Are College Degrees Still Protected from Automation? Measuring Curricular Exposure to Artificial Intelligence

Jacob Light*

May 31, 2025

Preliminary - please do not cite or circulate
[Click here for latest version]

Abstract

Using a new dataset of course descriptions from 900 U.S. colleges and universities, I adapt the patent-to-task overlap method in Webb (2019) to measure how strongly each college major is exposed to three technology classes: Robots, Software, and Artificial Intelligence (AI). Exposure to Robots and Software is modest, but AI exposure is pervasive: the median major ranks around the 75th percentile of the occupation-level AI-exposure distribution. STEM and Business/Economics majors are more exposed to AI than Social Science or Humanities fields. Within a given field, teaching-focused institutions exhibit greater AI exposure than R1 research universities. Finally, I outline an event study analysis that leverages the November 2022 release of ChatGPT to estimate how student enrollment and course supply adjust to a sudden, highly salient shift in AI capabilities.

JEL Classification: J24, I2, O3.

*Hoover Institution. 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 history, 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 return to "upskilling" through college coursework may erode faster than in previous eras. Moreover, curricular change in higher education is notoriously sticky (Light 2024); universities may struggle to realign programs before technologies have already leapt ahead.

This paper provides the first systematic measurement of how exposed college majors are to recent technological advances, with a focus on AI. Building on the task-based approach in Webb (2019), I combine more than twenty-five years of course descriptions, enrollment counts, and offerings for 900 U.S. colleges and universities (Light 2024) with the universe of U.S. patents classified as Robotics, Software, or AI. For each field of study, I extract verb-object pairs (e.g., "analyze data," "write essays") from course descriptions, construct a task-frequency distribution, and compute its overlap with the task profiles that characterize each technology class. The resulting *exposure score* is intentionally agnostic: it captures potential for either substitution or complementarity, leaving the sign of the welfare effect for subsequent analysis.

The paper documents three headline facts. First, exposure of college courses to AI is significantly greater than their exposure to previous waves of technological change, such as Robotics or Software patents. While only one field (Skilled Trades) has exposure at or above the 75th percentile of occupation exposure to Robots, the median field has exposure to AI above the 75th percentile of the occupation-level AI exposure distribution. STEM fields — especially Statistics and Data Science — are most exposed to AI, followed by Economics and Business. Second, AI exposure of college courses has risen steadily since 2010, even within fields of study. The growing exposure is driven largely by the incorporation of data analysis tasks into new and existing courses, both in and outside of the traditionally quantitative fields. Third, within a given field, exposure varies systematically by institution type. Teaching-oriented and regional public universities offer curricula that peak at more introductory levels and, therefore, overlap with current AI capabilities more than the comparatively advanced sequences at R1 research universities. However, the single most exposed field-institution cell is Statistics/Data Science at R1 institutions.

In preparing these results, I conduct a series of validation exercises to demonstrate that the task-based measure of human capital development in universities produces sensible and meaningful patterns. The tasks extracted from course descriptions recover intuitive relationships between fields of study and the skills emphasized in those fields. For example, tasks such as "analyze data" and "solve equations" appear most often in quantitative fields like Math and Statistics, whereas tasks such as "write essays" or "interpret texts" are prominent in English and Humanities courses. Moreover, the tasks emphasized in each field correlate with the tasks performed in the occupations graduates typically enter. When I repeat the main exposure analysis using this occupational linkage, rather than the direct field-to-technology overlap, the exposure rankings across fields are very similar. This finding suggests that the task-based approach captures a real channel through which students' coursework aligns with the skills they deploy in the labor market.

In ongoing exploratory work, I intend to explore how new technology influences student enrollment and course offerings. I outline an event study design around the sudden public release of ChatGPT in November 2022. In theory, the release of ChatGPT and similar LLM models creates potentially opposing pushes on student enrollment. In the short-term, the availability of these tools complements (or substitutes for) student effort in courses whose tasks overlap with AI capabilities, thereby reducing the cognitive cost of taking these courses. However, the same overlap may impact students' expected return on investment in human capital that overlaps with new technology's capabilities, potentially reducing the long-term benefits of taking these courses. Comparing courses with higher pre-release AI exposure to otherwise similar courses within the same department, I will estimate how course enrollment and supply adjust when the technology frontier shifts. Evidence from this analysis will provide new insight into factors guiding student and institution preferences.

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, the event study framework should provide new evidence on the elasticity of both student demand and course supply to fast-moving technological shocks.

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]

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.

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 outlines the proposed event study analysis to measure enrollment responses to the release of ChatGPT. Section 7 concludes.

2 Data

The primary data sources for this project are patent text from the US Patent Office, covering all patents filed between 2000-2024, and comprehensive course offerings data from 900 US

[1] Although I follow Webb's use of patents here, the same methodology could also be applied to other sources of technology data, like GitHub commits or academic papers, to expand the measure's coverage as AI evolves.

colleges and universities.

2.1 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 2024). I collected these data by scraping online course catalogs and course schedules, capturing 50 million course sections offered since 1996 across 900 institutions.

The dataset forms an unbalanced panel that includes the complete set of course offerings for each institution-year. For each course section, I observe 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]

2.2 Patent data

I obtain patent titles and abstracts for all patents filed with the US Patent and Trademark Office between 2000-2024. 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).[3] An example of a patent title is "Analyzing flight data using predictive models" (Desell et al. (2019)). The final patent dataset includes approximately 135,000 patents covering Robotics, Software, or 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 Software, I adapt the classification of Bessen and Hunt (2007) and include patents mentioning "software" or both "computer" and

[2]See Light (2024) for a detailed discussion of coverage and representativeness.

[3]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.

“programming,” excluding those also mentioning “circuit,” “chip,” “semiconductor,” “bus,” “antigen,” or “chromatography.”

For AI, I expand Webb’s definition 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, 80.225 Software patents, and 26,136 AI patents. Approximately half of the patent titles contain at least one verb-object pair, resulting in around 12,000, 16,000, and 43,000 verb-object pairs for the three technology classes, respectively.

2.3 Supplemental data sources

For validation exercises, 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. These occupation details are gathered through firm and worker surveys to maintain up-to-date information. For my analysis, I use task summaries linked to occupations in the O*NET data.[4] I use O*NET ratings of task importance and relevance to weight verb-object pairs in constructing exposure measures.

To link fields of study to occupations, I use field-occupation mappings estimated from the American Community Survey (ACS). Specifically, I use ACS surveys from 2009-2019 and restrict to employed workers ages 25-65 who hold a four-year degree. This mapping serves as an external validity check for my task-based exposure measures, allowing me to compare the skills emphasized in college courses to those performed in jobs linked to these courses.

3 Methodology

The core methodology for this project involves three main steps: first, extracting “tasks” (verb-object pairs) from course descriptions and patent titles; second, constructing distributions of these tasks for each field of study and technology class; and third, calculating an overlap statistic that reflects the exposure of each field (or field-institution cell) to each technology. This exposure score captures the degree to which the core tasks students perform in their courses are also performed by new technologies.

[4] 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.”

3.1 Defining and extracting tasks

Text data encode rich information, but the core challenge is reducing this high-dimensional, interconnected content into a tractable form that captures what is economically meaningful. In this project, I must extract from course descriptions — documents that include administrative, pedagogical, and contextual details — only the parts relevant for measuring a course's technology overlap. A similar process occurs, implicitly, in most economic analyses: for example, researchers often control for an individual's race, gender, or age in labor studies, but not for their quarter of birth or year of marriage unless relevant. With text data, the separation between relevant and irrelevant information is harder to anticipate, so applications typically rely on data-driven approaches to isolate the pieces of text most pertinent for analysis.

I define a *task* as a verb-object pair. For example, consider the description for Stanford's COMM 247D (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[5], and lemmatize the verb and object to a common form (e.g., "analyzing" and "analyze" 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 first remove "boilerplate" sentences (such as prerequisites and enrollment restrictions). Next, I calculate chi-squared statistics for each verb-object pair, comparing their occurrence frequency in course descriptions versus patents. I drop pairs that are among the 100 pairs with the highest chi-squared statistic,

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

suggesting that these pairs are broadly related to course descriptions as a document type rather than the specific tasks students perform or human capital they develop.[6]

To further ensure that the verb-object pairs are closely related to course tasks, I project the cleaned task dictionary onto the union of task pairs that appear in either the patent data or the O*NET task inventory (excluding previously dropped pairs). This step ensures that exposure scores only reflect tasks that are relevant to technological change or to job tasks more broadly.

The final dictionary includes 27,612 unique verb-object pairs. Table 3 summarizes how many pairs are dropped at each step. In Section 4, I summarize validation exercises, including manual review of top pairs by field, to confirm that the resulting task dictionary captures the main activities taught in those courses.

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 weighted average of task shares:

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

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. In later analyses, I extend this

[6]For example, "introduce student" or "receive credit." Examples of dropped pairs are listed in Appendix Table 2.

exposure calculation to field-institution cells to account for differences in course sequencing and teaching emphasis across different types of colleges and universities.

4 Validation

I conduct a series of validation exercises to confirm that the task-based exposure measures capture meaningful differences across fields of study and align with intuitive and observable patterns in the data.

4.1 Distinctive tasks by field of study

To 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 across fields. For example, English courses emphasize tasks such as "reading texts" and "writing essays," Education courses emphasize "teaching lessons" and "developing curriculum," and Arts courses emphasize "creating images" and "performing music." At the same time, some tasks — particularly those involving data analysis — appear across multiple fields, reflecting the growing role of data skills across a range of disciplines.

Columns 5-7 of Table 1 show the technology exposure of these most important tasks, calculated as a weighted average of task frequency in course descriptions and task relevance to each technology class. For readability, I multiply the scores by 100. The table reveals that data analysis tasks are highly exposed to AI across the board, with particularly high exposure for fields such as Economics and Statistics/Data Science where this form of analysis is emphasized. More broadly, exposure to Robots is generally the lowest across all fields, followed by Software. Where tasks overlap with AI, the extent of that overlap is often much greater than for other technology classes.

4.2 Linking field tasks to occupation tasks

As a second validation exercise, I assess whether the tasks extracted from course descriptions for a given field also appear in the occupations into which graduates from that field typically flow. This provides an external check that the task-based measure of field exposure captures tasks that matter in the labor market.

To do so, I use field-to-occupation mappings developed from the ACS for prime-age workers (ages 22-65).[7] I then construct task vectors for each occupation and each field of study, based respectively on the importance of verb-object pairs in O*NET and their share in course descriptions. For this exercise, I restrict the task space to only those pairs that appear in any O*NET task descriptions. I measure the similarity between each field-occupation pair by calculating the cosine distance in task space. To ensure that I focus on meaningful matches, I drop any field-occupation pair with either an empirical share below 0.01% in the ACS and no task overlap.[8] Finally, I compute the rank-rank correlation between the empirical field-to-occupation shares and the task-based rankings.

The resulting rank-rank correlation is 0.31, indicating a meaningful positive relationship between the two measures. Given the sparsity of O*NET task lists and the fact that some fields map to multiple occupations imperfectly, this relatively large and positive correlation validates the relevance of my task-based measure. It also provides new evidence of how the tasks students learn in college align with the jobs they enter after graduation.

5 Results

5.1 Overall overlap between courses and technology

Figure 2 plots the distributions of technology exposure scores for each of the three technology classes (Robots, Software, and AI) using course descriptions from the 2023-2024 academic year. For each of 52 fields of study (e.g., Math, English, History), I calculate the average exposure score by pooling tasks from all undergraduate-level courses offered across institutions.

The results indicate that most fields of study have low exposure to Robots, with some modest variation. The field most exposed to Robots is Skilled Trades, which includes courses in vocational subjects like Construction and Auto Manufacturing, mainly offered at community colleges. Computer Science, Medicine, Engineering, and Agriculture also have relatively higher exposure to Robots, though the absolute levels are low. Similarly, exposure to Software is generally low across fields, with Library Science showing the highest overlap, followed by Statistics/Data Science, Computer Science, Engineering, Business, and Economics. However, overall exposure to both Robots and Software is minimal across most fields.

In contrast, the AI exposure distribution is wider and shows substantially higher values for almost every field. The most exposed field is Statistics/Data Science, with exposure scores nearly five times the average across fields. Other highly exposed fields include Linguistics,

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

[8] Dropping these occupation-field pairs is generally conservative for this validation exercise, as it excludes pairs where the occupation-to-field shares and task overlap are strongly correlated.

Computer Science, Economics, and Chemistry. Generally, STEM and Business fields are more exposed to AI, followed by Social Sciences, while Humanities fields (such as History, Ethnic/Cultural Studies, American Studies, and Religion) show the lowest exposure.

For reference, in each panel I plot the 75th and 95th percentiles of the occupational technology exposure distribution, using the Webb (2019) scoring.[9] For Robots and Software, most fields of study score well below the 75th percentile of occupational exposure. In contrast, the median field scores just above the 75th percentile for AI, and the most exposed field (Statistics/Data Science) exceeds the 95th percentile of occupational exposure to AI. This pattern is consistent with prior work showing that Robots and Software overlap more with low-skill jobs, whereas AI overlaps more with higher-skill activities.

There are two primary takeaways from this figure. First, the same fields that were most exposed to previous waves of technology through this task-based measure are the fields most exposed to AI. This may be due to overlapping task intensity across the three technologies (for example, 'analyze data" is a task common across the three technologies, potentially due to misclassification, but is substantially more important for AI than other technology categories). However, second, AI exposure is markedly higher across almost all fields compared to previous waves of technological change, especially when compared to occupation-level exposure to these technologies.

5.2 Time trend in technology exposure

Figure 3 shows how exposure to each technology has evolved over the past two decades. I estimate these trends by regressing field-technology exposure scores on year dummies, controlling for institution-by-field fixed effects. This approach isolates changes within a field of study rather than differences driven by compositional changes across fields.

The figure reveals a steady increase in AI exposure across nearly all fields since approximately 2012-13. On average, courses across fields have increased in their AI exposure by 0.00003 points (equivalent to 0.5 standard deviations in the occupation-AI exposure distribution) over the last decade. This trend is most pronounced in fields that have incorporated more data analysis tasks into their curricula. In contrast, exposure to Robots and (to a lesser extent) Software has remained largely flat over time, suggesting that the growing AI

[9] Projecting the course verb-object pair dictionary onto the union of the patent and occupation pair dictionaries produces a distribution of exposure scores for courses that is similar in magnitude to the occupation exposure scores. However, some of the pairs that appear in course descriptions (particularly those related to administrative details) do not overlap with on-the-job activities. Projecting only onto the patent and occupation verb-object pairs thus introduces an upward bias in the exposure scores. As a result, what matters for interpretation is not the absolute level of these scores, but rather the relative scores across fields within a given technology class, and how field-level exposure compares to different percentiles of the occupation exposure distribution.

exposure reflects the specific expansion of AI-overlapping tasks into a broader set of courses.

Although there is no clear exogenous shock in 2012-13 to explain the inflection point, this timing coincides with broader shifts in student demand toward Computer Science and STEM majors and away from Education and Humanities (Light 2024). Universities may have responded by incorporating more data-intensive content into their courses to align with these demand shifts, indirectly driving up AI exposure across multiple disciplines.

5.3 Heterogeneity by institution type

Institutions differ in the types of courses they offer and how they structure those courses, which may influence patterns of technology exposure. From a policy perspective, understanding this heterogeneity can identify which types of institutions and students are most exposed to growing AI capabilities. Additionally, the results in this section provide new insight into how instruction in the same field of study differs at different types of institutions.

I calculate exposure scores separately for field-technology-institution type cells. I classify institutions into four categories based on Carnegie classifications and degree offerings: R1 universities (very high research activity), R2 universities (high research activity), teaching-focused four-year universities, and two-year community colleges. For each cell, I pool course descriptions to calculate task distributions and measure their overlap with the patent-derived task distributions for each technology.

Figure 4 plots the distributions of exposure scores across these institution categories. Exposure to Robots and Software is uniformly low across all types of institutions, with R1 universities generally having the lowest exposure to these technologies. In contrast, AI exposure has greater variation across institution types and fields. On average, teaching-focused universities have higher AI exposure than both research universities and community colleges. This higher exposure is not due to differences in how the same courses are taught at these institutions — for example, the AI exposure of Principles of Economics is similar regardless of institution type. Instead, the difference arises from the types of courses offered. Teaching-focused universities often offer introductory-level courses and lack the advanced electives more common at research universities. For example, among teaching-focused universities in my sample, 75% offer pre-calculus courses and only 65% offer advanced courses such as Real Analysis. At R1 universities, these shares are 60% and 94%, respectively. This suggests that as AI's capabilities move up the ladder of cognitive tasks, they will match the level of courses at teaching-focused universities before reaching the more advanced material taught at research universities.

Finally, the single most exposed field-institution cell is Statistics/Data Science at R1 universities, followed by the same field at R2 universities. The creation of new Data Science programs and the broader expansion of applied statistics courses appear to drive this high level of AI exposure.

6 Enrollment responses (ongoing)

In ongoing work, I aim to study how students and institutions have adjusted course enrollment and offerings since the release of OpenAI's ChatGPT in November 2022. ChatGPT is a large language model (LLM) trained to generate text and answer questions in a conversational style. As the first LLM widely accessible to the public, ChatGPT represented a step change in the practical availability of AI-powered tools. Two months after its release, ChatGPT passed 100 million users.[10] Its surprise release represents an inflection point in the availability of generative AI across educational and professional settings.

The release of ChatGPT may have multiple, potentially offsetting, effects on student enrollment preferences. On one hand, by automating routine coursework tasks, ChatGPT may reduce the cognitive cost of courses that emphasize those tasks, thereby increasing student demand for AI-exposed courses. On the other hand, the same overlap that makes these courses easier may also reduce the long-term payoff to the human capital developed in them if students expect that the tasks they learn will be automated in the labor market as well.

To test these competing predictions, I plan to implement an event study analysis that exploits variation in AI exposure at the course level. My empirical strategy leverages within-institution and within-field-of-study comparisons before and after the ChatGPT release. Building on the finding that task profiles are highly comparable across institutions (Section 5.1), I assign courses to common categories (e.g., Calculus I, Labor Economics, Organic Chemistry) and pool course descriptions within each category to estimate the task distributions associated with those courses.

I will then estimate an event study regression at the institution-course-semester level, with exposure measured by the degree of overlap between each course's task distribution and AI task capabilities. This design allows me to identify changes in course enrollment and supply decisions that coincide with the sudden shift in perceived AI capabilities following ChatGPT's release.

[10] For comparison, Facebook passed 100 million users four and a half years after its public release (Porter 2023).

7 Conclusion

This paper adapts the approach of Webb (2019) to provide new measures of the exposure of college courses to emerging technologies, including Artificial Intelligence. Whereas Webb maps patents to O*NET tasks to identify which jobs are technologically exposed, I map the same patents to the tasks embedded in college courses and majors. This shift moves the focus from labor demand adjustment to the upstream investment (and potential misallocation) of human capital.

The descriptive results confirm predictions in the literature that AI is distinct from earlier technological waves in the degree to which it overlaps with high-skill tasks. Exposure of college courses to Robots and Software is generally low, but the median field of study has AI exposure comparable to the 75th percentile of the occupation-level AI exposure distribution. STEM and Business/Economics courses are the most exposed to AI, as are courses at teaching-focused universities compared to research institutions.

Next steps will focus on the event study analysis outlined in Section 6. A central challenge is the identification assumption of parallel enrollment trends for highly versus less exposed courses within the same field — an assumption that may be complicated by Covid-19 related enrollment shocks in the immediate pre-period. If this assumption does not hold in the data, alternative strategies will be needed. Additionally, I plan to develop a simple conceptual model to formalize the competing forces of technology-induced demand and substitution in course-taking decisions. Finally, to move closer to a welfare calculation, I will explore back-of-the-envelope estimates of how the concentration of AI exposure in high-earning fields could compress college major-earnings gaps.

More broadly, an open question is whether (and in what ways) task-technology overlap indicates a risk of substitutability versus an opportunity for productivity gains. I remain intentionally agnostic in this draft about whether exposure implies a threat of automation or a complementarity that could enhance human capabilities. This question is first-order for interpreting the policy implications of these findings and represents an important direction for continued research.

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.

Bessen, James and Robert M Hunt (2007). “An empirical look at software patents”. In: *Journal of Economics & Management Strategy* 16.1, pp. 157–189.

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.

Desell, Travis, James Higgins, and Sophine Clachar (2019). *Analyzing flight data using predictive models*. US Patent 10,248,742.

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 (2024). “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.

Porter, John (2023). “ChatGPT continues to be one of the fastest-growing services ever”. In: *Verge*. URL: https://www.theverge.com/2023/11/6/23948386/chatgpt-active-user-count-openai-developer-conference.

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 144
Lecture: 01
Units: 5
Class#: 31846

Winter 2021

Family and Society

Department of Economics

Lecture, Discussion
1/11/21 - 3/19/21

To be Scheduled
12:00 AM - 12:00 AM
Remote

Instructor: Persson, P.

Enrollment Status

Open Seats: 0
Enrolled: 81 Waitlist: No waitlist
Capacity: 80 Waitlist Max: No waitlist

Course Description

The family into which a child is born plays a powerful role in determining lifetime opportunities. This course will apply tools from economics and related social sciences to study how the functioning of families is shaped by laws, social insurance, social norms, and technology. Topics will include intergenerational transmission of wealth and health, the importance of the early family environment, partnership formation, cohabitation and marriage, teen pregnancy and contraception, assisted reproduction, Tiger Moms and Helicopter Parenting, and the employment effects of parenthood. In the context of these topics, the course will cover social science empirical methods, including regression analysis, causal inference, and quasi-experimental methods. Throughout the course, we will think critically about the role of the government and how the design of public policy targeting families affect our ability to solve some of the most important social and economic problems of our time. Prerequisites: Econ 50

Grading basis

Letter or Credit/NoCredit Exce

Class level

Undergraduate

Instructional mode

Remote: Asynchronous

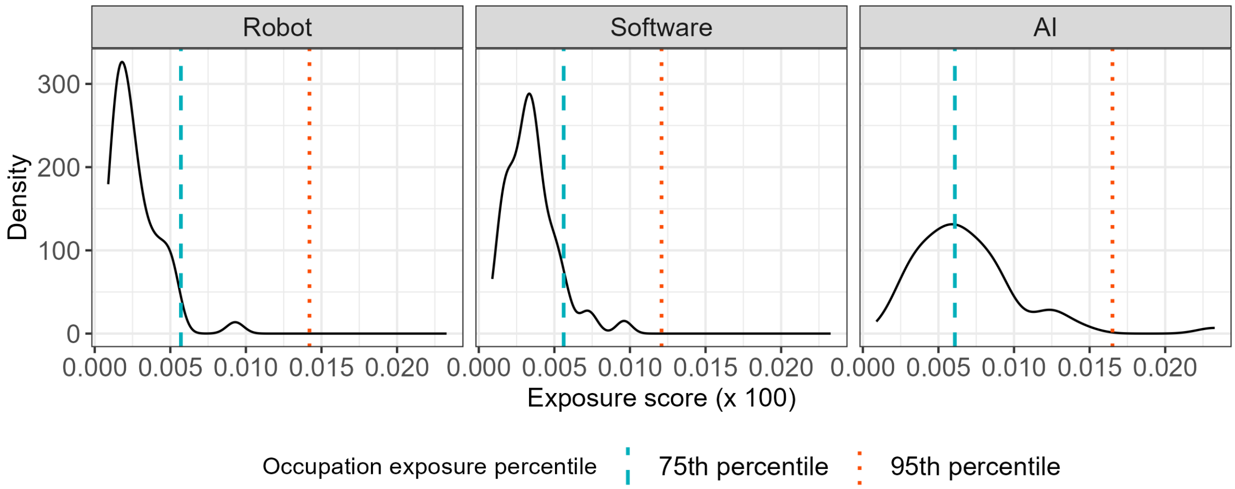
Final exam

Meets Requirement(s)

WAYS - Applied Quantitative Reasoning (AQR)
WAYS - Social Inquiry (SI)
SYMBO-BS Subplan: Computational Social Science
ECON-BS Core Program Requirements
ECON101 ECON101 Prerequisite

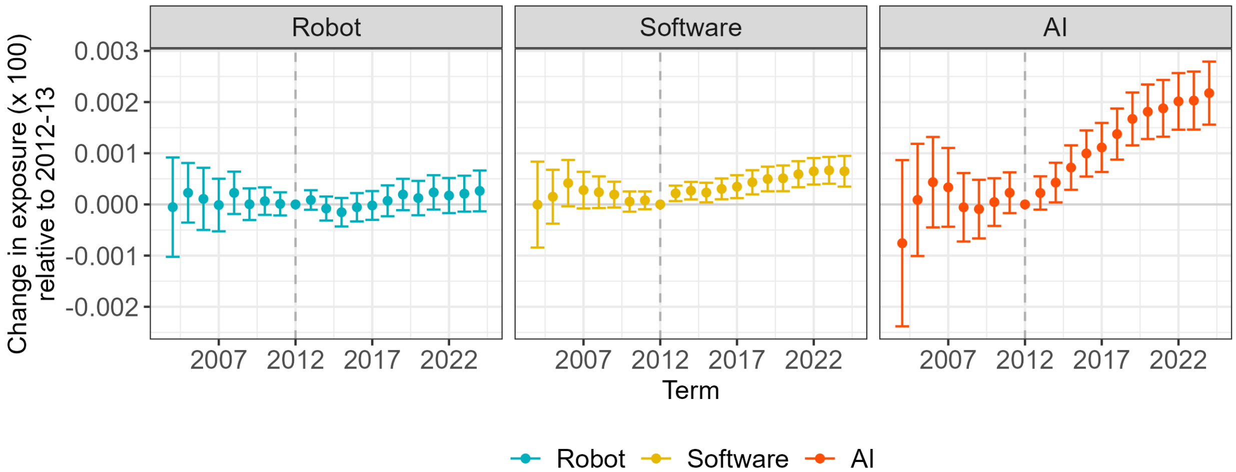
Source: Stanford University.

Figure 2. Technology exposure distribution across fields



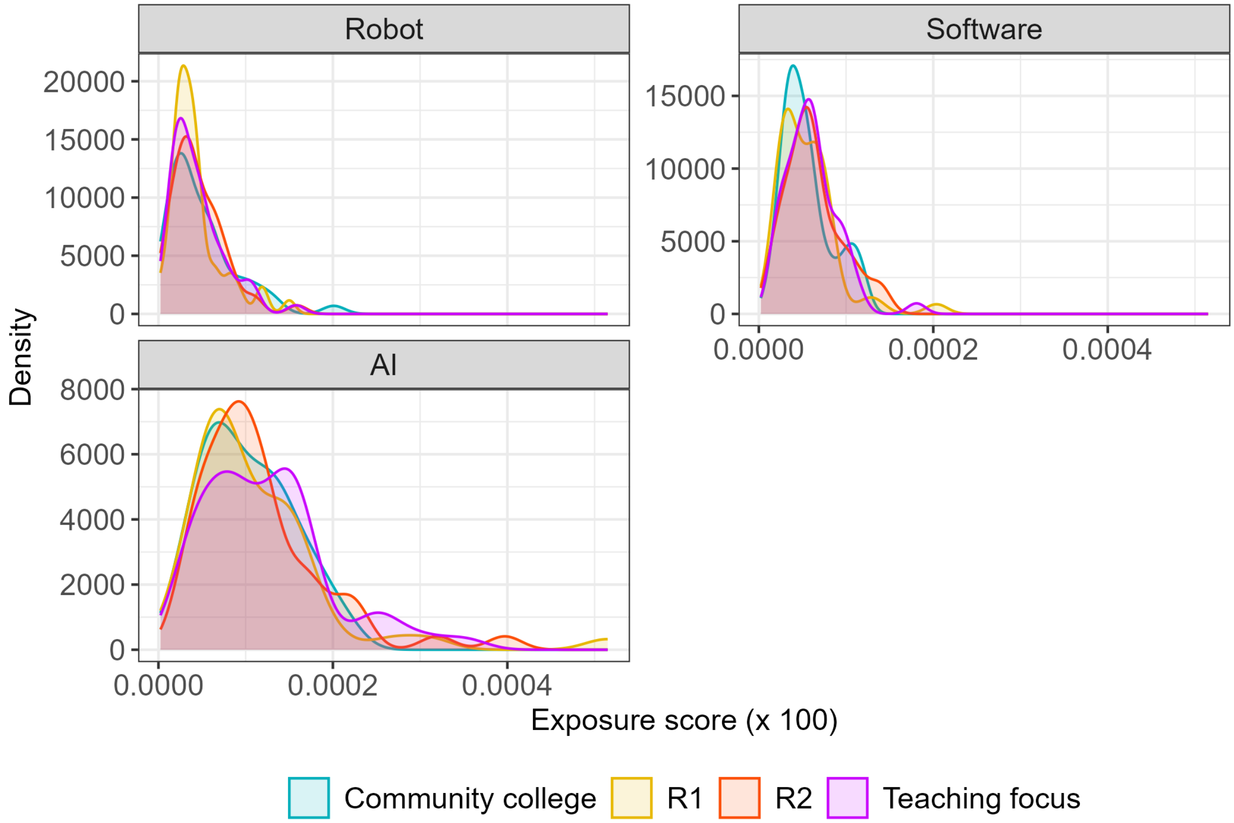
Notes: The figure plots density plots of field-technology exposure scores, split by technology class. Courses are partitioned into 52 fields (e.g., Math, Business, History). Vertical lines represent the 75th and 95th percentiles of the corresponding occupation-technology exposure distributions.

Figure 3. Technology exposure trend across fields



Notes: The figure plots the time trend in average technology exposure by field. Estimates come from separate regressions of technology exposure score on a vector of time dummies, controlling for institution-by-field fixed effects. Observations are at the institution-field-year level, and standard errors are clustered at the institution-field level. Estimates are relative to the average field-technology exposure in 2012-13.

Figure 4. Technology exposure distribution across fields, by institution category



Notes: The figure plots overlapping density plots of field-technolgoy exposure scores, with institutions broken out by degree level and research intensity (Carnegie classification). Exposure scores are calculated using course descriptions for the 2023-24 academic year. Within each school category-field cell, course descriptions are pooled together and verb-object pairs are extracted to calculate separate verb-object pair distributions.

Table 1. Common verb-object pairs for select fields

Field (1)	Verb (2)	Object (3)	% of tasks (4)	AI (5)	Robot (6)	Software (7)
Arts	develop	technique, concept, approach, appreciation	0.023	0.000	0.000	0.000
Arts	create	work, project, artwork, image	0.023	0.004	0.002	0.006
Arts	teach	music, lesson, skill, technique	0.018	0.000	0.000	0.001
Chemistry	perform	experiment, calculation, analysis, research	0.067	0.015	0.004	0.012
Chemistry	develop	technique, concept, perspective, principle	0.038	0.000	0.000	0.000
Chemistry	qualitative	analysis	0.035	0.000	0.000	0.000
Computer Science	develop	application, solution, system, program	0.019	0.003	0.000	0.020
Computer Science	process	spreadsheet, database, communication, graphic	0.019	0.000	0.000	0.003
Computer Science	create	application, page, visualization, plan	0.010	0.007	0.008	0.014
Economics	analyze	datum, problem, behavior, issue	0.045	0.124	0.028	0.050
Economics	develop	model, tool, theory, technique	0.033	0.019	0.004	0.003
Economics	apply	tool, method, analysis, model	0.032	0.038	0.009	0.006
Education	teach	lesson, skill, art, child	0.025	0.000	0.000	0.001
Education	learn	strategy, environment, disability, style	0.022	0.013	0.000	0.000
Education	develop	proposal, relationship, approach, philosophy	0.012	0.000	0.000	0.000
English	write	composition, requirement, analysis, fiction	0.063	0.000	0.000	0.000
English	read	text, literature, page, skill	0.034	0.000	0.000	0.001
English	develop	process, expertise, paper, question	0.023	0.000	0.000	0.001
Ethnic/Cultural Studies	examine	relationship, aspect, method, content	0.023	0.000	0.000	0.000
Ethnic/Cultural Studies	analyze	impact, representation, issue, study	0.021	0.003	0.000	0.002
Ethnic/Cultural Studies	develop	tool, appreciation, proposal, sense	0.016	0.004	0.000	0.001
Math	solve	equation, system, variety, model	0.194	0.000	0.002	0.005
Math	analyze	datum, graph, problem, relationship	0.028	0.091	0.021	0.038
Math	graph	system, analysis	0.025	0.007	0.000	0.000
Stats/Data Science	analyze	datum, output, set, result	0.078	0.214	0.052	0.088
Stats/Data Science	present	datum, result, concept, model	0.040	0.015	0.000	0.025
Stats/Data Science	test	analysis, power, model, application	0.017	0.001	0.000	0.002

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 the share of all verb-object pairs represented by these selected pairs. Columns 5-7 multiply these shares by the technology weights for each technology class.

Additional figures and tables

Table 2. Excluded verb-object pairs highly distinctive of course descriptions

Verb	Object	Verb	Object
introduce	student	satisfy	requirement
provide	student	take	course
develop	skill	solve	problem
provide	opportunity	conduct	research
receive	credit	include	study
prepare	student	provide	experience
fulfill	requirement	include	analysis
provide	introduction	enable	student
provide	overview	offer	semester
develop	understanding	meet	requirement

Notes: The table shows the 20 verb-object pairs (of the 100 excluded from the pair dictionary) that are most distinctive of course catalogs. The pairs are ranked by the chi-square statistic comparing their frequency shares in course descriptions to their frequency shares in patents.

Table 3. Summary of task counts in course description data

	Catalog [1]		Occupation [2]		Patent [3]		Occupation + patent [4]	
	Unique	Total	Unique	Total	Unique	Total	Unique	Total
1 - full corpus	248,088	3,759,797	12,092	25,697	122,027	461,387	129,691	487,084
2 - remove uncommon pairs	112,735	3,576,337	12,092	25,697	30,180	348,955	31,819	367,463
3 - remove catalog-specific pairs	112,645	3,078,562	12,058	25,123	30,161	348,493	31,788	366,428
4 - project onto patent/occupation pairs	10,017	719,479	12,058	25,123	30,161	348,493	31,788	366,428
5 - remove stopwords	8,110	452,331	10,942	21,564	26,150	254,375	27,612	269,258

This table reports counts of unique and total verb-object pairs at each stage of data processing. Counts are shown for course descriptions (1), occupation tasks from O*NET (2), patents (3), and the union of the O*NET and patent corpora. Row (1) covers the full corpora. Row (2) excludes extremely uncommon pairs (fewer than three appearances) in the catalog and patent data (not applied to O*NET, which is sparser and has higher signal). Row (3) removes 100 pairs that are highly distinctive of course descriptions. Row (4) projects the catalog pairs onto the dictionary of pairs found in either O*NET or patents. Row (5) removes stopwords.