How exposed is higher education to Artificial Intelligence?*

Jacob Light†

June 13, 2026

[Click here for latest version]

Abstract

How exposed is higher education to artificial intelligence, and how are institutions and students responding? I construct a novel measure of curricular exposure to AI by applying the task-based framework of Webb (2019) to a comprehensive dataset of course offerings from more than 1,000 U.S. colleges and universities. Extracting verb-object pairs from course descriptions and patent texts, I measure the overlap between what students are trained to do and what AI and other technologies can do. I find that college curricula are more exposed to AI than to prior technological shocks, and that exposure has increased over the past decade, driven by the diffusion of data analysis coursework across disciplines. Exposure is highest in fields intensive in writing and data analysis. Despite high and growing exposure, I find limited evidence of adjustment: using a panel of course syllabi spanning the pre- and post-ChatGPT periods, instructors rapidly adopt AI policies but make only modest changes to course structure, and student enrollment patterns show little response. These findings suggest that higher education may adjust slowly to technological change, even when the skills it produces are highly exposed.

JEL Classification: J24, I2, O3.

*Thanks to David Autor, Nick Bloom, David Figlio, Caroline Hoxby, and Chris Karbownik for advice and feedback, and to seminar participants at the ASSA Annual Meeting, AEFP, Columbia University, the Hoover Institution, and the Stanford Graduate School of Education. Special thanks to Michael Webb for sharing code and providing feedback. Nadia Chung provided excellent research assistance. All errors are my own.

†Hoover Institution, Stanford University. jdlight@stanford.edu

1 Introduction

Technological disruption has long been a feature of labor markets. Each wave of innovation expands the production possibilities frontier but may simultaneously render specific skills — and the workers who supply them — less valuable. In recent decades, the brunt of automation has fallen on routine, lower-skill jobs (Autor et al. 2003). Under such skill-biased technological change, a standard policy prescription has been to encourage workers at risk of displacement to upskill through higher education or lifelong learning on the premise that higher education equips workers with harder-to-automate capabilities. That logic may now be in question.

A new generation of artificial intelligence (AI) tools differs from earlier technologies in three notable ways. First, their potential application space is extraordinarily broad, spanning language, vision, prediction, and code. Second, many models operate with minimal human supervision, performing entire task bundles rather than single modular steps. Third — and central to this paper — the tasks at which modern AI excels overlap heavily with those performed by workers in traditionally high-skill occupations (Brynjolfsson and Mitchell 2017). If AI increasingly competes on the same cognitive terrain that universities teach, the value of "upskilling" through college coursework is no longer assured: depending on whether these tools substitute for or augment the skills that coursework builds, the return to college training in exposed fields could erode or be amplified relative to previous eras.

This paper provides the first systematic measurement of how exposed college coursework is to modern AI capabilities. Building on the task-based framework of Webb (2019), I combine a novel dataset of course descriptions, enrollments, and offerings from 1,019 U.S. colleges and universities (Light 2026) with two measures of technological capabilities. The first uses patent data to characterize the distribution of AI capabilities at the technological frontier. The second, and preferred measure, uses task-level estimates of exposure to large language models (LLMs) from Eloundou et al. (2023), capturing the capabilities of tools that are already widely used by students and instructors.

For each field of study, I extract verb-object pairs (e.g., "analyze data," "write essays") from course descriptions, construct a task distribution, and measure its overlap with the task distributions implied by each technology. The resulting exposure score captures the extent to which tasks taught in college can be performed more efficiently using a given technology. The measure is intentionally agnostic about valence: it takes no stand on whether AI substitutes for or complements these tasks in the labor market, nor on whether the tools substitute for or complement student learning. What exposure locates unambiguously is the classroom margin central to this paper — where students are most able to substitute AI for their own effort and, therefore, where the skill a course is designed to build is most at risk of going undeveloped and undetected.

The paper documents two main descriptive findings. First, college coursework is substantially more exposed to AI than to earlier waves of technological change. While exposure to robotics remains low across fields, exposure to AI — particularly under the LLM-based measure — is widespread, with many fields lying well above the corresponding distribution of occupational exposure. Second, exposure to AI has risen steadily over the past decade, driven by the diffusion of data analysis and related tasks across disciplines.

In the second part of the paper, I examine how higher education responds to a sharp technology shock: the release of ChatGPT in November 2022. A high exposure score does not by itself imply that the affected coursework has become less valuable. Schools teach many tasks that existing technology already performs (e.g., basic arithmetic) because performing them is foundational to higher-order learning, and because human capital may retain value even for tasks machines can complete, as when workers must supervise and evaluate AI output. The concern raised by AI is practical: because LLMs can complete exposed coursework directly, students in high-exposure fields are the most able to substitute AI for their own effort. This threatens human capital formation along two margins: students may acquire less of the skill a course is designed to build, and instructors may be less able to measure whether that skill was acquired, because the assessments most vulnerable to AI substitution (out-of-class essays, problem sets, projects) are precisely those that lose signal value when students can outsource them.

The instructor margin is where exposure binds most sharply. Exposure is necessary but not sufficient for this measurement problem: it marks where AI could substitute for student effort, but whether that ability actually erodes the signal value of assessment depends on how a course is assessed — a design choice the instructor controls. Course administration is therefore the institution's primary defense of human capital measurement against the substitution risk that exposure identifies, which makes whether and where instructors respond a first-order empirical question. The three pieces of the analysis map onto this logic: the exposure score identifies where the concern might be greatest; the grade component analysis asks whether the measurement of learning is keeping pace with that concern; and the interaction between the two asks whether the courses where the problem is potentially most acute are disproportionately adapting. I study responses along both this instructor margin and the student demand margin.

On the student side, I measure enrollment responses across fields with different levels of AI exposure. In aggregate, the short-run response is small: a one occupation-standard-deviation increase in LLM exposure (roughly the difference between Skilled Trade and Computer

Science) is associated with a statistically significant but economically modest 1.5 log point decline in enrollment between 2022-23 and 2025-26.

On the instructor side, I construct a panel of syllabi covering all courses offered at 25 colleges and universities to track changes in course design. I find rapid adoption of AI policies: by Fall 2025, a majority of courses explicitly regulate the use of AI tools, with higher rates of restriction in high-AI exposure fields, particularly writing-intensive Humanities and Social Sciences, than in STEM or Business. Changes in course structure, however, are modest. In particular, there is only limited reweighting toward in-class assessments such as exams and participation. These adjustments are small and gradual, especially when compared to the large and persistent changes in course design observed during the COVID-19 pandemic.

Taken together, these findings point to a gap between where the measurement of student learning is most threatened and whether the instruments that measure it are adapting. College coursework is highly exposed to AI broadly, and to LLM capabilities specifically, yet instructor adjustment to the assessments most vulnerable to AI substitution is modest, and student reallocation, though concentrated among the most exposed students at research universities, remains limited in aggregate. The contrast with the pandemic is striking. COVID made existing modes of instruction infeasible and forced an immediate, dramatic, and persistent restructuring of how learning was measured; the ChatGPT period degrades the informativeness of out-of-class assessment without making any existing mode of instruction infeasible, and no comparable restructuring appears. This contrast suggests that institutional frictions, uncertainty about the long-run role of AI, and the difficulty of redesigning curricula may slow the responsiveness of higher education to technological change.

I validate the task-based measurement approach using several complementary exercises. The tasks extracted from course descriptions align closely with intuitive differences across fields. For example, "analyze data" and "solve equations" are concentrated in quantitative disciplines, while "write essays" and "interpret texts" are more prominent in the Humanities. Moreover, the task distributions implied by course descriptions correlate strongly with the tasks performed in the occupations associated with each field. These patterns suggest that the exposure measure captures meaningful variation in the content of human capital formation.

The paper makes three central contributions. First, it extends task-based exposure metrics from occupations to the human capital formation in college, offering a forward-looking benchmark for policy discussions around the value of college majors during a period of potential AI disruption in the labor market. Second, the analysis reveals substantial within-major heterogeneity across institution types, driven largely by differences in the level of courses offered rather than differences in skills/tasks emphasized in the same course offered across institutions. The findings demonstrate how the "same degree" can differ meaningfully depending on where it is earned and reveal limitations of data sources that record information only on an individual's field of study. Third, it reframes AI exposure as a problem of human capital measurement and provides new evidence on how higher education defends against it. The novel object is a measure of where the classroom's capacity to verify learning is most at risk; the behavioral analysis then asks whether course administration adapts to protect it.

The paper builds on a large literature in labor economics studying how technology affects employment and occupational choice. Autor et al. (2003) introduced the idea of modeling occupations as bundles of tasks, with technology potentially complementing or substituting for workers in those tasks. Subsequent work has shown that different technologies vary in how they alter demand for skill — some technologies complement cognitive or non-routine work, while others automate routine tasks. For example, technological change in the late 20th century was skill-biased: it disproportionately substituted for workers in jobs with lower education requirements while complementing highly educated workers, thereby contributing to the growing earnings gap between more and less educated workers (Autor et al. 2006). Although evidence on the labor market effects of AI remains nascent because of its recent development and, as yet, limited deployment (McElheran et al. 2024), there is growing empirical support for these task-based predictions. Establishments that have adopted AI tend to demand more workers who can integrate AI into production processes, while simultaneously reducing demand for workers in occupations that AI can readily perform (Acemoglu et al. 2022).

Several recent papers have developed methods to measure the exposure of occupations to new technologies. Felten et al. (2021) and Brynjolfsson et al. (2018) rely on crowdsourced scoring: Brynjolfsson et al. (2018), for example, use a standardized rubric to assess whether tasks are suitable to machine learning based on how clearly inputs map to outputs and how easily performance can be measured. Felten et al. (2021) focus on perceptions of overlap between specific AI capabilities and occupational abilities. In contrast, Webb (2019) uses a data-driven approach: he measures exposure by comparing the text of patents (which describe what technologies do) with O*NET task descriptions (which describe what workers do). This method has several advantages for my project. First, it relies on textual similarity between technology capabilities and tasks, rather than subjective judgments of suitability. Second, because it is based on patents rather than realized wage changes or job postings, it provides a forward-looking measure that reflects the technological frontier, rather than current market conditions. Finally, the method can be easily updated as new patents emerge, making it a flexible way to track technological change over time.[1]

[1] Although I follow Webb's use of patents here, the same methodology could also be applied to other

My paper applies this logic one step earlier in the labor market pipeline — mapping patents not to occupations, but to the tasks embedded in college courses. This shifts the unit of analysis from workers to students and moves focus from firms' hiring decisions to universities' curricular decisions and students' human capital investment.

Applications of a task or skills framework to research in higher education are more limited, largely because of the need for rich data that records skills or tasks at a field- or occupation-by-field level. Recent work in this area has relied on either course descriptions or course syllabi to extract institution- and course-level detail on course content. Closest in spirit to my project is work by Javadian Sabet et al. (2024), who link course descriptions from a large corpus of syllabi at numerous colleges and universities to detailed work activities in the labor market. Their data tool offers an alternative source for assessing course or field exposure to AI, although it is limited for measuring exposure of more recent courses to technological change because its data cutoff is 2017. Biasi and Ma (2022) also use this syllabus data to track how courses incorporate material at the frontier of academic research. While this shares my goal of linking course content to cutting-edge knowledge, their focus is on courses' inclusion of academic frontier material, whereas my analysis asks whether courses incorporate tasks that overlap with what technologies at the frontier can do. Finally, a recent white paper by Timmerman (2025) measures field-level AI exposure by mapping majors to occupations and applying occupational AI exposure scores. My approach differs in measuring exposure directly at the field-of-study level, rather than using occupations as a mediator. It also benefits from rich data that allow me to measure heterogeneity in exposure within the same field across institutions and over time. My analysis confirms findings in contemporaneous work by Salitel (2026) and Chirikov (2026), who describe AI policy statements in course syllabi, while extending the question of how instructors respond to ChatGPT's release by considering more broadly how course structure has changed.

The rest of the paper proceeds as follows. In Section 2, I summarize the data sources used for this project. Section 3 outlines the text analysis used to measure field-technology exposure. I walk through validation exercises in Section 4 to stress-test the task-based framework as applied in this new setting. Section 5 summarizes results in the measurement of field-technology exposure, and Section 6 summarizes instructor and student responses to ChatGPT's release. Section 7 concludes.

sources of technology data, like GitHub commits or academic papers, to expand the measure's coverage as AI evolves.

2 Data

The primary data sources for this project are patent text from the US Patent Office, covering all patents filed between 2000-2025, and comprehensive course offerings data from 1,019 US colleges and universities.

2.1 Course data

2.1.1 Course catalog data

The course catalog dataset contains detailed information on course offerings at a large, nationally representative panel of US colleges and universities (Light 2026). The data are assembled by scraping online course catalogs and semester course schedules: catalogs supply the course description and permanent attributes of each course, while schedule data link each section to an instructor, instruction format, and section-level headcount enrollment. The resulting dataset captures 50 million course sections offered since 1996 across 1,019 institutions.

The dataset forms an unbalanced panel that includes the complete set of course offerings for each institution-year. For the typical course section, I observe headcount enrollment, instructor(s), instruction format (e.g., in-person or virtual), and a brief text description of course content. Figure 1 is an example of the information contained in a typical observation. Most crucial for this project are the course descriptions, from which I extract tasks, and the enrollment counts.

The sample is broadly representative of US colleges and universities. It covers 37% of non-profit four-year institutions and 31% of non-profit two-year institutions — including 58% and 43% of enrollment, respectively. It also mirrors national distributions of key characteristics such as selectivity, tuition costs, and endowment levels, although it skews somewhat toward larger, public institutions because it misses many very small, often religiously affiliated private colleges.[2] This compositional skew is unlikely to affect the cross-field patterns documented in the measurement section, which pool institutions; it results in estimates that more accurately reflect the experience of the typical college student (who attends a school much larger than average) rather than a student at the typical college (which, by virtue of the large number of small private institutions, is smaller than the average school in my sample).

[2]See Light (2026) for a detailed discussion of coverage and representativeness.

2.1.2 Syllabus data

Measuring the evolution of coursework in response to ChatGPT requires detailed information on the structure of individual college courses. I construct a panel of course syllabi from 20 US colleges and universities spanning the pre- and post-ChatGPT periods, with coverage as early as 2005. In total, the sample contains approximately 1.1 million syllabi covering 400,000 unique courses.[3] Benchmarking against the course catalog data, the syllabus sample covers approximately 50% of undergraduate course sections offered at these institutions, accounting for 65% of enrollment.[4]

The sample contains publicly available syllabi from universities that publish full syllabus archives online and not behind a login wall. I only include institutions for which a reasonably complete archive is available. The majority of institutions in the sample are located in Texas, where Texas HB 2504 requires universities to post course syllabi publicly. While the law mandates only a two-year archive, many institutions maintain records dating back to the law's passage in 2009, providing a pre-period of sufficient length to establish pre-trends and benchmark the COVID-19 shock. The remaining institutions are primarily research universities from other states that voluntarily publish complete syllabus archives. The sample therefore skews toward larger, public four-year institutions, ranging from highly selective research universities to less selective regional teaching institutions. Because the main identifying variation in the syllabus analysis is within-course over time — tracking how the same course changes before and after ChatGPT — the cross-sectional composition of the sample is less critical for internal validity than for external generalizability.

For this paper, I use an automated approach to extract information related to course structure and administration from each syllabus. Specifically, I extract grade components with their associated weights, and course policies — including any policies regarding the use of AI tools — using the OpenAI API.[5]

For the data to be useful, it is essential that the data extraction accurately reflects the information in the syllabus. Of particular concern is the risk of AI hallucinations producing output that seems plausible but does not match the source document. I take two steps to guard against this. First, I have the LLM output the raw text from the syllabus alongside

[3] Appendix Table A-1 summarizes the data availability for the schools in the syllabus sample.

[4] Excluding independent study, study abroad, private lessons, and other non-classroom-based courses. Incomplete coverage typically arises because of imperfect compliance with university syllabus uploading policies at the instructor level. This increases the likelihood that I observe the same course over time, which is essential for my empirical analysis. Bias would be introduced if instructors who are more likely to comply with the university policies are differentially responsive in updating their syllabi following ChatGPT's release. Analysis restricting only to institutions with high compliance produce similar estimates to those summarized in this paper, suggesting that this is not a first-order concern.

[5] See Appendix A.1 for details on the LLM-based extraction.

the processed grade component name and weight, allowing me to verify that the extracted information matches the syllabus verbatim. Second, I validate the accuracy of the LLM output using a human evaluator on a sample of 250 syllabi. The LLM accurately extracts the full grading scheme in 85% of cases.[6]

I classify AI policy stances into four categories: *Prohibited* (no AI use permitted), *Limited* (AI permitted for specified tasks only), *Allowed* (AI use permitted without restriction), and *Unclear* (policy present but ambiguous). Syllabi with no mention of AI tools are coded as having no policy. The analysis in Section 6 uses a binary indicator for whether a syllabus contains any AI policy.

I classify grading components into seven categories using string matches on the raw text of each component: exams (tests, quizzes, etc.), homework, participation, projects, presentations, labs, and essays. Components that do not fit into one of these seven categories are classified as "other." For the main grade component analysis, I aggregate these categories into two groups based on the degree to which AI tools can substitute for student effort: *AI-substitute* assessments, comprising essays and homework, which students can complete outside of class and with access to AI tools; and *AI-non-substitute* assessments, comprising exams, participation, presentations, and labs, which require in-person demonstration of knowledge or skills and are therefore less amenable to AI assistance. This aggregation motivates the central prediction tested in Section 6: if instructors respond to ChatGPT by trying to preserve the signal value of assessment, they should shift weight toward AI-non-substitute components and away from AI-substitute ones.[7]

2.2 Technology data

2.2.1 Patent data

I obtain patent titles and abstracts for all patents filed with the US Patent and Trademark Office between 2000-2025. Titles typically provide brief descriptions of a technology's core function, while abstracts provide longer summaries of the technology's development and applications. For my analysis, I use patent titles to extract the tasks a technology can perform, following Webb (2019).[8] An example of a patent title is "Analyzing flight data

[6]Errors typically take the form of not reporting any grading information or adding an extra grading component that is the total of different sub-components: for example, a syllabus may list Assignment 1 and Assignment 2 as grade components and include an Assignment Subtotal line in the grade composition table. In post-processing, I remove any grade item that includes the string "total." These errors do not appear to be systematic for specific grading components or in specific fields or years. Thus, the consequence of these inaccuracies will primarily be to introduce measurement error into the estimates.

[7][NTD: provide descriptive statistics on the distribution of grading components in Appendix Table X.]

[8]Although abstracts offer richer detail, they often include background or development information unrelated to direct applications. Results using abstracts are qualitatively similar to those using titles.

using predictive models." The final patent dataset includes approximately 135,000 patents covering Robotics and Artificial Intelligence.

To classify patents into technology classes, I apply a keyword filter similar to Webb (2019). For Robots, I include patents whose titles or abstracts contain the word "robot" or derivatives of "manipulate," excluding patents with CPC codes A61 (medical or veterinary science) and B01 (physical or chemical processes). For AI, I expand Webb's original definition, which pre-dates the widespread deployment of large language models and therefore omits the vocabulary of transformer-based systems, to include phrases associated with recent AI innovations: "artificial intelligence," "deep learning," "large language model," "llm," "neural network," "predictive model," "reinforcement learning," "supervised learning," or "transformer model."

The resulting dataset includes 28,165 Robot patents and 26,136 AI patents. Approximately half of the patent titles contain at least one verb-object pair.

I do not report results for software exposure, which is included in Webb (2019). Webb's software classification was developed for an earlier era in which software and AI were more clearly separable. In my data, the rank correlation between field-level AI and software exposure scores is 0.91, reflecting the increasing integration of AI functionality into software-related innovations in recent years. Because the two measures are nearly collinear at the field level, including both would introduce substantial multicollinearity without adding analytical clarity. I focus on Robots and AI, which remain meaningfully separable — their field-level exposure scores have a rank correlation of only 0.28 — and which correspond more directly to the conceptual distinction between automation of physical tasks and automation of cognitive ones.[9]

2.2.2 LLM capability data

My second measure of AI capabilities is based on Eloundou et al. (2023), who evaluate the extent to which large language models (LLMs), specifically GPT-4, can reduce the time required to complete tasks in O*NET. Using a combination of crowdsourced human assessments and model-based evaluations, they classify each task into three categories. Tasks are assigned "no exposure" if GPT-4 produces little or no reduction in completion time; "direct exposure" if GPT-4 alone can reduce completion time by at least 50 percent; and "indirect exposure" if achieving this reduction requires GPT-4 in combination with additional tools or software.

[9] Appendix Figure A-1 shows the full matrix of pairwise correlations across all technology measures. The Robot-OpenAI correlation is −0.38, confirming that the two measures capture opposite ends of the skill spectrum.

To integrate these data into my framework, I follow Webb (2019) and extract verb-object pairs from each O*NET task description. I then map the categorical exposure classifications into a continuous measure by assigning each task a score between 0 and 1, following Eloundou et al. (2023). Finally, I aggregate exposure to the verb-object pair level by averaging across all tasks containing a given pair.

This procedure produces a measure of exposure that captures the extent to which individual task components can be performed more efficiently using LLMs. Unlike patent-based measures, which reflect the distribution of capabilities embodied in technological innovations, the LLM measure directly reflects the ability of currently available models to substitute for or augment specific task inputs. For this reason, I treat the LLM-based measure as my preferred specification in the behavioral analysis of Section 6, where the relevant question is how exposed coursework is to the tools that students and instructors are already using. The patent-based AI measure captures the broader frontier of AI-related capabilities and serves as an alternative specification and robustness check.[10]

2.3 Supplemental data sources

The validation exercises in Section 4 draw on two additional sources. First, I use occupational task descriptions from O*NET, a database maintained by the US Department of Labor that summarizes worker attributes and job characteristics for nearly 1,000 occupations.[11] I use O*NET ratings of task importance and relevance to weight verb-object pairs in constructing exposure measures. Second, to link fields of study to occupations, I use field-occupation mappings estimated from the American Community Survey (ACS), restricting to employed workers ages 25-65 with a four-year degree using surveys from 2009-2019. Both sources are used exclusively for validation and do not enter the main exposure or behavioral results.

3 Methodology

The core methodology for this project involves three steps: extracting "tasks" (verb-object pairs) from course descriptions and patent titles; constructing distributions of these tasks for each field of study and technology class; and calculating an overlap statistic that measures

[10]The LLM measure also has a conceptual advantage over many existing exposure measures. Because a task is coded as exposed if a *single* LLM capability substantially reduces the time required to complete it, exposure does not require overlap with many distinct AI capabilities. Patent-distribution approaches such as Webb (2019) and ability-based approaches such as Felten et al. (2021) similarly place greater weight on tasks that map to multiple AI capabilities, which is not obviously the right restriction when a single capability is sufficient to automate or substantially assist a task.

[11]For example, an O*NET task for Economists is "Compile, analyze, and report data to explain economic phenomena and forecast market trends, applying mathematical models and statistical techniques."

each field's exposure to each technology. This exposure score captures the degree to which the core tasks students perform in their courses are tasks that new technologies can also perform. In the second part of the paper, field-level exposure scores serve as the treatment intensity in a difference-in-differences framework, allowing me to ask how instructor course design and student enrollment respond to ChatGPT across fields that differ in how much of their task content overlaps with LLM capabilities.

Throughout, exposure is defined at the verb-object pair level and then aggregated to the course and field level; all reported results reflect these aggregated measures.

3.1 Defining and extracting tasks

The core challenge in working with course description text is isolating the economically relevant content from surrounding noise. A course description contains at least three distinct types of information: administrative boilerplate (prerequisites, credit hours, enrollment restrictions), pedagogical framing (learning objectives, teaching philosophy, course rationale), and task content (what students are specifically trained to do). The exposure measure is designed to capture the last category only. The first two types of content, if included, would either add noise — administrative phrases have no task content — or introduce systematic bias, since pedagogical framing language may vary predictably across institution types in ways unrelated to actual task emphasis.

I define a *task* as a verb-object pair. For example, consider the description for Stanford's COMM 274D (Public Affairs Data Journalism II): "Learn how to find, create and analyze data to tell news stories with public service impact. Uses relational databases, advanced queries, basic statistics, and mapping to analyze data for storytelling..." (Stanford 2025). The description includes multiple verb-object pairs: "find data," "create data," "analyze data," "tell stories," "use databases," "use queries," "use statistics," and "use mapping." These pairs provide a concise representation of the main activities taught in the course. The motivation for the analysis in this paper is that if patents for new technologies also perform these tasks, the course content itself — and therefore the human capital students develop in the course — may be exposed to potential automation or augmentation by that technology.

Given the large size of the text corpus and the complexity of natural language, manual review is not feasible. Instead, I use common Natural Language Processing (NLP) techniques to clean and parse the text at scale. I pre-process the text by removing punctuation and standardizing capitalization. Next, I extract verb-object pairs using the `spaCy` package in Python, a widely used NLP tool (Honnibal et al. 2020). I then remove pairs in which the verb or object is a stopword[12] and lemmatize the verb and object to a common form (e.g.,

[12] For example, "use" or "have," which lack meaningful economic content (Jurafsky and Martin 2014).

"analyzing" and "analyzes" become "analyze").

Course descriptions are noisier than patent titles or O*NET tasks, as they often include administrative details or generic phrases that are not relevant for measuring the human capital students develop in these courses. To reduce noise, I take two steps. First, I remove "boilerplate" sentences, such as prerequisites and enrollment restrictions, using a rule-based filter applied to sentence structure and content. Second, I calculate chi-squared statistics for each verb-object pair, comparing occurrence frequency in course descriptions versus patent texts. I drop the 100 pairs with the highest chi-squared statistics; these pairs appear disproportionately in course descriptions relative to patents, indicating they reflect the conventions of the course-description document type rather than meaningful task content.[13]

As a final filter, I restrict the task dictionary to verb-object pairs that appear in at least one patent title or O*NET task description. This step removes any pairs not demonstrably relevant to technological change or labor market tasks, ensuring the exposure score is grounded in content with external validity. The step is conservative — it removes pairs from course descriptions that do not appear elsewhere — but guards against idiosyncratic course catalog language inflating exposure scores.

The final dictionary includes 27,612 unique verb-object pairs.

3.2 Exposure measure

I calculate the exposure of a field of study s to a technology class t as the frequency-weighted average of the tasks students perform in that field, weighted by the importance of those tasks to the technology. Let D denote the dictionary of verb-object pairs d. For field s, the share of task d is:

$$x_s(d) = \frac{\sum_{\delta \in D_s} \mathbb{I}(\delta = d)}{|D_s|} \tag{1}$$

where D_s is the set of all task observations in field s. For example, in Economics courses offered in 2023-24, "analyze data" accounts for 1.8% of all verb-object pairs.

Similarly, for each technology class t, I compute the share $w_t(d)$ of task d among all tasks mentioned in patents for that technology:

$$w_t(d) = \frac{\sum_{\delta \in D_t} \mathbb{I}(\delta = d)}{|D_t|} \tag{2}$$

The exposure of field s to technology t is then the inner product of the two task distributions:

[13] For example, "introduce student" or "receive credit."

$$exposure_{s,t} = \sum_{d \in D} x_s(d) \cdot w_t(d) \tag{3}$$

Intuitively, this measure captures how much of the "task footprint" of a given field overlaps with the tasks that a given technology is already able to perform. Higher values of $exposure_{s,t}$ indicate that a larger share of the tasks emphasized in a field's coursework are tasks that new technologies claim to be able to do. The primary specification uses raw frequency weights for both $x_s(d)$ and $w_t(d)$.[14] The LLM-based exposure measure follows the same framework but replaces the patent-derived technology weights $w_t(d)$ with the task-level LLM exposure score from Eloundou et al. (2023), aggregated to the verb-object pair level as described in Section 2.2.2.

In the measurement section that follows, I compute $exposure_{s,t}$ at the field level by pooling verb-object pairs across all institutions offering courses in field s. In later analyses, I extend the exposure calculation to field-by-institution cells, allowing exposure to vary within a field across institution types. In the behavioral analysis of Section 6, I normalize each field's LLM exposure score to have mean zero and standard deviation one, so that coefficients on exposure interactions have a natural per-standard-deviation interpretation. The normalized score enters the difference-in-differences specifications as a continuous treatment intensity, capturing how the relationship between enrollment or grade component weights and time changes differentially for fields at different points in the exposure distribution.

4 Validation

I conduct three validation exercises to confirm that the task-based exposure measures capture meaningful variation across fields of study. The first tests whether fields with similar task content cluster together in task space, independent of any exposure scores (construct validity). The second tests whether the most common verb-object pairs in each field align with intuitive descriptions of what those fields teach (face validity). The third tests whether field task profiles predict real-world occupational flows, providing a labor-market anchor for the exposure measure (external validity). Taken together, these exercises support the view that the verb-object pair framework produces measures of human capital content that are internally consistent, substantively interpretable, and relevant to the labor market.

[14] In Figure 5, I calculate exposure scores at the institution-field-year level. Because these scores rely on fewer verb-object pairs, the estimates are inherently noisier. When working at the institution-field level, I weight observations by the number of verb-object pairs to reduce the influence of noise.

4.1 Task-space similarity across fields

The most direct test of whether the verb-object pair extraction is capturing genuine disciplinary structure is to ask whether fields cluster in task space the way disciplinary knowledge suggests they should — without using the exposure scores themselves. Figure 2 presents this exercise. For each field of study, I construct a vector of verb-object pair frequencies from 2022-23 course descriptions, apply TF-IDF weighting to downweight pairs that appear uniformly across fields, reduce the resulting high-dimensional representation to 50 dimensions via truncated singular value decomposition, and project to two dimensions using multidimensional scaling. Fields that are positioned close together share similar task content; fields that are far apart emphasize meaningfully different tasks.

The resulting map recovers disciplinary structure with high fidelity. The quantitative sciences (e.g., Mathematics, Physics, Chemistry, and Statistics/Data Science) cluster tightly in the upper portion of the figure, reflecting their shared emphasis on tasks such as "solve equations," "analyze data," and "model systems." Computer Science and Engineering sit nearby but are clearly separated, consistent with their more applied orientation. At the opposite end, the Humanities (e.g., English, History, Philosophy, Religion, and Ethnic/Cultural Studies) cluster to the left, unified by tasks such as "read texts," "write essays," and "examine arguments." Social Science fields occupy an intermediate position, consistent with their mix of writing-intensive and quantitative tasks. Economics is notably separated from both the core Humanities and core quantitative clusters, sitting closer to the quantitative fields, a pattern that reflects the increasing emphasis on data analysis and empirical methods in economics coursework. Health and professional fields (Nursing, Medicine, Pharmacy, Rehabilitation) form their own cluster to the right, sharing a vocabulary of clinical and applied tasks.

Two features of the map are especially relevant for interpreting the exposure results that follow. First, the fields with the greatest pairwise task similarity are those within the same broad disciplinary category, while fields across categories are substantially more distant — exactly what one would expect if verb-object pair extraction is recovering genuine task content rather than surface-level textual noise. Second, the fields that are most distant in task space (e.g., Mathematics and English) are also the fields whose task profiles diverge most sharply in Table 1, providing a consistency check across the two validation exercises.

4.2 Distinctive tasks by field of study

To further demonstrate how the task extraction approach captures variation across fields, Table 1 summarizes the most common verb-object pairs for a representative sample of fields.

For each field, I select the three verbs that appear most frequently and, for each verb, list the (up to) four most frequent objects.

The table highlights distinctive and intuitive task patterns. English courses emphasize tasks such as "read texts" and "write essays"; Education courses emphasize "teach students"; Arts courses[15] emphasize "create images" and "examine art." These patterns are what one would expect from knowledge of these fields' subject matter, and they align closely with the clustering structure visible in Figure 2. At the same time, some tasks — particularly those involving data analysis — appear across multiple fields, reflecting the growing role of empirical methods across disciplines. The task "analyze data" appears prominently not only in Statistics and Economics, but also in Social Science and, increasingly, in Business and Health fields.

4.3 Linking field tasks to occupation tasks

The third validation exercise asks whether the tasks extracted from course descriptions for a given field also appear in the occupations that graduates from that field typically enter. This provides an external check that the task-based measure captures labor-market-relevant content rather than idiosyncratic features of course catalog language.

To do so, I use field-to-occupation mappings from the ACS for prime-age workers (ages 25-65).[16] I construct task vectors for each occupation and each field of study, based respectively on the importance-weighted verb-object pairs in O*NET and their frequency-weighted share in course descriptions. I restrict the task space to pairs that appear in at least one O*NET task description, so the comparison is grounded in occupationally relevant tasks. I then measure the similarity between each field-occupation pair using the cosine distance in this shared task space. After dropping field-occupation pairs with an empirical ACS share below 0.01% and no task overlap,[17] I compute the rank-rank correlation between the empirical field-to-occupation shares and the task-based similarity rankings.

The resulting rank-rank correlation is 0.31. To benchmark this figure: under the null that field task profiles carry no information about occupational flows, the correlation would be near zero. At the other extreme, a correlation near one would require O*NET task lists to be dense and comprehensive, and field-to-occupation flows to be fully determined by task similarity, neither of which holds in practice. O*NET task lists are sparse (typically five to fifteen tasks per occupation), many fields flow into multiple occupations with overlapping

[15]The Arts category includes courses in Studio Arts, Music, Theater, Art History, and Dance.

[16]A robustness exercise using only early-career workers (ages 25-35) produces similar field-to-occupation mappings.

[17]This restriction is conservative: it excludes pairs where both the empirical share and the task-based similarity are near zero, which would mechanically inflate the correlation.

task content, and some occupational choice reflects factors unrelated to task match (earnings, preferences, geography). Given these constraints, a rank-rank correlation of 0.31 constitutes meaningful evidence that the verb-object pair profiles extracted from course descriptions reflect the same task content that characterizes the jobs graduates enter. Statisticians, Software Developers, and Data Scientists rank near the top of the AI distribution; Manual Laborers and Construction workers appear rank the top of the Robot distribution; and the rankings are reversed at the bottom of each.[18]

Taken together, the three exercises establish that the verb-object pair-based exposure measure has internal consistency, face validity, and external labor market relevance. The task map in Figure 2 confirms that the extraction procedure recovers genuine disciplinary structure. The distinctive-task table confirms that the most common pairs in each field align with substantive knowledge of those fields. And the field-to-occupation correlation confirms that the task profiles predict where graduates work. These properties provide the foundation for the cross-field and cross-institution comparisons in Section 5 and for the causal analysis of Section 6.

5 Higher education-technology exposure and implications

This section documents facts about curricular exposure to AI. First, college coursework is substantially more exposed to AI than to earlier waves of technological change. Second, AI exposure has risen steadily over the past decade, driven by the diffusion of data analysis tasks across disciplines. These findings motivate the behavioral analysis in Section 6: if exposure is high, growing, and concentrated in particular fields and institution types, then the release of ChatGPT provides a natural setting in which to ask whether instructors and students adjust their behavior in proportion to their exposure — or whether institutional and informational frictions suppress that response.

5.1 Overall overlap between courses and technology

Figure 3 summarizes field-level exposure to three technologies using course descriptions from the 2022-23 academic year. For each of 52 undergraduate fields of study, I pool verb-object pairs across institutions and calculate exposure to Robots, patent-based AI, and the LLM-based "OpenAI" measure. To facilitate comparison across technologies, I normalize each field-level score by the occupational exposure distribution for the same technology, so that

[18] For related evidence linking college major task content to occupational outcomes, see Javadian Sabet et al. (2024).

a score of zero means the field is as exposed as the average occupation and a score of one means the field sits one standard deviation above the mean occupational exposure.

The first panel shows that Robot exposure is low across nearly all fields of study. The most exposed fields are vocational and applied areas such as Skilled Trades, with modest additional overlap in Engineering, Agriculture, and some technical STEM fields. The typical field of study falls well below the mean level of occupational Robot exposure.

The next two panels show a different pattern for AI. Both the patent-based AI measure and the LLM-based measure indicate far greater exposure than Robots. The patent-based measure shows more dispersion across fields, with a long right tail among quantitative disciplines; the LLM measure shows a tighter distribution concentrated approximately half a standard deviation above the occupational mean, indicating that LLM-relevant capabilities are broadly diffused across the curriculum rather than concentrated in a handful of fields. This contrast reflects the measures' different scope: the patent-based measure captures overlap with the broader frontier of AI-related capabilities, while the OpenAI measure captures overlap specifically with the capabilities of deployed LLM tools.

Figure 4 complements the histograms by plotting the full exposure distribution as a CDF, with each field represented as a dot sized by enrollment share. Under the Robot measure, the CDF rises steeply near zero, while most fields cluster below the average occupation.[19] Under the patent-based AI measure, the distribution is more spread, with STEM fields at the upper end and Humanities and Skilled Trade fields at the lower end.[20] Under the LLM-based measure, Statistics/Data Science and Computer Science sit at the top of the distribution because of their emphasis on coding, data analysis, and writing, followed by English and Social Science fields, which combine writing tasks with varying degrees of quantitative work. The LLM-based AI curve is less steep than the patent-based AI curve, reflecting more uniform exposure across fields and fewer extreme outliers. Put differently, the patent-based measure captures the right tail of highly exposed quantitative fields most sharply; the LLM-based measure better captures the broad diffusion of LLM-relevant capabilities across the curriculum.

The central takeaway is straightforward: college coursework is far more exposed to AI, and to LLM capabilities specifically, than to earlier technology shocks. The pattern is consistent with the idea that modern AI differs from prior waves of automation in the extent to

[19]The somewhat elevated Robot exposure of STEM fields relative to non-STEM fields likely reflects a limitation of patent-based scoring: patentable activities overlap disproportionately with STEM subject matter, so some apparent alignment may reflect shared vocabulary rather than genuine task overlap.

[20]Reassuringly, the Statistician occupation sits approximately two standard deviations above the mean in the AI-patent occupational distribution, consistent with the high exposure of Statistics coursework coming through genuine task overlap rather than coincidental vocabulary.

which it overlaps with the cognitively demanding, language- and data-intensive tasks that universities teach.

5.2 Time trend in technology exposure

Figure 5 shows how exposure to each technology has evolved over the past two decades. I estimate these trends from regressions of institution-field-year exposure on year indicators, controlling for institution-by-field fixed effects. The estimates therefore isolate within-field changes over time, holding the composition of fields constant.

The clearest pattern is a steady rise in patent-based AI exposure beginning in the early 2010s and continuing through the 2020s, while Robot exposure remains low and flat throughout. The rise in patent-based AI exposure reflects the gradual diffusion of data analysis and related empirical tasks across fields — tasks such as "analyze data," "build model," and "apply method" appear with increasing frequency in course descriptions over this period, in fields ranging from Economics and Statistics to Public Health and Education.

5.3 Distributional implications

The preceding results imply that AI exposure is not evenly distributed across students. Figure 6 summarizes three dimensions of this heterogeneity using terciles of the LLM-based AI exposure distribution.

Three patterns are worth highlighting. First, highly exposed majors are disproportionately associated with higher earnings: average income rises monotonically across exposure terciles, reflecting the concentration of AI exposure in quantitative fields with strong earnings. This creates an ambiguity about the welfare consequences of exposure that the paper does not attempt to resolve: if AI augments productivity in high-exposure fields, returns to those fields may rise further; if AI substitutes for skill in those fields, the historically large earnings premium may erode.

Second, exposure is more concentrated among male graduates than female graduates, reflecting gender sorting across fields of study. Men are overrepresented in the high-exposure tercile relative to women, while women are overrepresented in the low-exposure tercile. To the extent that AI affects earnings differentially by exposure level, these patterns imply uneven incidence across genders.

Third, high-exposure majors are earned across a range of institution types, not exclusively at research universities. R1, R2, and liberal arts institutions all contribute meaningfully to the high-exposure tercile, while teaching-focused institutions are underrepresented there relative to their share of overall enrollment.

These descriptive patterns do not establish welfare effects; the exposure measure is intentionally agnostic about whether AI substitutes for or complements the skills it overlaps with. They do, however, clarify who stands to be most affected by whatever adjustment occurs. The concentration of exposure among higher-earning, male-skewing, and research-university fields shapes both the stakes and the incidence of the behavioral responses examined in Section 6.

6 Instructor and student responses to ChatGPT release

Having documented that college coursework is substantially and increasingly exposed to AI, this section examines how higher education responds to a sharp, exogenous technology shock: the public release of ChatGPT in November 2022. The rapid and widespread diffusion of LLM tools that followed provides a natural setting to assess whether instructors and students adjust their behavior in proportion to their exposure.

The release of ChatGPT has two immediate implications for higher education that motivate the analysis. First, it lowers the effective cost of completing coursework in tasks that overlap with LLM capabilities. Students can use these tools to augment or substitute for their own effort on out-of-class assignments — essays, problem sets, take-home assessments — in ways that were not previously possible at scale. This creates a direct problem for instructors who use such assignments to evaluate student learning: if AI can complete the assignment, the assignment no longer reliably signals what the student knows. Second, ChatGPT introduces uncertainty about the labor market returns to skills developed in AI-exposed courses. If employers come to view AI-substitutable skills as less valuable, or if AI tools make those skills more productive, students may rationally adjust their enrollment toward or away from high-exposure fields.

These two mechanisms generate a set of testable predictions that I evaluate in order. First, instructors should rapidly acknowledge the shock by adopting AI policies at much higher rates after ChatGPT than before. Second, if instructors seek to preserve the signal value of assessment, they should restructure course grading away from components which students can complete outside of class using AI tools (essays and homework, which I collectively label as "AI-substitutes") and toward components which require in-person demonstration of knowledge (exams, participation, and presentations, which I collectively label as "AI-non-substitute"). Third, students in fields with high AI exposure should adjust their enrollment patterns relative to students in low-exposure fields, with the direction of adjustment reflecting their net expectations about costs and returns.

I evaluate these predictions using event study specifications that compare within-course changes in outcomes before and after Fall 2022, exploiting variation in field-level LLM exposure to assess heterogeneity in responses.

6.1 Adaptation of course design

6.1.1 AI policy adoption

To document that instructors were cognizant of ChatGPT's release and concerned about its implications for course administration, Figure 7 plots the share of syllabi containing an explicit AI policy over time. The regression controls for course fixed effects, tracking within-course changes over time. The figure shows a discontinuity: prior to Fall 2022, fewer than 1% of syllabi contain any mention of AI tools, a level that had been stable for nearly a decade. By Fall 2025, more than half of all syllabi include an AI policy, representing a near-complete transformation of course administration norms within three academic years.

The majority of policies imposed restrictions or regulations on student AI use rather than permitting or encouraging it, consistent with instructors viewing AI as a threat to the integrity of existing grading practices rather than as a tool to be integrated into coursework. Appendix Figure A-2 shows that adoption rates rose across all fields but were highest in the most AI-exposed fields, particularly in the immediate post-ChatGPT period. By Fall 2025, adoption rates had converged across exposure terciles to around 60%, suggesting that the shock ultimately affected course administration practices broadly regardless of field exposure.

6.1.2 Assessment restructuring

The AI policy findings establish that instructors were aware of and concerned about ChatGPT. The natural follow-up question is whether that concern translated into substantive course redesign. To assess this, I estimate event study regressions of the form:

$$y^q_{i,s,\tau} = \alpha^q_{i,s} + \sum_{\tau \in \mathcal{T}} \beta^q_\tau + \varepsilon^q_{i,s,\tau},$$

where $y^q_{i,s,\tau}$ is the weight on grade component q in course i at institution s in semester τ, and $\alpha^q_{i,s}$ are course fixed effects that absorb any time-invariant differences across courses. The coefficients β^q_τ trace out changes in grading weights in each semester relative to Fall 2022, the last pre-ChatGPT term.

Figure 8 presents the results across six grade component categories. The most salient feature of the figure is not what changes after ChatGPT but what does not. In sharp contrast to the COVID-19 period — when courses shifted rapidly and dramatically away from exams and toward homework-based assessments beginning in Spring 2020, with effects visible as large discrete breaks in the figure — the post-ChatGPT period shows only gradual and modest adjustments. Essay weights decline slowly over the post-2022 period; participation weights increase modestly. By Fall 2025, the decline in essay weight corresponds to a 11% change relative to the Fall 2022 baseline, and increases in participation and presentation weights account for a 7% change. These magnitudes are economically modest relative to the baseline variation in grading schemes across fields and institutions, and relative to the shifts observed during COVID.[21]

Importantly, the direction of the modest post-ChatGPT shift is consistent with the prediction: instructors are moving weight from AI-substitute toward AI-non-substitute components, as the theory predicts. The COVID pandemic posed a comparable challenge to grade component integrity — remote instruction also made it difficult to verify student work — and instructors responded immediately and dramatically. The ChatGPT period poses a structurally similar challenge to grade component integrity through out-of-class assignments, yet the response is an order of magnitude smaller and slower.

Heterogeneity analysis suggests that, to the extent instructors are adjusting course administration, the adjustments are concentrated in fields where tasks overlap most with AI's capabilities and at more research-intensive institutions. Appendix Figures A-4 and A-3 plot estimates of separate interrupted time series regressions, splitting courses by field-level AI exposure and R1 vs non-R1 classification, respectively. The figures suggest small increases in AI non-substitute shares in fields with the highest tercile of AI exposure. Fields in the middle tercile, on the other hand, decreased the weight on AI non-substitutable evaluations in 2023-24. Similarly, I detect a slight increase in the weight on AI non-substitutable grade components at R1 institutions, but levels remain far below pre-pandemic levels.

6.1.3 Exam modality

The grade component weight results may understate the full extent of course restructuring if instructors respond along the intensive margin of exam format rather than the extensive margin of grade weights. Figure 9 presents evidence on this channel, plotting the share of syllabi describing an unproctored (take-home, open-book, or web-based exam) over time.[22] The

[21] Given that the majority of schools in the syllabus sample are in Texas, which offered in-person instruction in Fall 2021, the COVID-era estimates likely understate the national magnitude of pandemic-era course restructuring.

[22] The share of courses with an unproctored or out-of-class exam ranges from 5-15% of courses in 2022-23. Unproctored exams are more common in upper-level courses and at more selective institutions. For this analysis, I restrict to courses offering in-person instruction. Web-based courses face different constraints in changing the exam modality, and the exam modality is often not indicated in the syllabus. At most of the universities in the syllabus sample, online courses represent a relatively small share of total course offerings; however, for some of the universities, online course shares increased during the pandemic and have remained

pattern is instructive: the share of unproctored exams rises sharply during COVID (consistent with the shift to remote administration) but then declines meaningfully after ChatGPT, falling below the pre-pandemic level by 2024-25. It is worth noting that the response size, although precisely estimated and in the expected direction, is relatively small and may reflect mean reversion from a post-pandemic spike as much as a response to ChatGPT's release.

This decline is not captured by the grade component weight analysis, since the weight assigned to exams need not change when the format shifts from take-home to in-person. The modality evidence thus provides a complementary indicator of course restructuring: instructors are modestly pulling exams back toward proctored, in-person formats in the post-ChatGPT period, even as the weight assigned to exams remains roughly stable.

Taken together, the policy adoption, grade component weight, and exam modality findings tell a consistent story. Instructors responded quickly along low-cost margins — adding AI policies requires minimal course redesign — and have made modest but generally directionally consistent adjustments along higher-cost margins, shifting grade weight away from AI-substitutable assignments and moving exams toward formats that are harder to complete with AI assistance. The scale of these adjustments, however, is substantially smaller than the response observed during the COVID-19 pandemic, despite similar challenges to grade component integrity.

6.2 Course taking

Figure 10 provides context for the enrollment analysis by plotting the long-run trend in enrollment shares by broad field category from 2000 to 2025. Several features of the pre-ChatGPT period are relevant for interpreting the post-shock results. Computer Science enrollment share rose explosively from the mid-2000s through the early 2020s, while Humanities enrollment share declined steadily over the same period. These divergent trends were well underway before ChatGPT and reflect longer-run structural shifts in student demand that the event study framework absorbs through its pre-period controls.

To assess whether ChatGPT accelerated or altered these trends differentially by field AI exposure, I estimate event study models of enrollment at the institution-field-semester level. I interact year indicators with each field's standardized LLM exposure score and control for institution-by-field, institution-by-year, institution-by-semester, and field-by-semester fixed effects, so identification comes from within-institution, across-field variation in exposure over time.[23] The coefficient in each year captures the differential change in log enrollment associated with a one standard deviation increase in field-level AI exposure, relative to the

elevated relative to pre-pandemic levels.

[23]Results are substantively similar using the patent-based AI exposure measure.

2022-23 reference year.

Figure 11 presents the results. Pre-ChatGPT coefficients are small and statistically indistinguishable from zero, supporting the parallel trends assumption: fields at different points in the AI exposure distribution were not already on divergent enrollment trajectories before the shock. Beginning in 2023–24, coefficients turn negative and grow in magnitude, reaching approximately -1.5 log points per standard deviation of exposure by 2025-26. While statistically significant, these effects are economically modest: a one standard deviation difference in AI exposure, roughly the gap between Sociology and Computer Science, is associated with about a 1.5 log point differential decline in enrollment. The effect is also growing, and it is too early to determine whether it will stabilize or accelerate.

The aggregate coefficient masks meaningful heterogeneity in what is driving the decline. Longer-run enrollment shifts away from writing-intensive fields, particularly English, are partially responsible and predate ChatGPT, as visible in Figure 10. The more novel development in the post-ChatGPT period is the reversal in Computer Science enrollment: 2025-26 marks the first decline in Computer Science enrollment share in twenty years, halting a trend of explosive growth. Offsetting these declines are enrollment increases in fields lower in the AI exposure distribution (Business, Engineering, Biology, and Skilled Trades).

The direction of the enrollment response is informative even though its magnitude is small. Under a revealed preference interpretation, the negative post-ChatGPT coefficient is consistent with students updating downward on the labor market returns to AI-exposed skills, anticipating that skills which AI can perform may be less valuable to develop. However, the sign of the effect is not unambiguous: high AI exposure also lowers the cost of completing coursework, which would tend to increase enrollment in high-exposure fields. The negative coefficient suggests that the return-updating channel dominates the cost-reduction channel in net, but the two cannot be separated with the available data.

6.3 Discussion

The results across the instructor and student analyses point in the same direction: adjustment to the ChatGPT shock has been real but limited. On the instructor side, AI policy adoption has been rapid and near-universal. Assessment restructuring has been slow, modest in magnitude, and relatively uniform across fields — response heterogeneity by extent of field-level exposure or across institutions of differing responses is directionally intuitive but not economically meaningful. On the student side, enrollment has begun to shift away from high-exposure fields, but the effects are small enough that the overall distribution of students across fields looks much the same in 2025-26 as it did in 2022-23.

The contrast with the COVID-19 pandemic is instructive. Figure 8 makes the comparison vivid: the 2020-21 course restructuring — a shift of roughly 2.1 percentage points in exam weight and 1.8 percentage points in homework weight — occurred over a single semester and persisted for years. The ChatGPT period shows changes of a fraction of that magnitude spread over three academic years. This asymmetry is not simply a matter of time: by Fall 2025, three years had elapsed since ChatGPT's release, roughly the same interval as between the pandemic onset and Fall 2023. The scale of response remains categorically different.

Why does the COVID shock generate rapid, large responses while the ChatGPT shock generates slow, small ones, despite both posing challenges to the integrity of out-of-class assessment? Three mechanisms may explain the gap, and the data speak to each differently. The first is the nature of the constraint. COVID made in-person instruction infeasible, forcing immediate and unambiguous action. ChatGPT makes AI-assisted out-of-class work easier but does not eliminate any existing mode of instruction. Instructors can continue teaching and assessing in the same way they always have; the problem is that out-of-class assignments are now less informative, not that they are impossible. This asymmetry in urgency likely accounts for much of the difference in response speed. The policy adoption data support this: instructors acknowledged the ChatGPT problem immediately (60% with AI policies by Fall 2025), but acknowledging a problem is much lower-cost than solving it.

The second is uncertainty about the long-run role of AI. Instructors may be uncertain whether AI tools will become standard professional equipment — in which case integrating them into coursework would be appropriate — or whether they will undermine the value of the skills currently taught — in which case restriction and restructuring would be warranted. This uncertainty justifies a wait-and-see strategy, particularly for instructors at the beginning of multi-year curriculum revision cycles. The near-uniform AI policy adoption across high- and low-exposure fields (Appendix Figure A-2) is consistent with this: policies signal awareness but do not commit to a substantive position.

The third is the cost of redesign. Restructuring grading schemes and developing new assessments is time-consuming, particularly for large courses with many sections. The modest differences by field-level exposure documented in Appendix Figure A-4 — showing only modest differential response by AI exposure — is notable here: if redesign costs were the binding constraint, one might expect high-exposure instructors to be most motivated to bear those costs. The similarity of responses instead suggests that constraints are not specific to high-exposure fields but are broadly shared across the curriculum.

Together, these mechanisms suggest that higher education's limited response to the ChatGPT shock is not primarily a failure of awareness or motivation, but reflects the genuinely different nature of the challenge AI poses relative to prior shocks. The adjustment may accelerate as uncertainty resolves and as the costs of redesigning courses decline through experience and shared resources. Whether it will do so fast enough to keep pace with continued AI capability growth is an open question that the data available through 2025-26 cannot yet answer.

7 Conclusion

This paper provides the first systematic evidence on the exposure of college curricula to artificial intelligence and on how instructors and students respond to a sharp AI technology shock. Building on the task-based framework of Webb (2019), I map the capabilities embodied in AI patents and in deployed LLM tools onto the task content of courses at more than 1,000 US colleges and universities, constructing field- and institution-level exposure scores that span two decades.

The measurement results establish three facts. First, college coursework is far more exposed to AI, and to LLM capabilities in particular, than to earlier waves of technological change. The typical field of study sits near the 75th percentile of the occupational AI exposure distribution, while Robot exposure is generally below the average occupation. Second, AI exposure has risen steadily since the early 2010s, driven by the diffusion of data analysis and empirical methods across disciplines — a trend that predates ChatGPT by more than a decade. Third, exposure varies substantially across institution types within the same field: R1 universities offer more AI-exposed curricula than teaching-focused institutions and community colleges, reflecting differences in course menus and curricular depth rather than differences in how individual courses are taught. These patterns imply that the "same" degree can correspond to meaningfully different task content depending on where it is earned.

The behavioral results reveal a gap between where the measurement of learning is most threatened and whether the instruments that measure it adapt. On the instructor side, AI policy adoption has been rapid and near-universal: fewer than 1% of syllabi mentioned AI tools before ChatGPT; more than half did by Fall 2025. But substantive course restructuring that would limit students' ability to substitute AI for their own work has been modest. Grade component weights shifted only marginally — essay weights declined slightly, participation weights rose slightly — with no differential response in high-exposure fields where the concern is most acute. The share of courses with unproctored exams declined modestly after ChatGPT, suggesting instructors are pulling assessments back toward in-person formats even when they are not dramatically reweighting grades, though unproctored or out-of-class exams remain more common in 2025-26 than they were in the year before the pandemic. These adjustments stand in sharp contrast to the large and immediate changes in course structure observed during the COVID-19 pandemic, despite the two episodes posing structurally similar challenges to grade component integrity.

On the student side, enrollment in high-AI-exposure fields has declined relative to low-exposure fields, but the effects are economically small — around 1.5 log points per standard deviation of exposure by 2025-26 — and growing slowly. Although enrollment shares in Computer Science declined in 2025-26 for the first time in nearly twenty years, the dip was not nearly as large as the sharp decline in Computer Science enrollment in the early 2000s, following the Dot-com bubble.

Together, these findings suggest that higher education's response to AI is constrained less by awareness than by the nature of the challenge itself. Unlike COVID, which made existing modes of instruction infeasible and forced immediate action, AI makes existing grading practices less informative without eliminating them. This distinction, combined with genuine uncertainty about whether AI will ultimately substitute for or complement the skills universities teach, creates conditions in which a wait-and-see posture is individually rational even as it may be collectively slow.

Two directions for future research follow naturally from these findings. First, the exposure measures constructed here — covering 52 fields, more than 1,000 institutions, and 25 years of course offerings — provide a reusable benchmark for studying how AI affects the returns to human capital investment. Linking these measures to wage and employment outcomes for graduates of high- versus low-exposure fields would allow researchers to assess whether AI is already affecting the labor market returns to college majors, and whether the modest enrollment responses documented here reflect rational updating or insufficient information. Second, a central open question is whether the gradual adjustment documented here will persist or accelerate. As AI capabilities continue to expand and as the costs of curriculum redesign fall with experience and shared resources, the response may look very different in 2028 than in 2025. Whether higher education can adapt quickly enough to remain a reliable signal of human capital in an era of capable AI tools is among the most consequential questions facing the sector.

Works cited

Acemoglu, Daron et al. (2022). “Artificial intelligence and jobs: Evidence from online vacancies”. In: *Journal of Labor Economics* 40.S1, S293–S340.

Autor, David H, Lawrence F Katz, and Melissa S Kearney (2006). “The polarization of the US labor market”. In: *American economic review* 96.2, pp. 189–194.

Autor, David H, Frank Levy, and Richard J Murnane (2003). “The skill content of recent technological change: An empirical exploration”. In: *The Quarterly journal of economics* 118.4, pp. 1279–1333.

Biasi, Barbara and Song Ma (2022). *The education-innovation gap.* Tech. rep. National Bureau of Economic Research.

Brynjolfsson, Erik and Tom Mitchell (2017). “What can machine learning do? Workforce implications”. In: *Science* 358.6370, pp. 1530–1534.

Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock (2018). “What can machines learn and what does it mean for occupations and the economy?” In: *AEA papers and proceedings.* Vol. 108. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 43–47.

Chirikov, Igor (2026). “How Instructors Regulate AI in College: Evidence from 31,000 Course Syllabi”. In: *Higher Education Working Paper Series* 26.1.

Eloundou, Tyna et al. (2023). “Gpts are gpts: An early look at the labor market impact potential of large language models”. In: *arXiv preprint arXiv:2303.10130* 10.

Felten, Edward, Manav Raj, and Robert Seamans (2021). “Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses”. In: *Strategic Management Journal* 42.12, pp. 2195–2217.

Honnibal, Matthew et al. (2020). “spaCy: Industrial-strength natural language processing in python”. In.

Javadian Sabet, Alireza et al. (2024). “Course-Skill Atlas: A national longitudinal dataset of skills taught in US higher education curricula”. In: *Scientific Data* 11.1, p. 1086.

Jurafsky, Dan and James H Martin (2014). *Speech and language processing. Vol. 3.*

Light, Jacob (2026). “Student demand and the supply of college courses”. In: *Available at SSRN 4856488.*

McElheran, Kristina et al. (2024). “AI adoption in America: Who, what, and where”. In: *Journal of Economics & Management Strategy* 33.2, pp. 375–415.

Salitel, Fernando (2026). “Openness to AI in Higher Education”. In.

Timmerman, Jean Xiao (2025). “Educational Exposure to Generative Artificial Intelligence”. In: *FEDS Notes* 2025-02, pp. 26–2.

Webb, Michael (2019). “The impact of artificial intelligence on the labor market”. In: *Available at SSRN 3482150.*

Figures and Tables

Figure 1. Example course

ECON 101
Lecture: 01
Units: 5
Class#: 7559

Winter 2026
Open

Economic Policy Seminar

Department of Economics

Lecture, Discussion
1/5/26 - 3/13/26

Mon, Wed
1:30 PM - 3:20 PM
Econ 140

Instructor: Light, J.

Enrollment Status

Open Seats: 7
Enrolled: 17 Waitlist: 0
Capacity: 24 Waitlist Max: No waitlist

Course Description

Capstone and writing in the major course open to Econ majors only. Economic policy analysis, writing and oral presentations will be large components of this course. Students may also complete group projects that include empirical economic analysis focused on a specific topic. The goal of this course is to enable students to utilize the skills they have acquired throughout their time in the major. Section topics vary by instructor. Enrollment limited to application at the start of each school year with student placement notifications before the term starts. Permission numbers will be provided to students. Limited to students applying to graduate in 2025-26. Enrollment by application: https://economics.stanford.edu/forms.

Grading basis
Letter (ABCD/NP)

Class level
Undergraduate

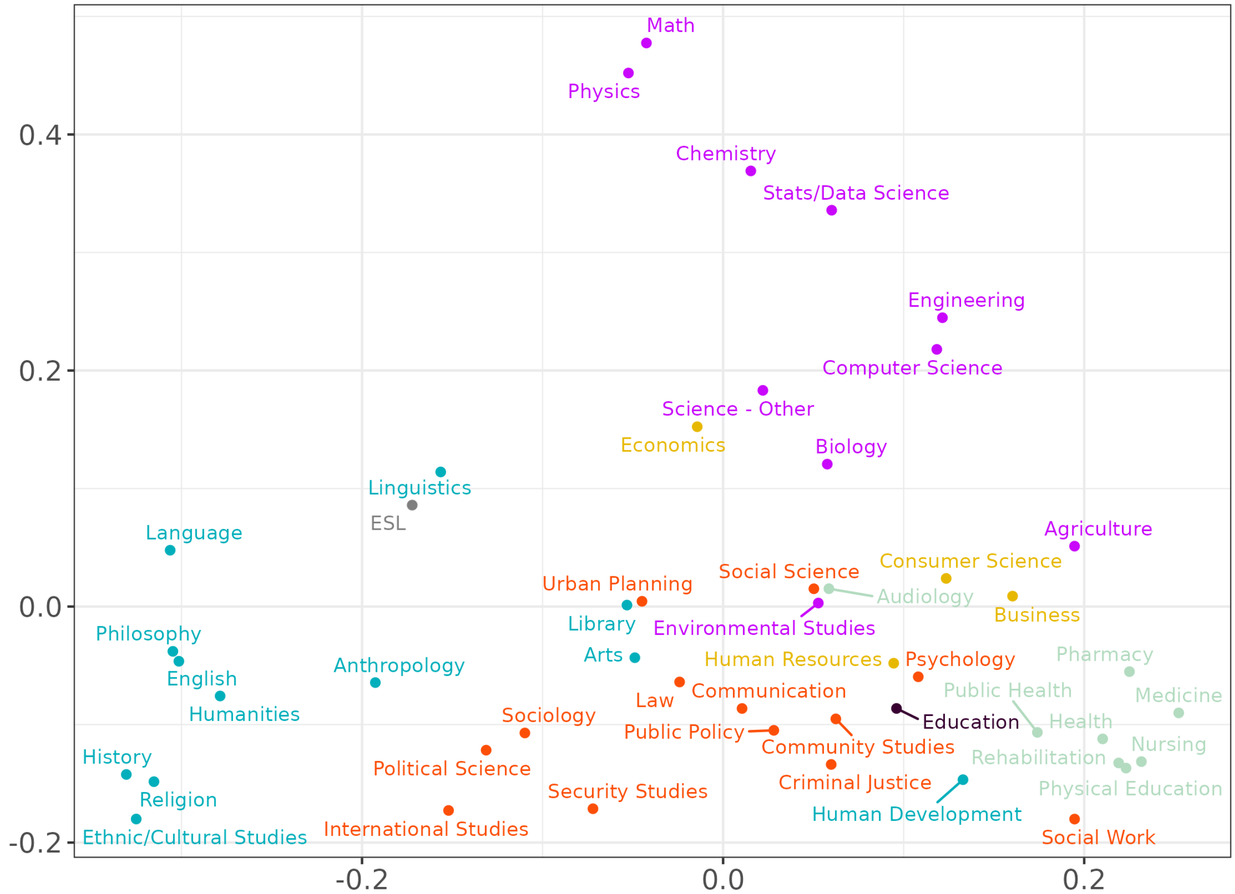
Instructional mode
In Person

Final exam
Wed, March 18th
3:30 PM - 6:30 PM
Econ 140

Gen Ed Requirement(s)
WAYS - Social Inquiry (SI)

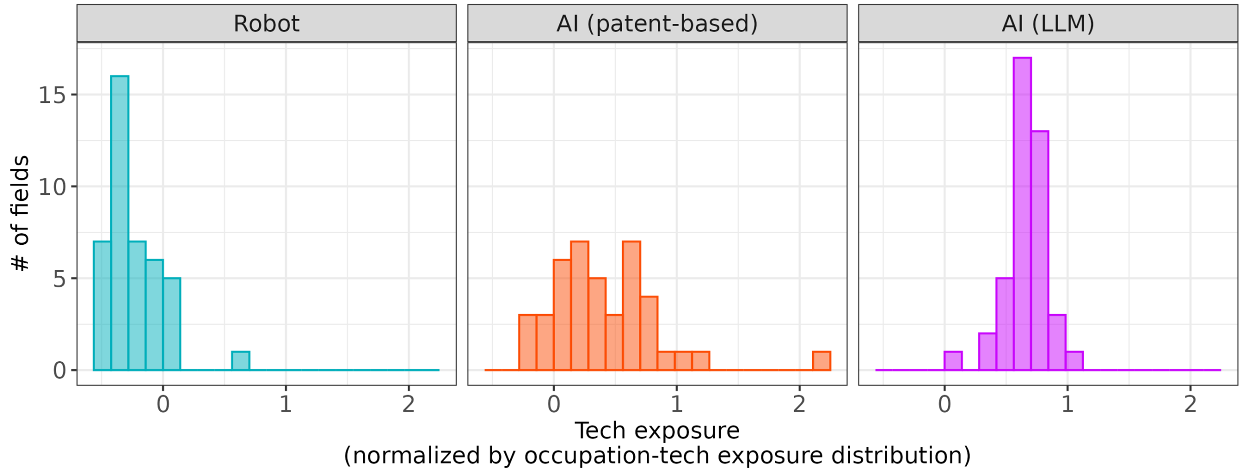
Source: Stanford University.

Figure 2. Verb-object pair relation between fields



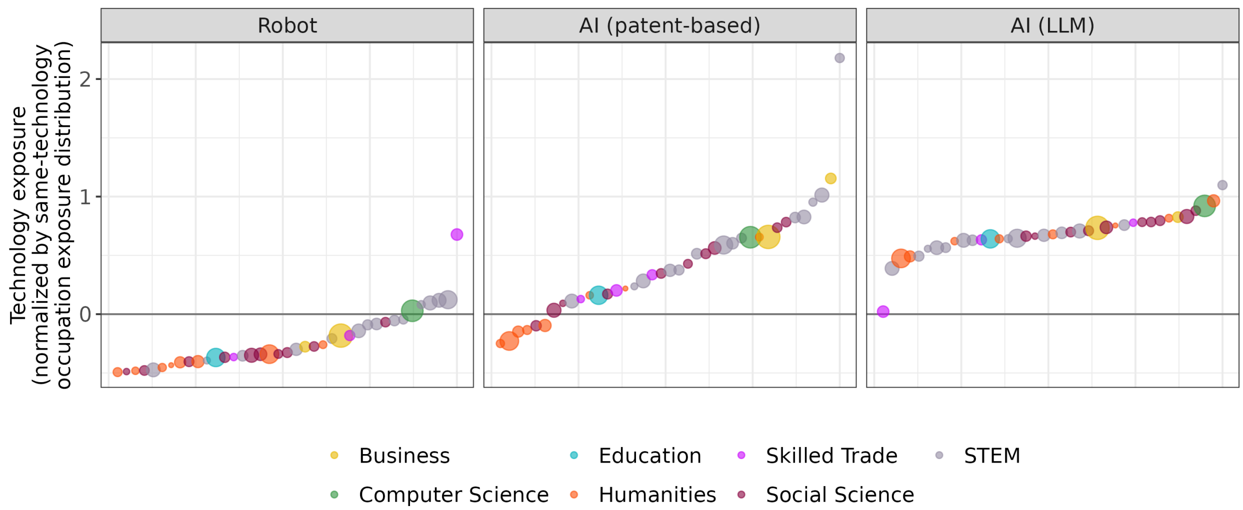
Notes: The figure plots the two-dimensional representation of fields of study based on the similarity of their verb-object pair distributions. Each point is a field of study, colored by broad disciplinary category. Distances are computed using cosine similarity on TF-IDF-weighted verb-object pair frequency vectors constructed from course descriptions in 2022-23, reduced to 50 dimensions via truncated singular value decomposition, and projected to two dimensions using multidimensional scaling. Fields that are closer together have more similar task content; fields that are farther apart emphasize more distinct tasks.

Figure 3. Technology exposure distribution across fields



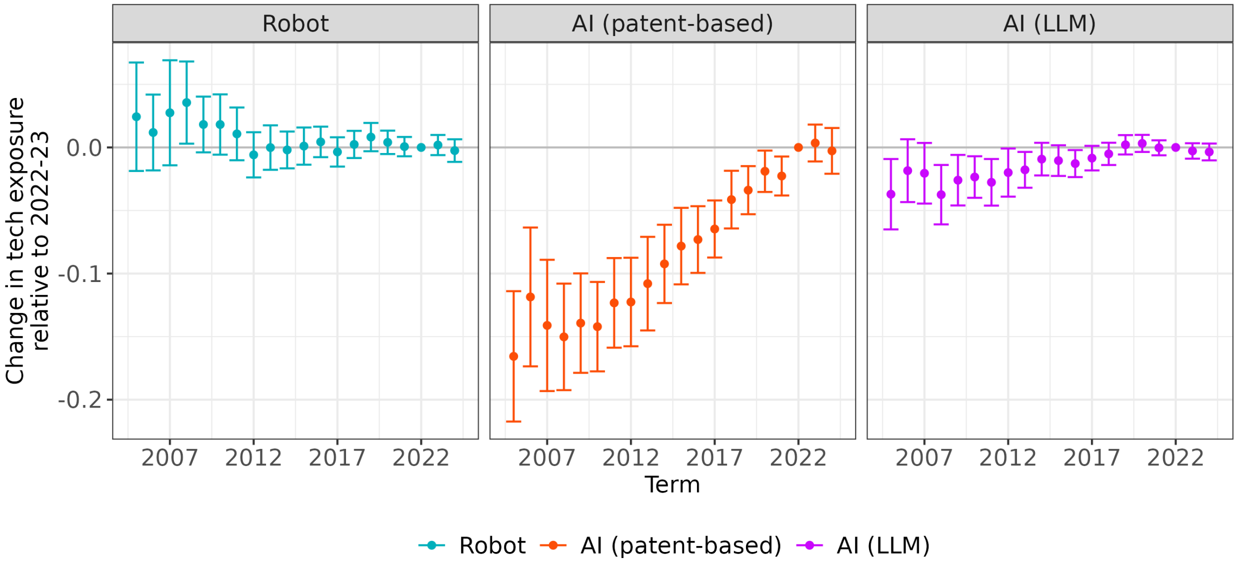
Notes: The figure plots density plots of field-technology exposure scores, split by technology class. "Robot" and "AI" measure patent-based overlap using the Webb (2019) exposure methodology; the "OpenAI" score measures overlap based on LLM capabilities measured by Eloundou et al. (2023). Courses are partitioned into 52 fields (e.g., Math, Business, History). Scores are normalized based on the occupation distribution of the same technology.

Figure 4. Technology exposure trend across fields - CDF



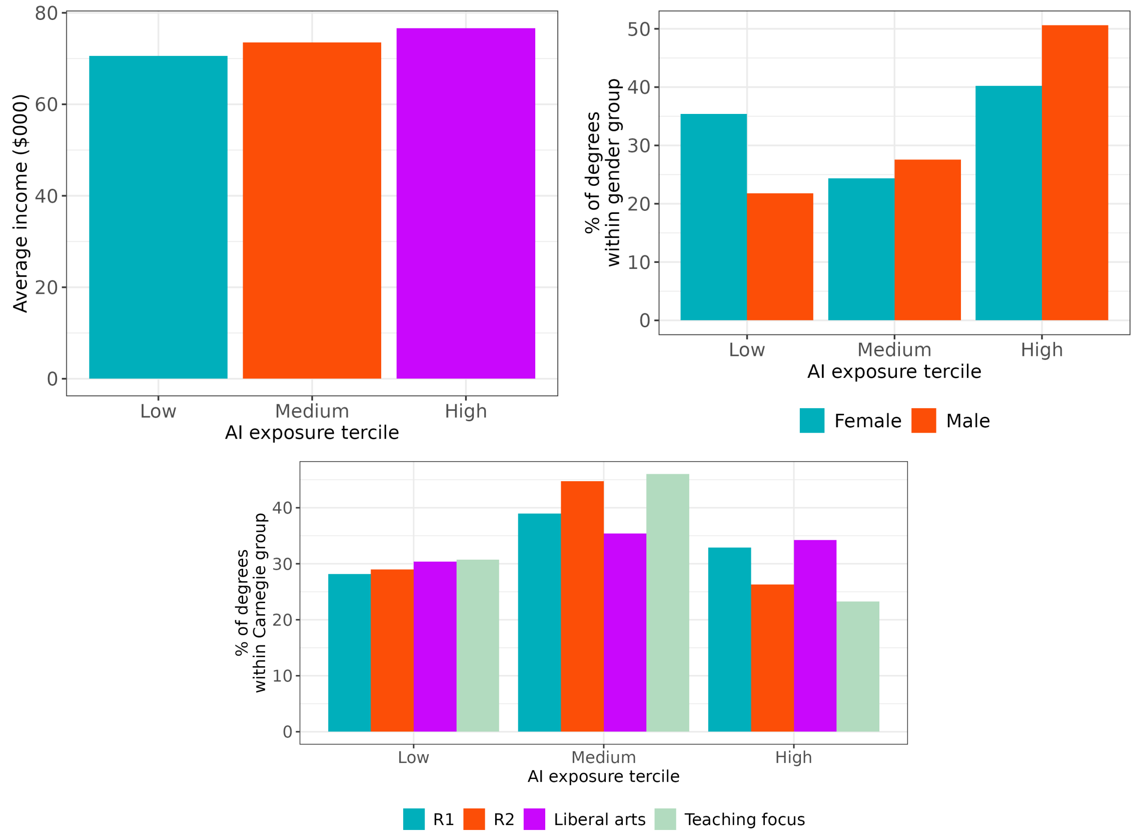
Notes: The figure plots density plots of field-technology exposure scores, split by technology class. "Robot" and "AI" measure patent-based overlap using the Webb (2019) exposure methodology; the "OpenAI" score measures overlap based on LLM capabilities measured by Eloundou et al. (2023). Courses are partitioned into 52 fields (e.g., Math, Business, History). Each field of study is represented as a dot, with size corresponding to the enrollment share in the field.

Figure 5. Time trend in technology-field of study exposure



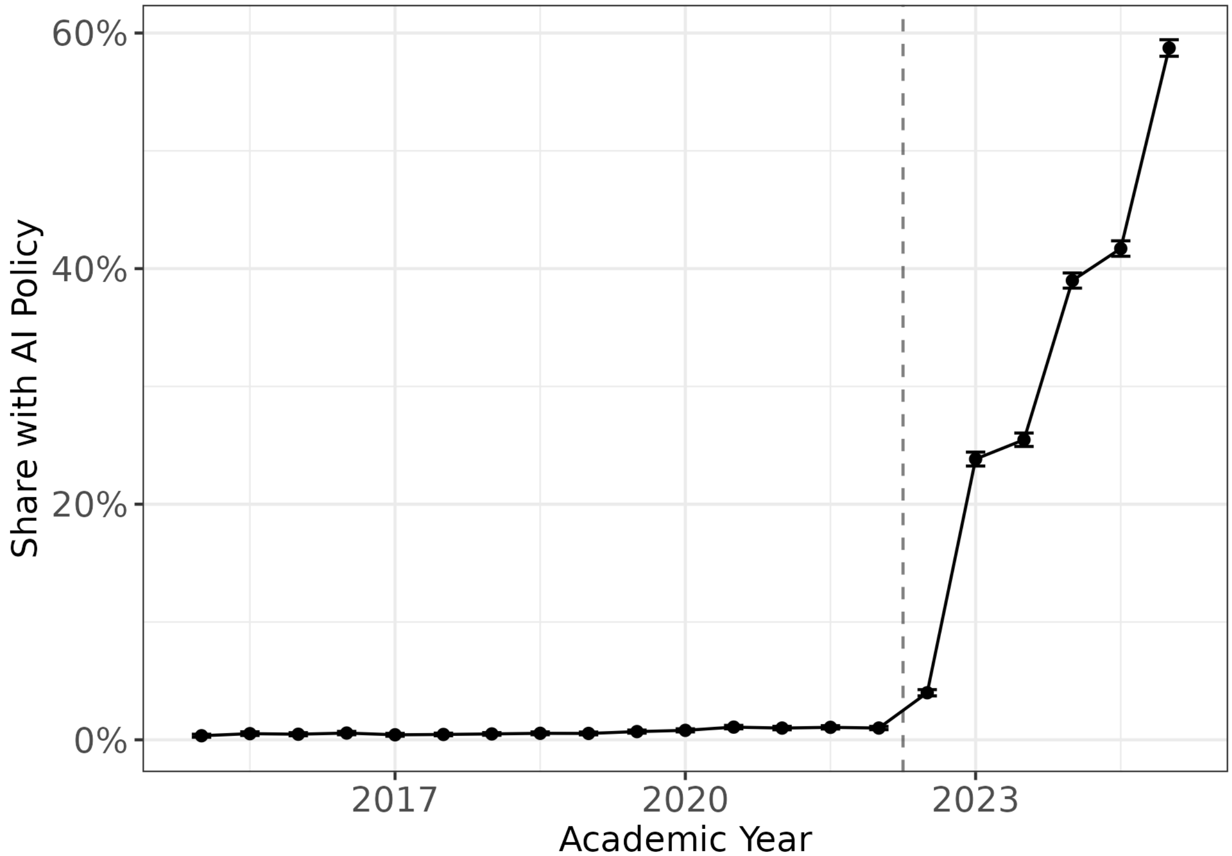
Notes: The figure plots density plots of field-technology exposure scores, split by technology class. "Robot" and "AI" measure patent-based overlap using the Webb (2019) exposure methodology; the "OpenAI" score measures overlap based on LLM capabilities measured by Eloundou et al. (2023). Courses are partitioned into 52 fields (e.g., Math, Business, History). Observations are at the institution-field-year level, estimates from a time series regression, with 2022-23 the reference year, controlling for institution-by-field fixed effects and weighting by the number of verb-object pairs in each cell. Scores are normalized based on the occupation distribution of the same technology.

Figure 6. Distributional implications of technology exposure



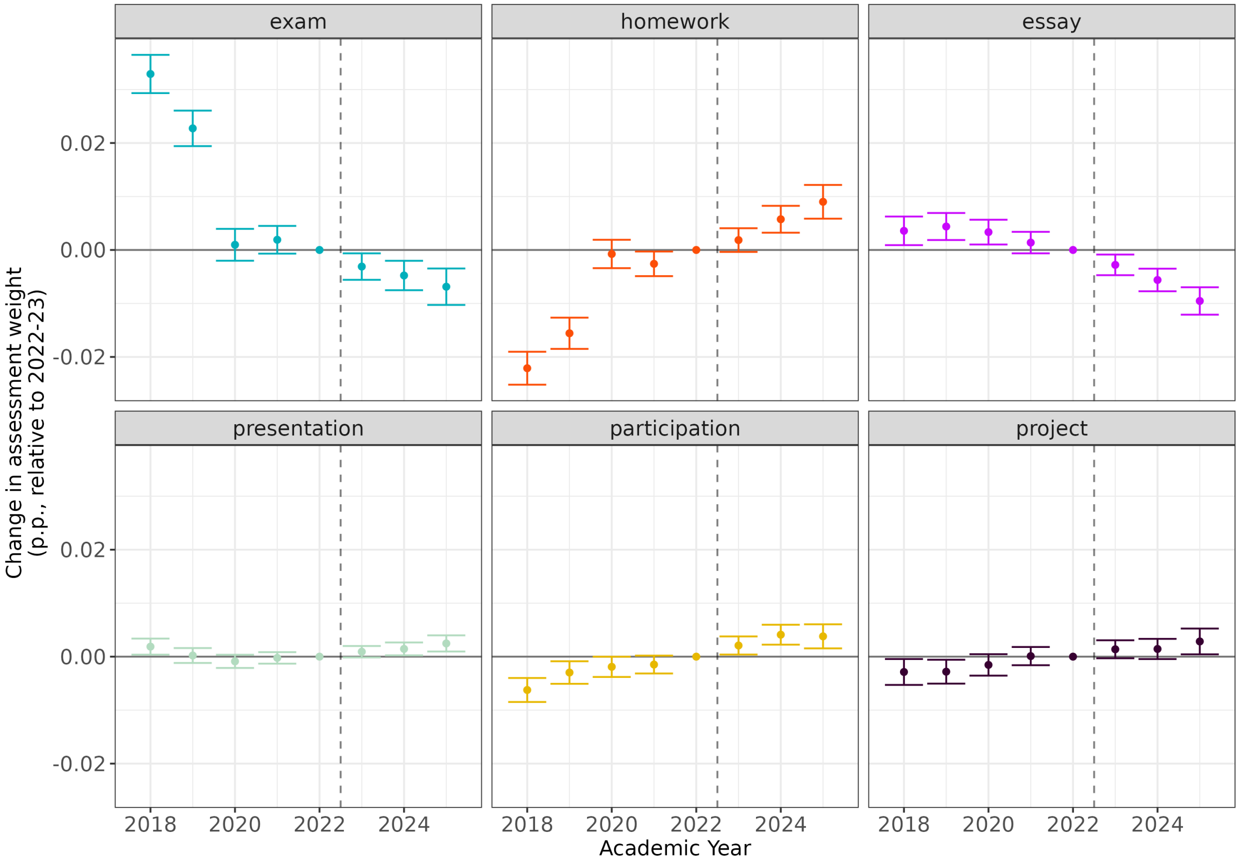
Notes: The three figures summarize differential technology exposure across three different groups, according to terciles of the "OpenAI" LLM-based AI exposure distributions. The first panel plots differential exposure for men and women based on the within-gender share of major completions at four-year institutions. The second panel plots average income ($000) by tercile of AI exposure. The third panel plots the within-Carnegie group share of major completions.

Figure 7. Trend in AI policies



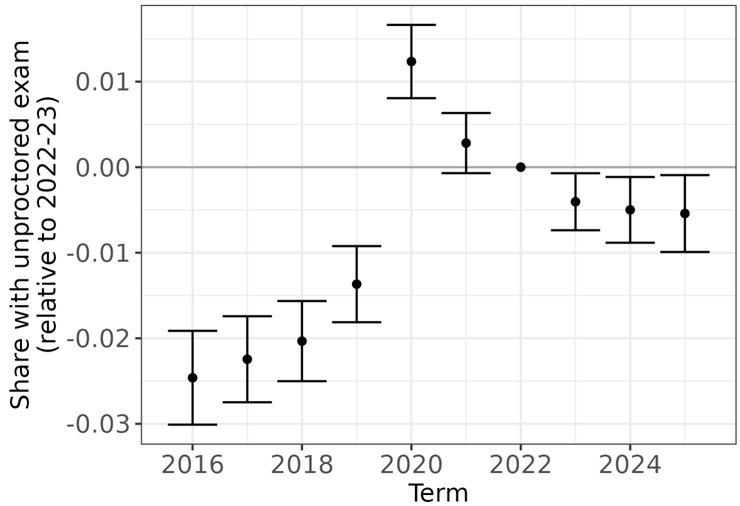
Notes: Estimates from interrupted time series with binary dependent variable (1 if the syllabus has an explicit AI policy), based on course-level observations from a sample of syllabi from 25 US colleges and universities. Fall 2022 is the omitted term. Time variable in academic years (thus, the 2022 point refers to the Fall 2022 semester, the 2022.5 point refers to Spring 2023, etc.). The dashed line indicates the public release of ChatGPT. AI policies pre-Fall 2022 are typically policies regarding the use of online translation tools in foreign language courses or related to AI-assisted exam administration during the pandemic. Standard errors clustered at the course level.

Figure 8. Changes to grade component weights around release of ChatGPT



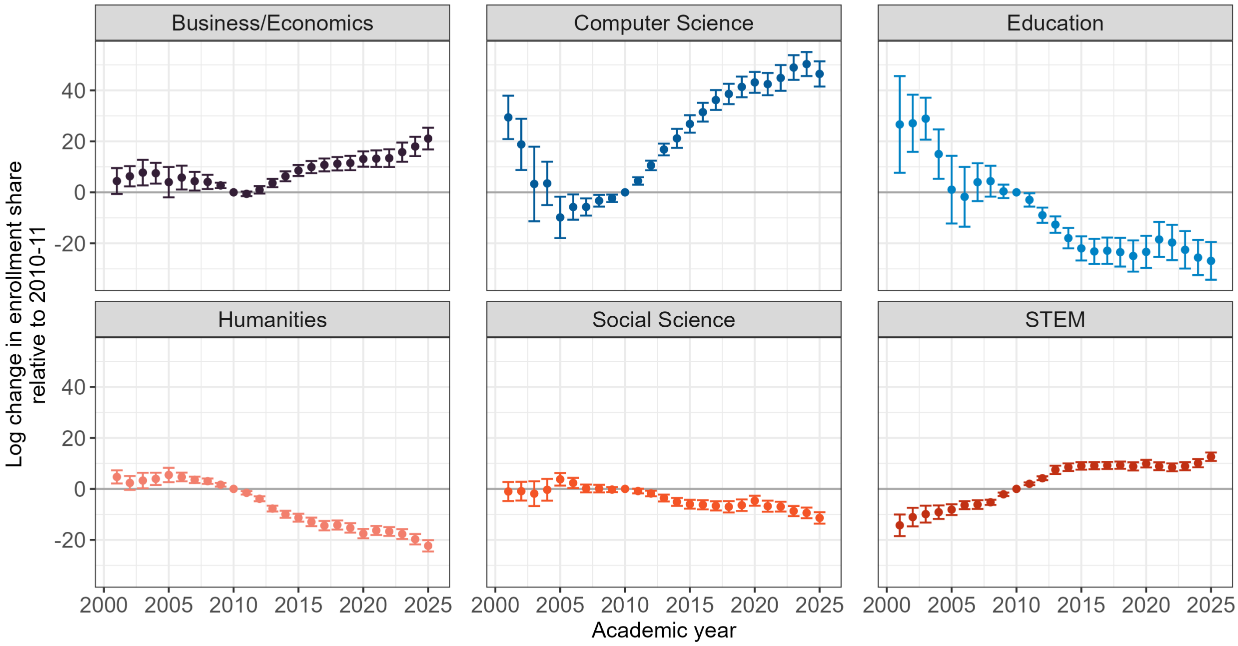
Notes: Grading weights from syllabi from a sample of 25 US colleges and universities. Estimates come from interrupted time series regressions of each grade component weight category around the release of ChatGPT in 2022-23. Observations are at the institution-course-semester level. Regressions control for course-semester fixed effects. Standard errors clustered at the course level.

Figure 9. Changes in exam modality



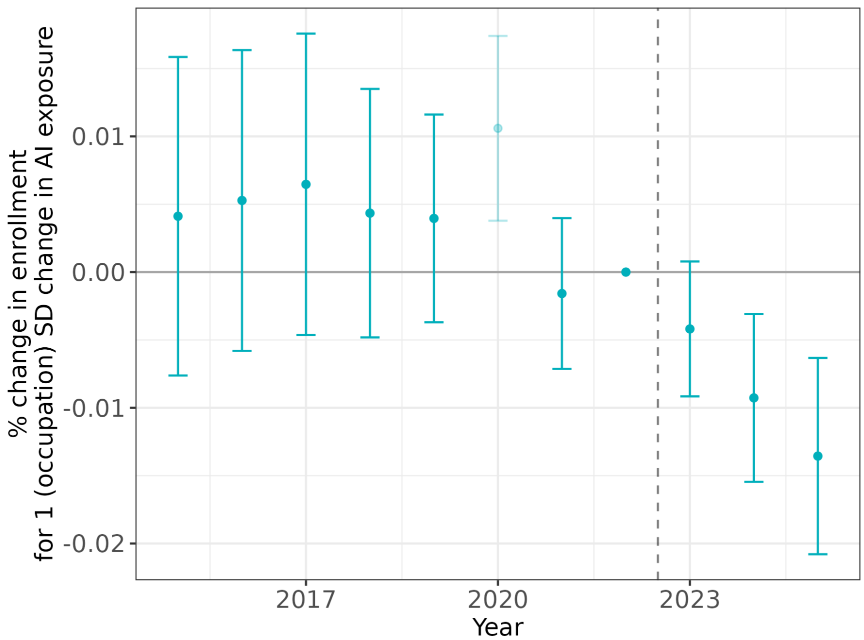
Notes: Estimates from an event study regression with a binary dependent variable equal to one if the syllabus describes an unproctored, take-home, open-book, or online exam. Observations are at the course-semester-year level, drawn from a sample of syllabi at 25 US colleges and universities. The reference year is 2022-23. The regression controls for institution-by-course and institution-by-semester fixed effects. The dependent variable is identified from syllabus text using a regular-expression classifier that detects phrases associated with unproctored exam administration (e.g., "take-home exam," "open-book," submission via an LMS). The dashed line indicates the public release of ChatGPT. Standard errors clustered at the course level.

Figure 10. Enrollment changes 2000-2025



Notes: The figure plots the change in the enrollment share of each broad field category at four-year institutions, relative to the 2010-11 academic year. Enrollment share is defined as the fraction of total undergraduate enrollment in a given institution-semester accounted for by each field category. Estimates come from separate regressions of log enrollment share on academic-year indicators, with institution-by-semester fixed effects, run independently for each field category. The reference year is 2010-11. Observations are weighted by total institutional enrollment in 2024-25. Standard errors are clustered at the institution level.

Figure 11. Enrollment trend following ChatGPT release



Notes: Estimates come from event study regressions of log course enrollment on year indicators interacted with field-level AI exposure (LLM measure), normalized across fields to mean zero and standard deviation one. The coefficient in each year is interpreted as the percentage change in log enrollment associated with a one standard deviation increase in a field's AI exposure, relative to the 2022-23 reference year. Observations are at the institution-field-semester-year level and restricted to undergraduate courses. Regressions control for institution-by-field, institution-by-year, institution-by-semester, and field-by-semester fixed effects. The 2020 estimates are shown at reduced opacity due to pandemic-related disruption. The dashed line indicates the public release of ChatGPT. Standard errors clustered at the institution level.

Table 1. Common verb-object pairs for select fields

Field (1)	Verb (2)	Object(s) (3)	Verb % (4)
Arts	develop	understanding, technique, project, awareness	0.074
Arts	examine	history, role, development, art	0.054
Arts	create	work, image, design, artwork	0.046
Arts	learn	technique, skill, fundamental, basic	0.031
Computer Science	develop	application, understanding, system, solution	0.052
Computer Science	solve	problem, strategy, student, topic	0.045
Computer Science	learn	technique, concept, skill, principle	0.044
Computer Science	examine	issue, opportunity, concept, principle	0.041
Economics	examine	design, model, role, modeling	0.091
Economics	analyze	datum, interaction, issue, policy	0.062
Economics	discuss	policy, relationship, system, method	0.051
Economics	understand	behavior, topic, application, role	0.050
Education	examine	issue, theory, role, research	0.081
Education	develop	understanding, knowledge, plan, strategy	0.071
Education	teach	student, science, study, strategy	0.058
Education	learn	disability, strategy, outcome, environment	0.049
English	write	assignment, paper, essay, skill	0.102
English	examine	work, production, discourse, history	0.071
English	read	text, work, novel, narrative	0.046
English	develop	idea, point, understanding, strategy	0.042
History	examine	history, development, experience, war	0.215
History	study	history, study, topic, movement	0.064
History	shape	history, life, identity, society	0.050
History	understand	character, history, diversity, violence	0.034
Math	solve	problem, equation, system, situation	0.215
Math	apply	technique, method, model, problem	0.039
Math	graph	equation, function, line, calculator	0.032
Math	learn	outcome, reading, student, skill	0.027

Notes: The table lists the three most common verbs in each field of study and, for each verb, the (up to) four most common associated objects. Column 4 reports verb (2)'s share (percentage points) of all verb-object pairs for the field.

A Appendix

A.1 Summary of data extraction procedure

This appendix describes the procedure used to convert raw syllabus documents into the structured, course-level records analyzed in Section 6. The procedure proceeds in three steps: (i) converting heterogeneous source documents into clean plain text, (ii) extracting structured fields (grade components and their weights, AI policies, and supporting metadata) from that text using a large language model, and (iii) mapping the extracted fields into the analysis categories defined in Section 2.1.2.

Institutions publish their syllabus archives in a mix of formats, primarily PDF, Microsoft Word, and HTML. I first convert each document to plain UTF-8 text. I apply a light cleaning pass to remove recurring page headers and footers (e.g., "Page 2 of 5"), horizontal rules, and redundant whitespace. I then extract structured fields from each converted syllabus using the OpenAI API.[24] I process each syllabus is independently against a fixed system and user prompt. The system prompt instructs the model to return a single JSON object conforming to a specified schema and to emit null or empty values for any field that is absent or ambiguous; the user prompt enumerates the target fields and directs the model to copy text verbatim rather than summarize. Consistent with the verb-object pair extraction described earlier, the prompt is deliberately minimal and context-free, with all downstream filtering and harmonization handled in post-processing rather than through elaborate prompt engineering.

The schema captures the course number and title, term and year, instructor name(s) and email(s), the course description and objectives, the grade breakdown, the AI policy, and binary indicators for the presence of a disability-accommodation statement and a land acknowledgment.[25] Two fields are central to the analysis. I record the grade breakdown as a list of components consisting of the model's verbatim label for the component (e.g., "Midterm Exam," "Reading responses"), the associated weight, and the raw source line from which it was drawn. The LLM indicates the existence of an AI policy through an indicator for whether the syllabus contains an AI policy, a stance classification (Prohibited, Limited, Permitted, or Unclear), and the literal text excerpt from the syllabus. To capture the location of the course description without requiring the model to reproduce long passages, the model returns short anchor strings — the first and last few words of the description — that locate the corresponding span in the source text.

[24] I perform the extraction using model `gpt-5-nano-2025-08-07`.

[25] I extracted these fields to conduct a placebo check of whether instructors were adding boilerplate non-AI policies to the syllabi around the release of ChatGPT. I omit this comparison from the paper; I detect no spike in the prevalence of these policies around the release of ChatGPT in Fall 2022.

I map the extracted grade component labels to the fixed set of categories described in Section 2.1.2 using string matching on the component text. This division of labor — free-form extraction by the model, deterministic classification in post-processing — keeps the mapping transparent and reproducible and makes it straightforward to revise the category definitions without re-querying the model. As noted in Section 2.1.2, I remove subtotal lines (components whose label contains the string "total") to avoid double-counting. I collapse the AI policy is collapsed to a binary indicator analysis.

Because the model returns the verbatim source text alongside each processed field, every extracted grade weight and policy classification can be checked directly against the underlying syllabus. The human validation exercise described in Section 2.1.2 confirms that the full grading scheme is recovered correctly in 85% of audited cases.

A.1.1 LLM Prompt

Extract the following information from the course syllabus below. Return a JSON object with the following fields:

- course_number (string)
- course_title (string)
- semester: "term": string, "year": integer
- instructor: list of "name": string, "email": string (just one is fine)
- description_objectives: list of "first_five": string, "last_five": string, "start_idx": integer—null, "end_idx": integer—null
  - Extract the actual course description/objectives text (do NOT summarize).
  - Exclude section headers like "Course Description" or "Objectives".
  - "first_five" = the first 5 whitespace-separated words from the start of that section, preserving original casing/punctuation. If the section has <5 words, return all available words.
  - "last_five" = the final 5 words from the end of that same section (or all available if <5).
  - For first_five and last_five: return ONLY these two literal substrings without ellipses or extra tokens.
  - "start_idx" = character index for start of provided description text

- "end_idx" = exclusive end index for provided description text
- If either index is uncertain, set it null

- grading_breakdown: list of "category": string, "weight_pct": number—null, "raw": string
- ai_policy: "present": boolean, "stance": "Allowed"—"Limited"—"Prohibited"—"Unclear", "excerpt": string
- land_acknowledgement (boolean): true if the syllabus includes a land acknowledgment, false otherwise
- disability_policy (boolean): true if the syllabus mentions a disability accommodation or accessibility policy, false otherwise

If numeric grading weights are given only in points, set "weight_pct" to numeric grading weight. If a field is missing or unclear, output null, false, or [] as appropriate. Return ONLY the JSON, no extra prose.

Syllabus text: {syllabus_text}

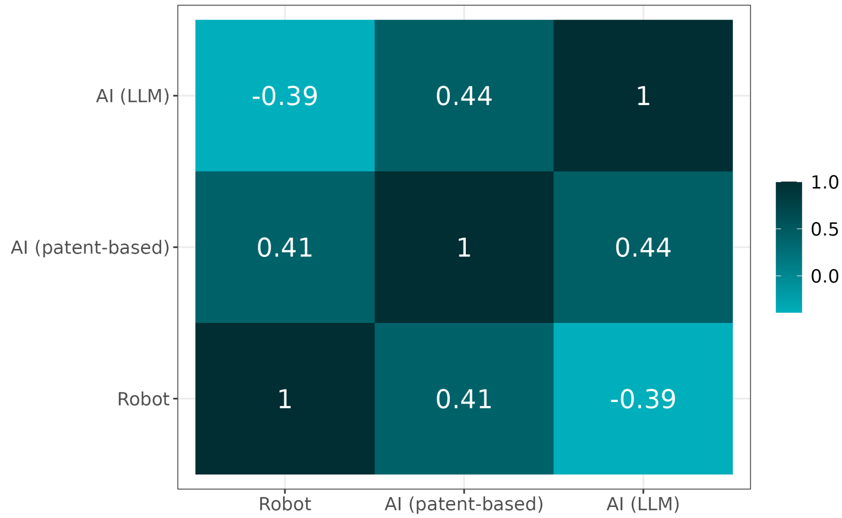
A.2 Additional figures and tables

Table A-1. Syllabus sample data coverage

Institution	Earliest Year	Latest Year	Syllabi Count	2022-23 Coverage % Courses	2022-23 Coverage % Enrollment
Angelo State University	2021-2022	2025-2026	4892	52	67
Arizona State University-Downtown Phoenix	2012-2013	2025-2026	10745	52	71
Arizona State University-Polytechnic	2011-2012	2025-2026	10788	51	66
Arizona State University-Tempe	2011-2012	2025-2026	79909	47	70
Arizona State University-West	2012-2013	2025-2026	8916	32	51
Brandeis University	2016-2017	2025-2026	9376	22	25
Clemson University	2013-2014	2024-2025	42478	31	—
Florida Gulf Coast University	2006-2007	2025-2026	18188	81	86
Sam Houston State University	2017-2018	2025-2026	16152	94	97
Stephen F Austin State University	2018-2019	2025-2026	14473	83	90
Texas A & M University-College Station	2013-2014	2025-2026	52715	65	—
Texas A & M University-Corpus Christi	2021-2022	2025-2026	2215	—	—
Texas A & M University-Kingsville	2016-2017	2025-2026	12101	88	89
Texas A&M University-San Antonio	2021-2022	2025-2026	4028	85	92
Texas A&M University-Texarkana	2011-2012	2025-2026	6993	64	68
Texas Southern University	2019-2020	2025-2026	8276	41	63
The University of Alabama	2005-2006	2023-2024	53819	46	65
The University of Montana	2010-2011	2024-2025	29803	18	21
The University of Texas at Austin	2010-2011	2025-2026	123576	53	—
The University of Texas at Dallas	2010-2011	2025-2026	27612	75	—
The University of Texas at El Paso	2017-2018	2025-2026	16502	66	82
The University of Texas at Tyler	2020-2021	2025-2026	2633	—	—
University of Florida	2018-2019	2025-2026	28492	21	—
University of Georgia	2014-2015	2025-2026	46953	29	47
West Texas A & M University	2020-2021	2025-2026	2564	—	—

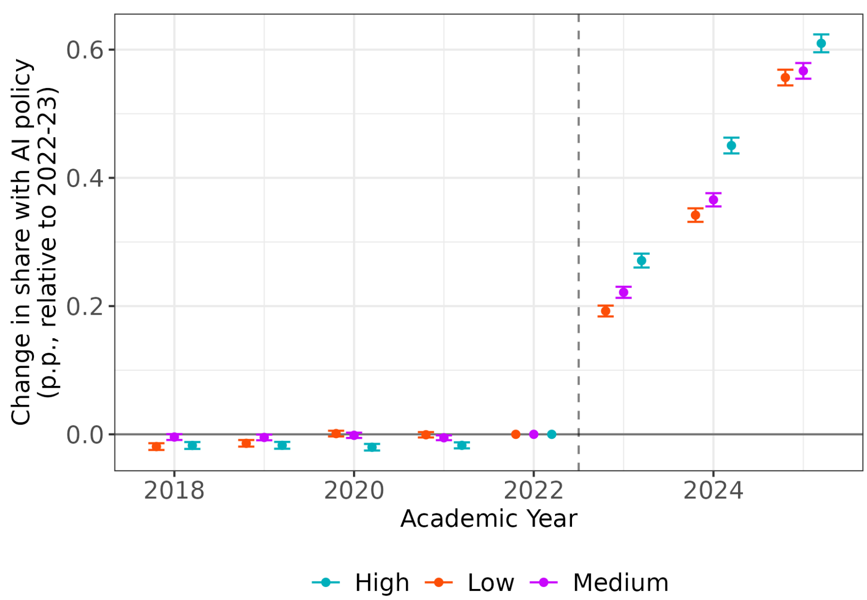
Notes: Syllabi collected by scraping online public syllabus repositories. 2022-2023 coverage shares calculated based on coverage of course offerings and enrollment data from Light (2026). Missing values indicate that section-level course offerings or section-level enrollment are not available in the course offerings data.

Figure A-1. Correlation of verb-object frequency scores by technology class



Notes: The figure reports the pairwise correlations of same field field-technology exposure scores for each technology type pair.

Figure A-2. AI policy split by AI exposure



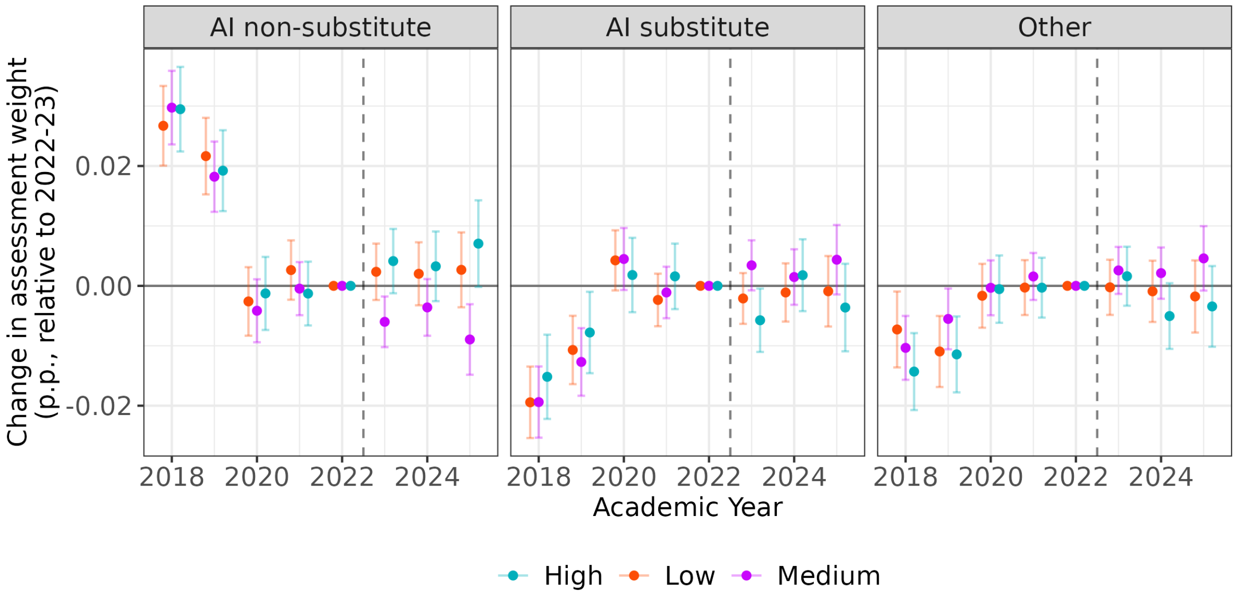
Notes: The figure plots the share of syllabi with an explicit AI policy over time, separately by tercile of AI exposure (based on the LLM exposure measure). Means are computed at the course-semester level using all syllabi with non-missing AI policy information. The dashed line indicates the public release of ChatGPT. Points are jittered slightly along the x-axis to reduce overlap across tercile groups.

Table A-2. Baseline grade component weights

type	exam	homework	essay	participation	project	presentation	lab	other
Overall	0.341	0.160	0.130	0.119	0.091	0.038	0.016	0.105
Business	0.469	0.167	0.076	0.086	0.092	0.032	0.005	0.074
Education	0.129	0.218	0.135	0.147	0.097	0.055	0.017	0.203
Humanities	0.209	0.128	0.195	0.176	0.095	0.054	0.006	0.136
STEM - CS	0.426	0.222	0.038	0.047	0.157	0.016	0.039	0.055
STEM - exclude CS	0.448	0.183	0.080	0.078	0.080	0.027	0.028	0.076
Skilled Trade	0.361	0.149	0.102	0.109	0.105	0.029	0.018	0.127
Social Science	0.299	0.133	0.185	0.138	0.081	0.041	0.008	0.115
Non-R1	0.361	0.165	0.116	0.117	0.085	0.033	0.018	0.104
R1	0.320	0.156	0.144	0.120	0.098	0.042	0.014	0.106

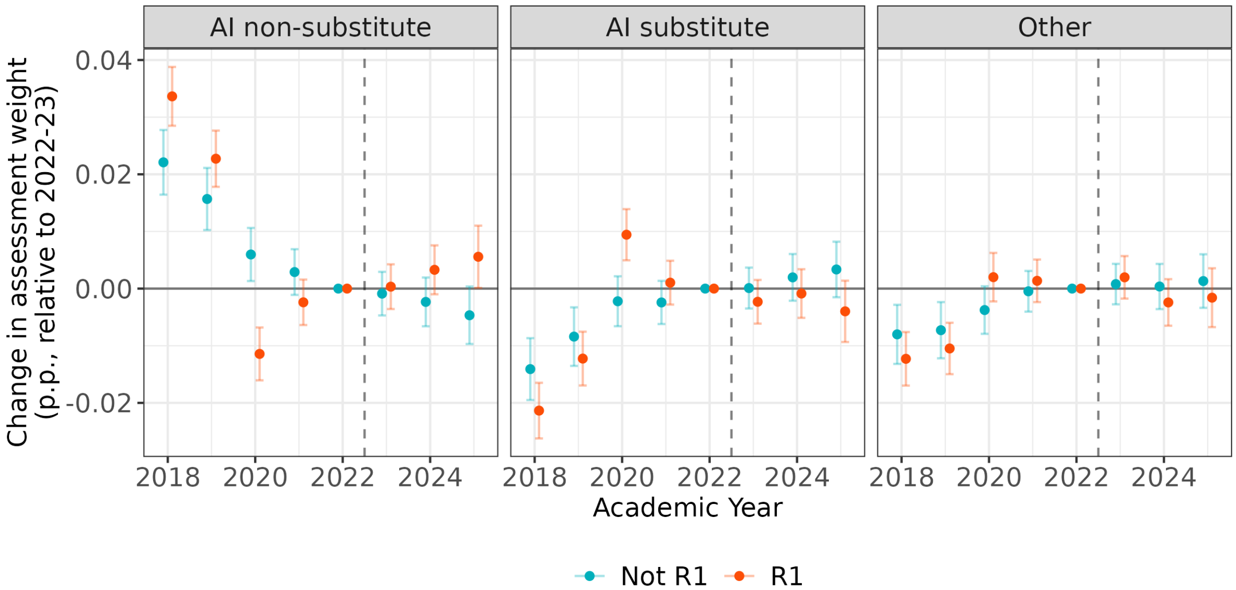
Notes: The table summarizes grade component based on course syllabi in 2022-23. Syllabi receive equal weight in the calculation. In the primary analysis, I classify essays and homework as “AI substitutes.” I classify exams, participation, presentations, and labs as “AI non-substitute.”

Figure A-3. Assessment event study split by AI exposure (LLM)



Notes: Each panel plots event study coefficients from a regression of the indicated aggregated grade component weight category on year indicators interacted with an indicator for R1 institution status, estimated separately for R1 and non-R1 institutions. "AI substitute" assessments are essays and homework; "AI non-substitute" assessments are exams, participation, presentations, and labs. The reference year is 2022-23. Observations are at the institution-course-semester-year level. Regressions control for institution-by-course-by-semester fixed effects. R1 institutions are defined as those with a Carnegie classification as doctoral universities with very high research activity as of 2018. The dashed line indicates the public release of ChatGPT. Standard errors clustered at the course level.

Figure A-4. Assessment event study split by institution type



Notes: Each panel plots event study coefficients from a regressions of the indicated aggregated grade component weight category on year indicators, estimated separately for R1 and non-R1 institutions in the syllabus sample. "AI substitute" assessments are essays and homework; "AI non-substitute" assessments are exams, participation, presentations, and labs. The reference year is 2022-23; the reference tercile group is Low. Observations are at the institution-course-semester-year level. Regressions control for institution-by-course-by-semester fixed effects. Year points are jittered slightly along the x-axis by tercile group to reduce overlap. The dashed line corresponds to the public release of ChatGPT. Standard errors clustered at the course level.