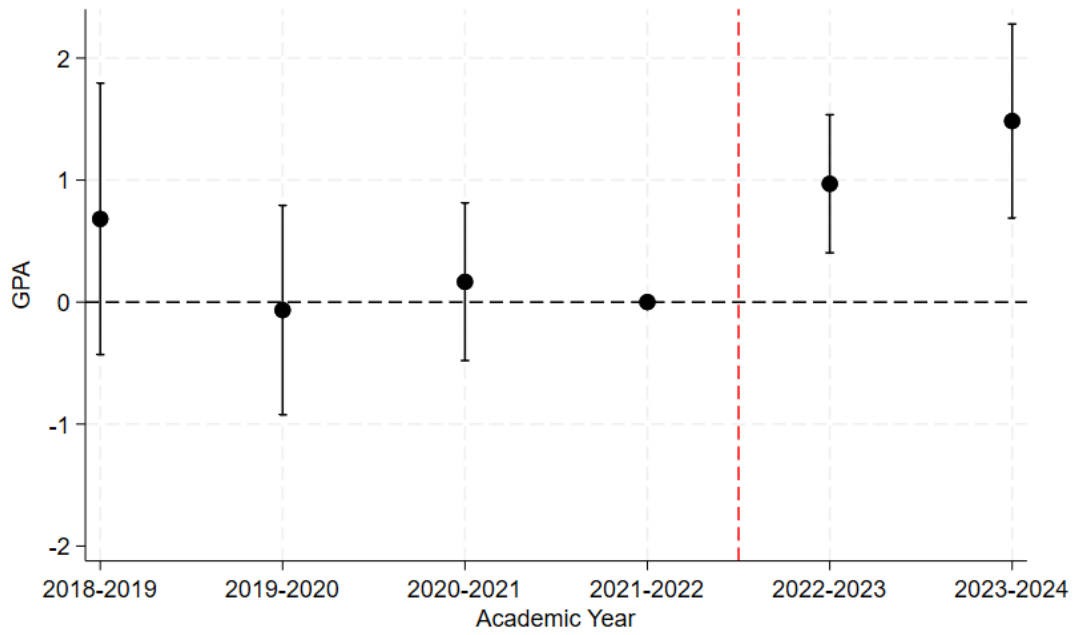
- Darvishi, Ali, Hassan Khosravi, Shazia Sadiq, Dragan Gašević, and George Siemens.** 2024. “Impact of AI assistance on student agency.” *Computers & Education*, 210: 104967.
- Dell’Acqua, Fabrizio, Edward McFowland III, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Kraye, François Candelon, and Karim 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).
- Demirci, Ozge, Jonas Hannane, and Xinrong Zhu.** 2025. “Who is AI replacing? the impact of generative ai on online freelancing platforms.” *Management Science*.
- Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock.** 2023. “GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.” Accessed: 2025-03-24.
- Georgieva, Kristalina.** 2024. “AI Will Transform the Global Economy. Let’s Make Sure It Benefits Humanity.”
- Gofman, Michael, and Zhao Jin.** 2024. “Artificial intelligence, education, and entrepreneurship.” *The Journal of Finance*, 79(1): 631–667.
- Koch, Michael, Ilya Manuylov, and Marcel Smolka.** 2021. “Robots and firms.” *The Economic Journal*, 131(638): 2553–2584.
- Lehmann, Matthias, Philipp B Cornelius, and Fabian J Sting.** 2024. “AI Meets the Classroom: When Does ChatGPT Harm Learning?” *arXiv preprint arXiv:2409.09047*.
- Lin, Luona, and Kim Parker.** 2025. “U.S. Workers Are More Worried Than Hopeful About Future AI Use in the Workplace.”
- Nie, Allen, Yash Chandak, Miroslav Suzara, Ali Malik, Juliette Woodrow, Matt Peng, Mehran Sahami, Emma Brunskill, and Chris Piech.** 2024. “The GPT Surprise: Offering Large Language Model Chat in a Massive Coding Class Reduced Engagement but Increased Adopters’ Exam Performances.” Center for Open Science.
- Noy, Shakked, and Whitney Zhang.** 2023. “Experimental evidence on the productivity effects of generative artificial intelligence.” *Science*, 381(6654): 187–192.
- Otis, Nicholas, Rowan P Clarke, Solene Delecourt, David Holtz, and Rembrand Koning.** 2023. “The uneven impact of generative AI on entrepreneurial performance.” Available at SSRN 4671369.

Figure 1: Definition of Treatment and Control Groups Based on Grade Components



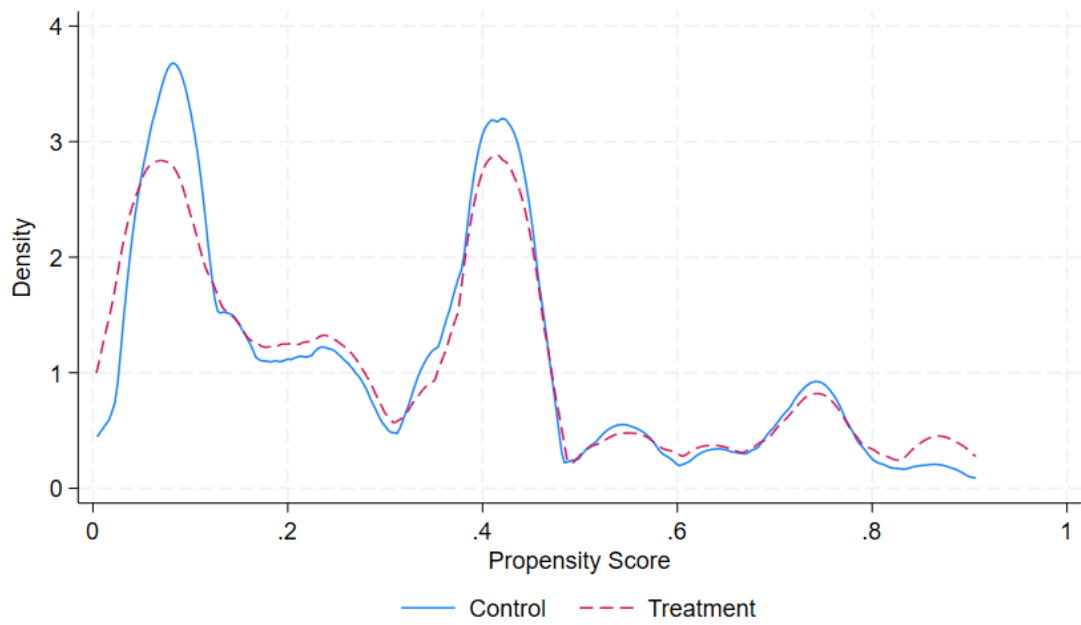
Notes: The histogram presents the distribution of courses according to the AI-incompatibility of their grade components. Courses in which all of the work is AI-incompatible are in the mass at 100. Courses in which all of the graded work is AI-compatible are in the mass at 0. The two vertical, dashed black lines represent cutoffs for treatment and control groups. Courses at and to the left of the first dotted line (60% AI-incompatible) are in the treatment group, and courses at and to the right of the second dotted line (90% AI-incompatible) are in the control group. Courses between the two dotted lines are omitted from the analysis. Grade components incompatible with AI usage are in-class exams and lab work.

Figure 2: Event Study: Effect of AI Availability on Grades



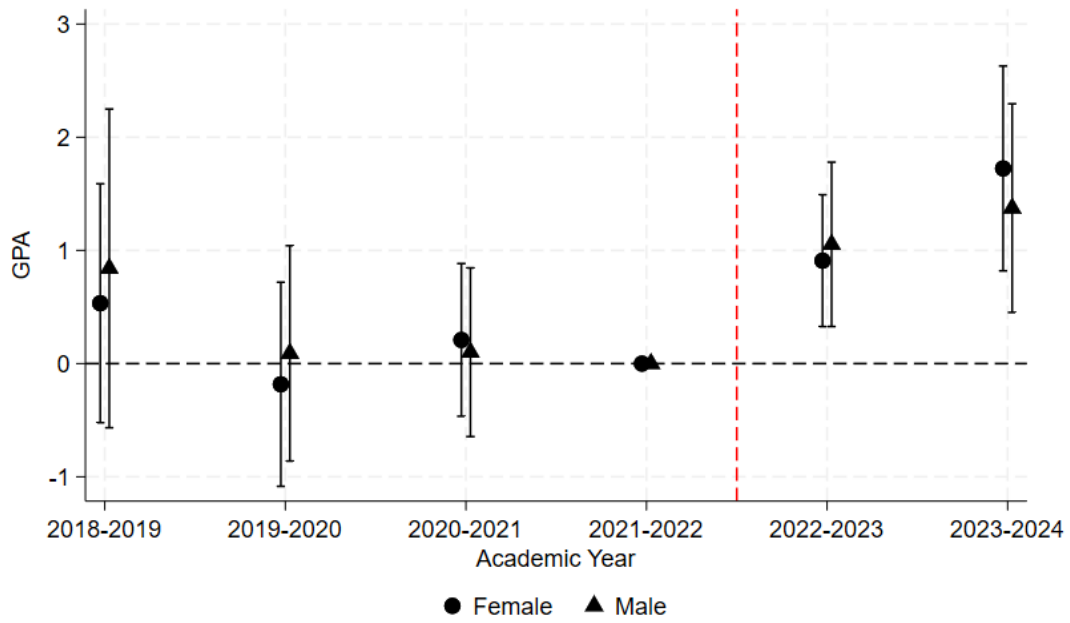
Notes: Points on the graph represent coefficients on $AI-comp*year$ interaction terms in Equation 1. GPA is measured on a scale from 0-100. Error bars reflect 95% confidence intervals. The vertical dashed line marks the beginning of the widespread availability of ChatGPT.

Figure 3: Distribution of Propensity Scores in Matched Sample



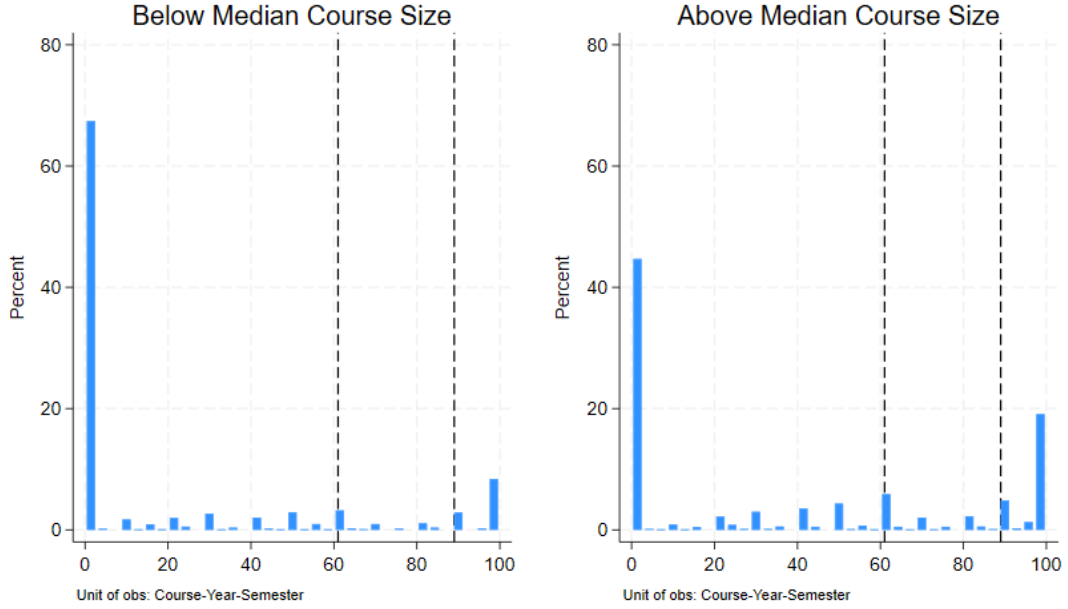
Notes: Figure shows the distributions of propensity scores for treated and control units in the matched sample.

Figure 4: Event Study: Effect of AI Availability on Grades, by Gender



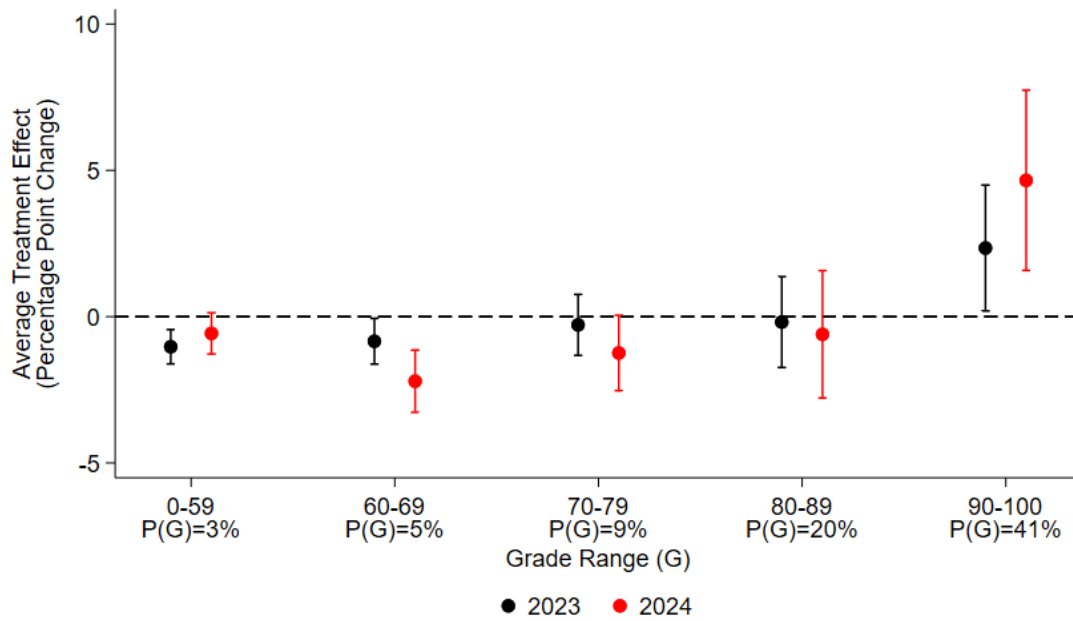
Notes: Points on the graph represent coefficients on $AI-comp * year$ interaction terms in Equation 1 when run separately for male and female students. GPA is measured on a scale from 0-100. Error bars reflect 95% confidence intervals. The vertical dashed line marks the beginning of the widespread availability of ChatGPT.

Figure 5: Distribution of Grade Components by Course Size



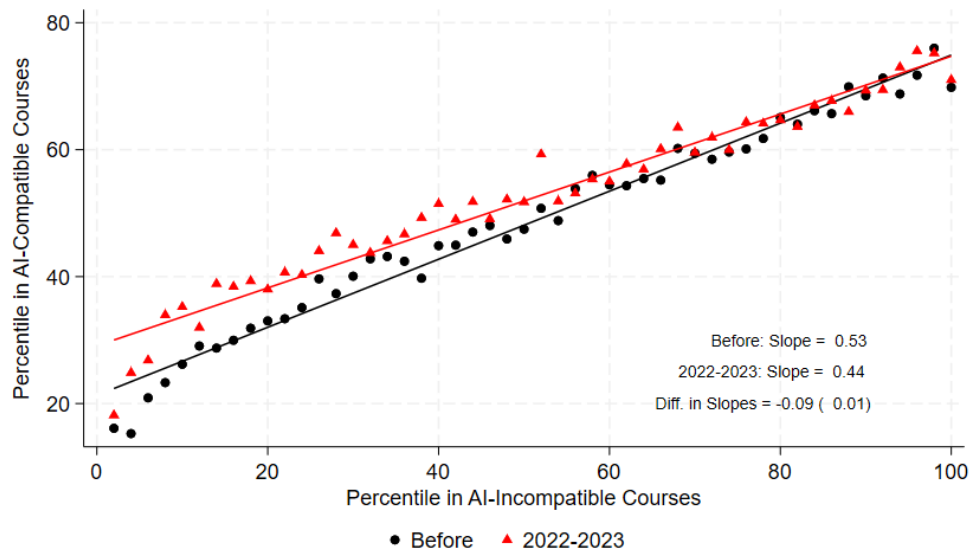
Notes: The histogram presents the distribution of courses according to the AI-incompatibility of their grade components in small versus large courses. Courses in which all of the work is AI-incompatible are in the mass at 100. Courses in which all of the graded work is AI-compatible are in the mass at 0. The two vertical, dashed black lines represent cutoffs for treatment and control groups. Courses to the left of the first dotted line (60% AI-incompatible) are in the treatment group, and courses to the right of the second dotted line (90% AI-incompatible) are in the control group. Courses between the two dotted lines are omitted from the analysis. Grade components incompatible with AI usage are in-class exams and lab work. Courses with a cumulative weight of non-AI compatible components below 60% are classified as treatment courses, while courses with a weight greater than 90% are classified as control courses.

Figure 6: Grade Distribution Effects: Probability of Scoring in each Grade Range

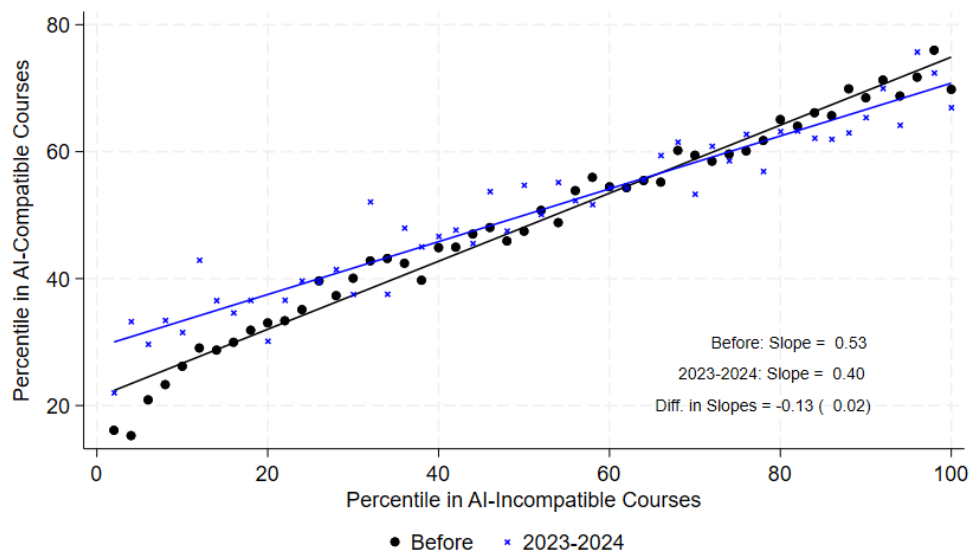


Notes: The figure presents coefficients on $AI-comp * year$ interactions from a version of Equation 1 in which the dependent variable is the probability of scoring in the given grade range. Error bars reflect 95% confidence intervals.

Figure 7: Within-Student Rank Consistency in AI-Compatible vs. AI-Incompatible Courses



(a) Before ChatGPT vs 1st Year After (2023)



(b) Before ChatGPT vs 2nd Year After (2024)

Notes: The figure plots within-student rank-rank relationships for student grade rankings across AI-compatible (y-axis) and AI-incompatible (x-axis) courses. Flatter slopes indicate less predictability of rank in one course type for rank in the other.

Table 1: Mean Treated and Control Course Characteristics

Variable	Treatment (AI-compatible)	Control (AI-incompatible)
<i>Course Characteristics</i>		
Avg. Grade	90	84
Class Size	36	78
Mandatory Class	44%	63%
<i>Fields</i>		
Humanities	22%	10%
Natural Sciences	15%	22%
Social Sciences	22%	12%
Agriculture	13%	34%
Law	5%	7%
Social Work	9%	6%
Engineering	6%	1%
<i>Students</i>		
Age	27	26
Male	40%	43%
Num. Course-Semester Obs 2,640		4,087
Num. Student-Course-Semester Obs	130,739	184,597

Table 2: Effect of AI Availability on Student Grades

	<i>Dependent Variable: Grade (0-100)</i>				
	Baseline DID		PS Matching DID		
	(1)	(2)	(3)	(4)	(5)
<i>Before:</i>					
2018-2019*AI-compatible	0.682 (0.567)	0.375 (0.858)	-0.286 (0.850)	0.135 (0.812)	0.117 (0.816)
2019-2020*AI-compatible	-0.066 (0.438)	0.530 (0.634)	0.296 (0.557)	-0.023 (0.563)	-0.048 (0.565)
2020-2021*AI-compatible	0.166 (0.330)	0.256 (0.490)	0.077 (0.527)	-0.162 (0.472)	-0.168 (0.473)
<i>After:</i>					
2022-2023*AI-compatible	0.970*** (0.289)	0.942** (0.387)	0.809** (0.365)	0.645* (0.345)	0.633* (0.344)
2023-2024*AI-compatible	1.484*** (0.406)	1.334** (0.540)	0.745 (0.545)	0.566 (0.486)	0.532 (0.483)
Course Char. Controls	✓	✓	✗	✓	✓
Match Group FE	✗	✗	✓	✓	✓
P-Score Control	✗	✗	✗	✗	✓
R^2	0.477	0.554	0.565	0.581	0.581
N	500,611	361,591	361,591	361,591	361,591

Note: An observation in the sample is a student-course-semester. Column (1) uses the full sample, while columns (2)-(5) use the reduced sample of courses that were matched with propensity score matching. The dependent variable is the student's course grade, from 0-100. The baseline (omitted) interaction is 2021-2022*AI-compatible. AI-compatible (treated) courses are defined as those in which in-class exams and lab work comprise less than 60% of the grade, such that AI can be used for much of the graded work. AI-incompatible (control) courses are defined as those in which in-class exams and lab work comprise more than 90% of the grade. Propensity score matching is NN-1, and only courses with constant treatment status over time are included. Course characteristic controls include fixed effects for class size quartile, department, number of credits, class type, mandatory class, language of instruction, and degree (BA/MA). All specifications include student FE and semester FE. Standard errors are clustered by course. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Heterogeneity in AI Effect on Grades

Dimension		No	Yes
Male	AI Compatible X 2022-2023	0.909*** (0.297)	1.053*** (0.370)
	AI Compatible X 2023-2024	1.724*** (0.462)	1.374*** (0.470)
	Observations	276,397	200,672
Young (< Age 26)	AI Compatible X 2022-2023	0.933*** (0.306)	1.132*** (0.358)
	AI Compatible X 2023-2024	1.287*** (0.398)	2.079*** (0.537)
	Observations	237,658	238,466
Advanced Course	AI Compatible X 2022-2023	0.999 (0.742)	0.843** (0.355)
	AI Compatible X 2023-2024	0.935 (0.876)	1.466*** (0.511)
	Observations	144,254	254,662
STEM	AI Compatible X 2022-2023	0.769*** (0.287)	1.332 (0.920)
	AI Compatible X 2023-2024	1.511*** (0.409)	1.269 (0.969)
	Observations	373,967	119,260
Business (vs. non-Business non-STEM)	AI Compatible X 2022-2023	0.769** (0.298)	0.188 (1.101)
	AI Compatible X 2023-2024	1.398*** (0.416)	0.522 (1.328)
	Observations	343,117	29,930
Senior Professor	AI Compatible X 2022-2023	0.791** (0.392)	1.093** (0.486)
	AI Compatible X 2023-2024	1.613*** (0.578)	1.105* (0.576)
	Observations	272,268	212,846
Large Course(>25 students)	AI Compatible X 2022-2023	1.123** (0.565)	1.075*** (0.327)
	AI Compatible X 2023-2024	2.423*** (0.813)	1.671*** (0.453)
	Observations	56,609	435,675

Note: Each cell in the table represents a separate regression illustrating heterogeneity in the treatment effects along the dimension indicated on the left. An observation in each sample is a student-course-semester. The dependent variable is the student's course grade, from 0-100. The baseline (omitted) interaction is 2021-2022*AI-compatible. AI-compatible (treated) courses are defined as those in which in-class exams and lab work comprise less than 60% of the grade, such that AI can be used for much of the graded work. AI-incompatible (control) courses are defined as those in which in-class exams and lab work comprise more than 90% of the grade. Course characteristic controls include fixed effects for class size quartile, department, number of credits, class type, mandatory class, language of instruction, and degree (BA/MA). All specifications include student FE and semester FE. Standard errors are clustered by course. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Effect of AI Availability on the Grade Distribution

<i>Dependent Variable:</i> Grade at:	Baseline DID			PS Matching DID		
	(1) p25	(2) p50	(3) p75	(4) p25	(5) p50	(6) p75
<i>Before:</i>						
2018-2019*AI-compatible	0.875 (1.077)	-0.157 (0.784)	-0.694 (0.678)	1.093 (0.993)	-0.083 (0.798)	-0.569 (0.674)
2019-2020*AI-compatible	0.081 (1.114)	-0.443 (0.788)	-0.594 (0.705)	-0.253 (1.182)	-0.720 (0.866)	-0.646 (0.745)
2020-2021*AI-compatible	1.620 (1.206)	0.270 (0.893)	-0.507 (0.523)	1.124 (1.293)	0.014 (0.957)	-0.498 (0.537)
<i>After:</i>						
2022-2023*AI-compatible	3.014*** (0.933)	1.272** (0.612)	0.738 (0.453)	2.618** (1.037)	1.044 (0.675)	0.719 (0.495)
2023-2024*AI-compatible	2.383** (0.978)	1.125 (0.752)	0.583 (0.511)	0.917 (0.877)	0.184 (0.686)	0.207 (0.519)
Dep. Var. Avg.	84.72	90.18	94.14	84.72	90.18	94.14
R^2	0.426	0.401	0.312	0.588	0.582	0.508
N	10,076	10,076	10,076	10,076	10,076	10,076

Note: An observation in the sample is a course-semester. All columns are estimated on the reduced sample of courses that were matched with propensity score matching. The dependent variable is the grade at the percentile indicated at the top of each column. The baseline (omitted) interaction is 2021-2022*AI-compatible. AI-compatible (treated) courses are defined as those in which in-class exams and lab work comprise less than 60% of the grade, such that AI can be used for much of the graded work. AI-incompatible (control) courses are defined as those in which in-class exams and lab work comprise more than 90% of the grade. Propensity score matching is NN-1, and only courses with constant treatment status over time are included. Course characteristic controls in all specifications include fixed effects for class size quartile, department, number of credits, class type, mandatory class, language of instruction, and degree (BA/MA). All specifications include student FE and semester FE. Standard errors are clustered by course. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: The AI Experience Effect on Advanced Course Grades

	AI-Compatible Advanced Course Grade			AI- Incompatible Advanced Course Grade		
	(1)	(2)	(3)	(4)	(5)	(6)
Second Year*AI Experience	1.194*** (0.267)	0.656*** (0.245)	0.928*** (0.229)	-9.232*** (0.353)	-0.431 (0.382)	-0.507 (0.353)
Course Char. Controls	✗	✓	✓	✗	✓	✓
Student FE	✗	✗	✓	✗	✗	✓
R^2	0.029	0.227	0.555	0.080	0.246	0.548
N	22,806	22,806	22,806	34,829	34,829	34,829

Note: An observation in the sample is a student-course-semester. Estimates presented are coefficients on the interaction of Second Year*AI Experience as described in Equations 4a and 4b. The dependent variable is the student's course grade, from 0-100, in the type of advanced course indicated. The sample is restricted to the years 2022-2024. In columns (1)-(3), the sample includes AI-compatible courses only from students' first and second year of studies. In columns (4)-(6), the sample includes AI-compatible courses from students' first year of studies and AI-incompatible courses from their second year. The "AI Experience" (treatment) group is defined as students who began their studies in 2022-23, once ChatGPT was available, and took at least one AI-compatible introductory course in that year. The control group is defined as students who began their studies in 2021-22, before ChatGPT was available, and took at least one AI-compatible introductory course in that year. Course characteristic controls include fixed effects for class size quartile, department, number of credits, class type, mandatory class, language of instruction, and degree (BA/MA). All specifications include semester FE. Standard errors are clustered by course. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix for Online Publication

A. Survey of Students on AI Adoption

- **Q1:** When did you first begin using AI (e.g., ChatGPT) to help you in any way with your coursework (e.g., to explain things to you or to help in producing assignments)?
Answer: _____ (Month/Year dropdown)
- **Q2:** For what percent of your home assignments in the last semester (Fall 2024-25) would you say you used AI to help in some way?
Answer: _____ (Percent dropdown)
- **Q3:** For what percent of your home assignments did AI reduce your own effort by more than 50% or produce more than half of the assignment?
Answer: _____ (Percent dropdown)
- **Q4:** What percent of your submitted assignments for which AI (e.g., ChatGPT) was used would be identified by an AI detector?
Note: AI detectors may have an accuracy of 99%.
Answer: _____ (Percent dropdown)
- **Q5:** Have you experienced an increase in your grades since you started using AI to aid with your schoolwork?
Answer: _____ (Yes / No / I don't know)
- **Q6:** To what extent do you think your use of AI for your schoolwork will make you more productive in the labor market after you graduate?
Answer: _____ (-5 = less productive; 0 = neutral; 5 = much more productive)
- **Q7:** To what extent do you think your use of AI for your schoolwork may be replacing learning/knowledge acquisition you otherwise would have done?
Answer: _____ (1 = not at all; 5 = to a great extent)
- **Q8:** To what extent do you think the knowledge you don't gain due to AI use is important?
Answer: _____ (1 = not at all important; 5 = very important)

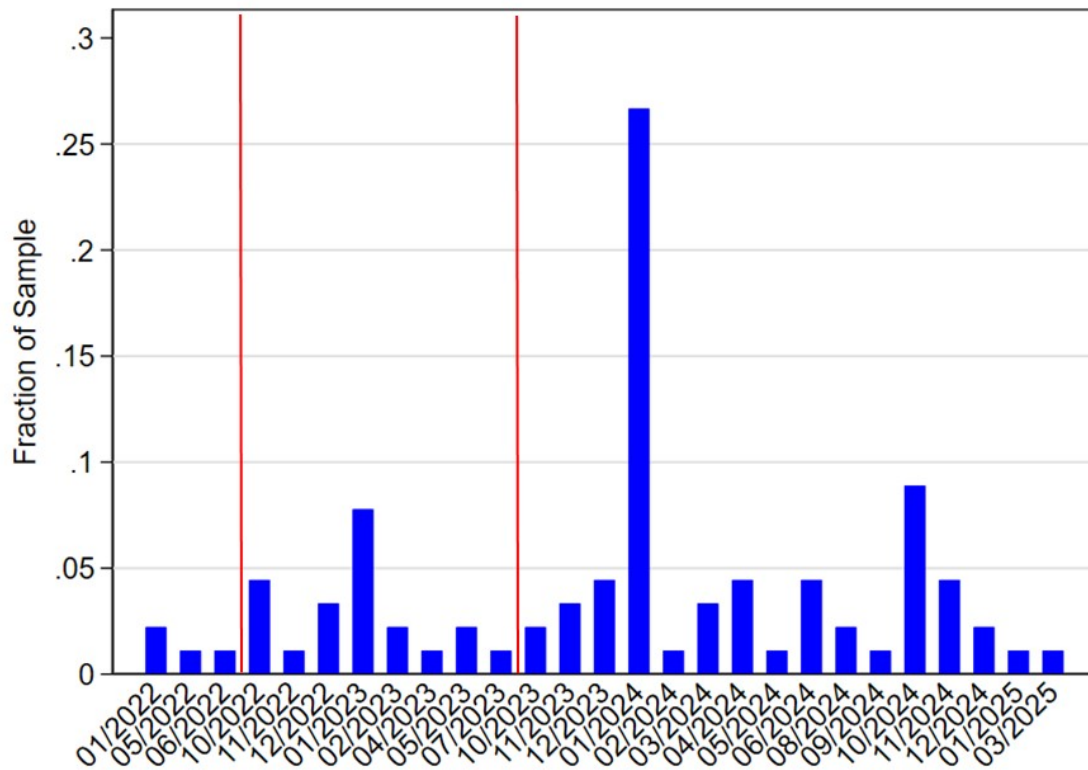
- **Q9:** To what extent do you expect to have access to AI for job tasks after graduation?
Answer: _____ (1 = not available for any tasks; 5 = available for all tasks)

Demographics

- **Sex:** _____ (Male / Female)
- **Age:** _____ (15–45 dropdown)
- **Ethnicity:** _____ (Jewish / Arab / Asian / Other Caucasian)
- **Immigration Year:** _____ (1970–2025 dropdown or N/A)
- **Program Level:** _____ (BA / MA / PhD)
- **Joint Degree with Another Faculty?**
(Law / Medicine / Computer Science / Statistics / Economics / Psychology / Other
Social Science / Other Hard Science / Other Professional / None)

B. Appendix Figures

Figure A1: Adoption Dates of ChatGPT Based on Survey Respondents



Notes: The histogram presents the distribution of ChatGPT adoption dates by month based on survey responses. The two vertical, red lines represent cutoffs for the first and second years of ChatGPT availability as defined in our analysis.

C. Appendix Tables

Table A1: Summary Statistics from Survey

Variable	Mean	P25	P50	P75	N
Demographics					
Male	0.516	0	1	1	91
Jewish	0.857	1	1	1	91
Arab	0.110	0	0	0	91
Age	26.21	23	25	28	91
New Immigrant	0.286	0	0	1	91
English Version	0.187	0	0	0	91
Survey Responses					
Early Adopter (by July 2023)	0.278	0	0	1	91
Adopter (by July 2024)	0.789	1	1	1	91
AI Use – Percent of Assignments	60.66%	40%	60%	80%	91
Usage Reduces Effort by at-least 50%	47.25%	20%	50%	70%	91
Beliefs about AI					
AI Use will Increase Work Productivity (Scale 1-5)	2.96	2	4	5	91
AI Replaces Learning (Scale 1-5)	2.63	2	3	4	91
Importance of Lost Knowledge (Scale 1-5)	2.38	1	2	3	91