Openness to AI in Higher Education*

Fernando Saltiel‡

This version: July 23, 2026
First version: March 26, 2026

Please click here for the latest version.

Abstract

This paper examines how AI is being integrated in higher education. I construct a dataset of 185,000 course syllabi from 50 public universities between 2024–2026 and use open-weight large language models to classify how instructors are incorporating AI into their teaching and evaluation. The LLM-based approach is reliable, showing strong agreement with human raters. By Spring 2026, sixty-five percent of courses included AI policies, yet only six percent had integrated it into their teaching. Integration is concentrated at selective, research-intensive universities, while less selective institutions rely more on explicit policies governing AI use. Instructors are the primary force shaping AI integration in the classroom. Assistant professors and research-active instructors are especially likely to integrate AI into their courses, particularly in STEM fields and at selective universities.

*I thank Maggie Jones, Fabian Lange, Matias Cortes, Brennan McLachlan and Jérôme Larivière for detailed comments and suggestions, as well as seminar participants at the BSE Summer Forum and McGill University for helpful feedback. Magnolia Miller and Aditi Ramanand provided excellent research assistance. All errors are my own.

†A companion website is available at: https://ai-higher-ed.netlify.app.

‡McGill University and IZA. Email: **fernando.saltiel@mcgill.ca**

1 Introduction

The emergence of generative artificial intelligence tools has prompted rapid adoption across the economy, with recent evidence documenting sizable effects on worker productivity, task allocation, and organizational design (Brynjolfsson et al., 2025; Bick et al., 2025). The diffusion of AI creates distinct challenges for universities, as they must decide how to integrate these tools into the classroom even as students are already using them in their coursework (Contractor and Reyes, 2025; Handa et al., 2025). These decisions matter not only because universities regulate AI use in the classroom, but also because they shape the skills that students develop in college. Understanding how universities and instructors are integrating AI in the classroom is therefore increasingly important.

In this paper, I analyze how artificial intelligence is being integrated into higher education. In particular, I examine how instructors are incorporating AI into their classes by constructing a dataset of course syllabi across public universities through Spring 2026. Syllabi are a natural context for studying this question, as they capture the content of college classes and the approaches that professors take in the classroom (Biasi and Ma, 2022). As such, they reveal how instructors are actually integrating AI into their classes, ranging from what they teach to how they evaluate student work.

To this end, I collect publicly available course syllabi, recovering 185,000 syllabi across 50 public universities in seven states between 2024 and 2026, covering a period with widespread AI use in higher education.[1] To understand which professors are most likely to integrate AI into the classroom, I construct a complementary dataset capturing their research productivity and academic rank from multiple data sources. I also recover institutional and major-level characteristics that allow me to identify which universities and majors are driving this integration.

The share of syllabi mentioning AI has risen sharply, from 42 percent in Fall 2024 to 54 percent by Spring 2026, up from fewer than three percent before ChatGPT. However, a syllabus that prohibits AI use and one that builds assignments around it would both be classified as mentioning the technology, despite taking opposite approaches in the classroom. I instead aim to recover how instructors are integrating AI into their teaching and how they use it to evaluate their students.

I thus construct four measures that capture how AI is being integrated into the classroom. On the teaching margin, I measure whether AI is part of what the course teaches, through its class sessions, topics, or course materials. On the evaluation margin, I measure whether at least one graded task requires students to engage with AI. I also consider whether the syllabus includes an explicit policy around the use of AI, and whether it affirmatively permits students to use it. Lastly, I validate these measures with human raters, finding high inter-rater agreement across all four questions (Dell, 2025; Haaland et al., 2025).

I classify whether these measures are present in each syllabus using large language models (LLMs), which are well-suited for this task as they can read unstructured syllabus

[1] For UT-Austin and UT-Dallas, I use syllabi dating back to 2019, which allows me to analyze how AI integration in the classroom has evolved since the release of ChatGPT.

text in context (Ash and Hansen, 2023). I implement the classification approach using local open-weight models, which enhance the reproducibility of the results (Korinek, 2023).[2] Specifically, I classify each syllabus using three independent open-weight models, taking the modal answer across them. To ensure that this approach can reliably classify the integration measures at scale, I compare the LLM classifications against the human raters. The LLM-based approach passes the Alternative Annotator Test, as it matches the consensus of human raters as often as the individual raters do (Calderon et al., 2025).

I use the LLM-based classification to assess how AI has been integrated into higher education. I find that sixty-five percent of courses had explicit AI policies by Spring 2026. However, just 28 percent of syllabi actually allowed students to use AI. This suggests that instructors are reticent to allow this technology in the classroom. In fact, just six percent of courses incorporate AI into their teaching. Professors are even less likely to incorporate AI into student evaluation, as only three percent of classes include assignments requiring students to engage with this technology. Altogether, a small share of students are being required to actively engage with AI in their college classes.

At the same time, AI integration could vary across universities, given differences in resources, research productivity and student populations. Eight percent of courses at selective institutions have already integrated AI into the classroom, compared to just five percent at less selective universities. Moreover, institutions with larger graduate sectors, research expenditures and higher faculty salaries are more likely to have integrated AI into their classes. This evidence is thus consistent with prior work showing that AI adoption is concentrated at larger and more research-intensive firms (McElheran et al., 2024; Babina et al., 2024). Within these universities, graduate courses integrate AI at higher rates than undergraduate classes, consistent with the fact that they cover content closer to the research frontier (Biasi and Ma, 2026).

Since majors differ significantly in their exposure to AI (Eloundou et al., 2024), I also analyze how adoption varies across fields. Business majors are most likely to have incorporated AI into their classes, including 16 percent of courses in Marketing. By contrast, courses in the Social Sciences have the lowest rates of classroom adoption: just two percent of Economics courses have integrated AI. Importantly, the majors most exposed to AI show the highest rates of classroom adoption. Instructors in these fields may be adapting their courses to a labor market increasingly shaped by AI (Bick et al., 2025).

Given the large differences in AI integration across universities and majors, I estimate a variance decomposition that separates the contributions of instructors, courses, and institutions. Professors are the main drivers of whether AI is adopted in the classroom, as they account for 25 to 30 percent of the variation across the teaching and evaluation margins. Moreover, course-level differences account for a comparable share of the variation in these outcomes. On the other hand, universities and majors play a limited role in shaping whether AI is integrated into the classroom.

Since professors are the main drivers of whether AI is adopted in the classroom, I exam-

[2]For each syllabus, the LLMs classify each of the four questions independently, to ensure that the response to one question does not influence its response to the others.

ine which instructors lead this adoption. Assistant professors are far more likely than full professors to have incorporated AI in their classes, conditional on the courses they teach and the broader institutional environment. Research productivity also predicts adoption outcomes, as instructors with publication records are more likely to integrate AI in their classes. Adoption is even stronger among instructors whose own research engages with machine learning. These effects are largely for integration on the teaching margin, and are robust to additional controls and fixed effects. Taken together, the leading edge of classroom AI integration is concentrated among young, research-active faculty members.

These differences hold in STEM and non-STEM fields alike. The career-stage differences, however, are concentrated at selective universities, where assistant professors are especially likely to integrate AI into the classroom. At the same time, cumulative measures of research productivity, including publication and citation counts, do not significantly predict AI integration beyond the main set of characteristics. This differs from Biasi and Ma (2022), who show that more productive researchers are more likely to bring frontier knowledge into the classroom. AI integration follows a different logic, since it requires instructors to experiment with a new technology and to redesign their courses around it. That helps explain why young, research-active faculty members are more likely to integrate AI into the classroom.

This paper makes several contributions to the literature. First, I contribute to a rapidly growing literature documenting the adoption of artificial intelligence across firms, workers, and occupations (Acemoglu et al., 2022; McElheran et al., 2024; Bonney et al., 2024; Brynjolfsson et al., 2025; Bick et al., 2025; Humlum and Vestergaard, 2025; Dillon et al., 2025; Hartley et al., 2025). I contribute to this literature by showing how AI is being adopted in higher education. I do so by developing an approach that allows me to recover how professors are integrating AI into the classroom from unstructured syllabus data. This allows me to uncover multiple margins of integration, and to distinguish actual classroom adoption from broader AI policy language (Chirikov, 2026). As such, I also contribute to recent work on how students are using artificial intelligence in higher education (Contractor and Reyes, 2025; Handa et al., 2025).

I further contribute to a literature examining how instructors shape the content of what is taught in higher education. Most closely related to this paper is Biasi and Ma (2022), who develop the approach of using course syllabi to recover curricular content at scale and show that research-active instructors are more likely to teach frontier knowledge in their courses. I contribute to this work by introducing an approach that uses large language models to reliably measure how instructors integrate AI into the classroom. Moreover, I show that AI integration has been led by young, research-active faculty members, a pattern that holds across fields and is most pronounced at selective universities.[3] I also find that professors are the main drivers of AI integration in the classroom, fitting in with a broader literature showing how instructors shape outcomes in higher education (Hoffmann and Oreopoulos, 2009; Carrell and West, 2010; Braga et al., 2016; De Vlieger et al., 2016; Feld et al., 2020).

[3]This aligns with recent evidence showing AI adoption is concentrated among younger workers (Humlum and Vestergaard, 2025; Bick et al., 2025).

Lastly, I contribute to a growing literature using large language models to recover nuanced information from text across different contexts (Bartik et al., 2024; Link et al., 2024; Braghieri et al., 2024; Bursztyn et al., 2025; Lagakos et al., 2025; Clayton et al., 2025; Chyn et al., 2025; Ash et al., 2025). Here, I show that large language models can reliably recover measures that require interpretation from unstructured syllabus text. I implement the approach using multiple local open-weight models, which makes the measurement approach transparent and reproducible (Korinek, 2023; Ludwig et al., 2025). I also validate the resulting measures against independent human raters to show that these models can reliably measure AI integration in higher education (Dell, 2025; Haaland et al., 2025). This paper thus also fits into a broader literature using text-as-data methods in economics (Baker et al., 2016; Gentzkow et al., 2019; Ash and Hansen, 2023), as I recover rich information from course syllabi on how instructors adapt their classes to a new technology.

The rest of the paper is structured as follows. Section 2 describes the data sources and presents descriptive statistics on the sample of university syllabi. Section 3 introduces the classification framework and measurement approach. Section 4 documents the prevalence and evolution of AI integration, along with its heterogeneity across majors and universities. Section 5 studies the drivers of AI integration, including the decomposition and course-level patterns. Section 6 examines who integrates AI into the classroom. Section 7 concludes.

2 Data Sources and Descriptive Statistics

To understand how AI has been integrated into higher education, I take advantage of course syllabi across public universities in the United States. College syllabi typically include detailed information on course content, topics covered, specific assignments and course policies, which allows me to observe how instructors incorporate AI into their classes. Following the rapid rollout of generative AI tools in higher education, I assemble a new dataset of publicly available syllabi from 50 public universities between 2024 and 2026. Moreover, I draw on a variety of external data sources to recover detailed characteristics of instructors, universities, college majors, and courses, to examine which characteristics drive AI adoption in the classroom.

2.1 University Syllabi

I construct this dataset by using publicly available syllabi from public universities across seven states.[4] I recover course syllabi for 50 universities, covering selective research universities such as UT-Austin and the University of Georgia, as well as broad-access institutions such as Florida A&M and West Texas A&M. For each syllabus, I recover the main information for the course, drawing on both structured fields in each university's data system and the syllabus text itself.

The main estimation sample includes syllabi for the courses offered between Fall 2024 and Spring 2026. I focus on these four academic terms, which cover the period of widespread

[4] Appendix B.1 describes the data sources and the construction of the main sample.

AI availability in higher education. Each syllabus in the main sample corresponds to a unique course taught by a specific instructor in a given semester. As such, the main sample includes 185,196 unique syllabi covering 82,129 unique courses taught by 65,843 different instructors in the 50 universities in the main sample (Table A1). For UT-Austin and UT-Dallas, I additionally recover 32,200 and 17,700 syllabi between Fall 2019 and Spring 2024, respectively, which I use to analyze outcomes before and after the rollout of ChatGPT.

Beyond the main course characteristics recovered from the syllabi, I extract a set of text-based features from each syllabus that capture relevant information about the course and the instructor (Appendix B.3). As a first pass at understanding how AI is being integrated into the classroom, I identify syllabi that include AI-related terms such as artificial intelligence, generative AI, large language model, and ChatGPT (Appendix B.3.6). This allows me to provide an initial description of the general prevalence of AI-related content across courses.

2.2 Instructor Characteristics

To examine which instructors are incorporating AI into their courses, I complement the syllabus data with a rich set of instructor characteristics. I draw on publicly available salary records, bibliometric databases, instructors' CVs, faculty directories, and institutional profiles to construct a detailed instructor-level dataset. These data allow me to examine how academic rank, gender, research productivity, and career experience drive AI adoption in the classroom.

To measure instructors' research productivity, I match the instructor roster to OpenAlex, an open-access bibliometric database (Priem et al., 2022). I match 39 percent of instructors in the sample (Appendix B.2), with cross-university variation in match rates reflecting institutional differences in faculty composition. From the matched profiles, I recover publication counts and dates, citation counts, and h-index as measures of research productivity and impact. Furthermore, I use the titles and abstracts of professors' publications to measure the content of their research. I also recover information on instructors' educational backgrounds from ORCID education records, faculty CVs, and departmental profile pages. Altogether, these data sources allow me to identify whether an instructor has any publications, along with their PhD institution and completion year.

I also recover academic rank from state salary databases, which record instructors' official institutional titles. I classify each instructor as a full, associate, or assistant professor, or as non-tenure-track faculty, which includes lecturers, adjuncts, and instructors of practice. For instructors' gender, I use direct records from the Texas salary database, along with a name-based probabilistic assignment using Social Security Administration baby-name data. All in all, I recover gender and academic rank for 82 and 69 percent of instructors in the sample, respectively (Appendix B.2).

2.3 University, Major, and Course Characteristics

To understand the extent to which the integration of AI into the classroom varies across universities, I also assemble a set of institutional characteristics from multiple data sources (Appendix B.1). First, I use data from the Integrated Postsecondary Education Data System (IPEDS) to capture measures of research intensity, faculty compensation, and average SAT scores. I additionally recover the median parental income for each institution from the Mobility Report Cards (Chetty et al., 2020). I further classify the 50 universities into three selectivity tiers based on the average SAT scores of their entering students, ranging from selective research universities to broad-access institutions. Table A2 presents these characteristics, showing sizable differences in research resources and faculty compensation across selectivity tiers.

I similarly recover a set of major-level characteristics from multiple data sources. In particular, I capture the mid-career median earnings, unemployment rate, gender composition, and share of graduates who earn a graduate degree for each major from the Federal Reserve Bank of New York's College Labor Market data (Abel et al., 2014). I also construct a measure of average SAT scores in each major from the National Postsecondary Student Aid Study (Appendix B.3). To understand the extent to which different majors are exposed to AI, I use the exposure measure proposed by Felten et al. (2023), which maps AI exposure scores from occupations to college majors. This measure intuitively captures the fact that AI exposure is higher for English and Marketing majors than for those in Nursing.

Since the 50 universities in the main sample use more than 2,700 distinct department codes, I harmonize them into 42 majors. For instance, I classify courses from Neuroscience, Microbiology, and Biochemistry departments into a harmonized Biology major (Table B4). I further aggregate the 42 majors into five macro-fields, covering STEM, Humanities, Social Sciences, Business, and Other. The harmonized majors show significant heterogeneity in AI exposure, earnings, and SAT scores across fields.

I also recover the academic level of each course from its university's course numbering conventions. Since nearly all universities follow a standard system in which the leading digit of the course number indicates the year of study, I classify courses as freshman, sophomore, junior, senior, or graduate (Appendix B.3). I use this five-level classification as the main variable, and also classify courses as basic undergraduate, advanced undergraduate, or graduate following Biasi and Ma (2022).[5]

2.4 Descriptive Analysis

To assess whether the main sample is representative of the courses offered at the 50 universities, I match the Spring 2026 syllabi to each university's published course schedule for the same term. I find that the syllabi in the sample cover 72 percent of offered courses at the median institution, and the field composition of the syllabi closely tracks that of the

[5] I also use the Florida Statewide Course Numbering System, which provides harmonized course identifiers across Florida universities and allows me to compare syllabi for the same course across institutions.

Table 1: Descriptive Statistics

	Share of Sample (1)	**Mentions AI** (2)	**No AI Mention** (3)
Instructor Gender			
Female	48.3	57.7	42.3
Male	51.7	50.6	49.4
Instructor Characteristics			
Tenure-Track	48.3	53.2	46.8
Any Publications	41.2	53.3	46.7
University Characteristics			
Selective	25.8	55.0	45.0
Moderately Selective	46.9	60.9	39.1
Broad Access	27.3	42.4	57.6
Major Characteristics			
Business	9.7	55.9	44.1
STEM	38.1	50.4	49.6
Social Sciences	20.8	56.8	43.2
Humanities	28.9	57.7	42.3
Course Characteristics			
Undergraduate	80.7	55.1	44.9
Graduate	19.3	51.6	48.4
Observations	102,905	55,934	46,971

Table 1 shows descriptive statistics for the Spring 2026 cross-section across the 50 universities in the main sample. Column (1) shows the share of the sample belonging to each row. Columns (2) and (3) give, within each row group, the share of syllabi that mention and do not mention AI. For the Female and Tenure-Track rows, I focus on syllabi with non-missing instructor information. University selectivity follows the three-tier classification. Business, STEM, Social Sciences, and Humanities denote the harmonized macro-fields, and Undergraduate combines basic and advanced undergraduate courses.

offered courses (Table B6). As such, the sample tracks the courses offered at each university (Appendix B.4).

Table 1 presents descriptive statistics for the main sample, focusing on the Spring 2026 cross-section across the 50 universities. Half of the courses in the sample are taught by tenure-track faculty, while 37 percent are taught by instructors who have published at least one paper, consistent with the composition of the instructor workforce at these institutions.

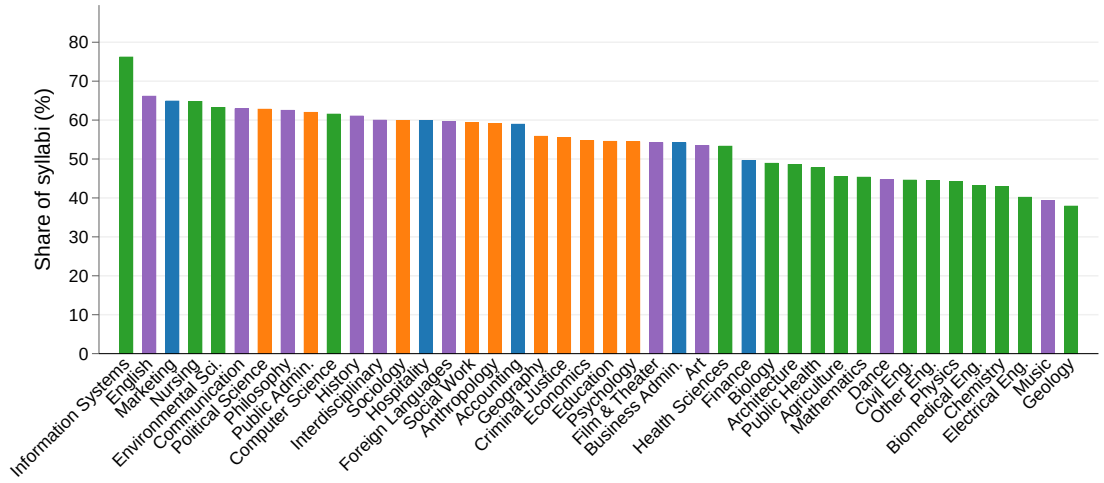
The next two columns split the sample by whether the syllabus mentions artificial intelligence, which 54 percent of syllabi in this semester do. Women are more likely to mention AI in their course syllabi than men, but there are no corresponding differences in this measure across professors' tenure-track status. Lastly, undergraduate courses are more likely to mention AI than graduate courses, and this is also more common at broad-access universities.

Figure 1 shows there is significant heterogeneity in the prevalence of AI mentions across the harmonized majors in the sample. For instance, close to 80 percent of courses in Sociology discuss AI in the syllabi, compared to about half in Economics. Meanwhile, course

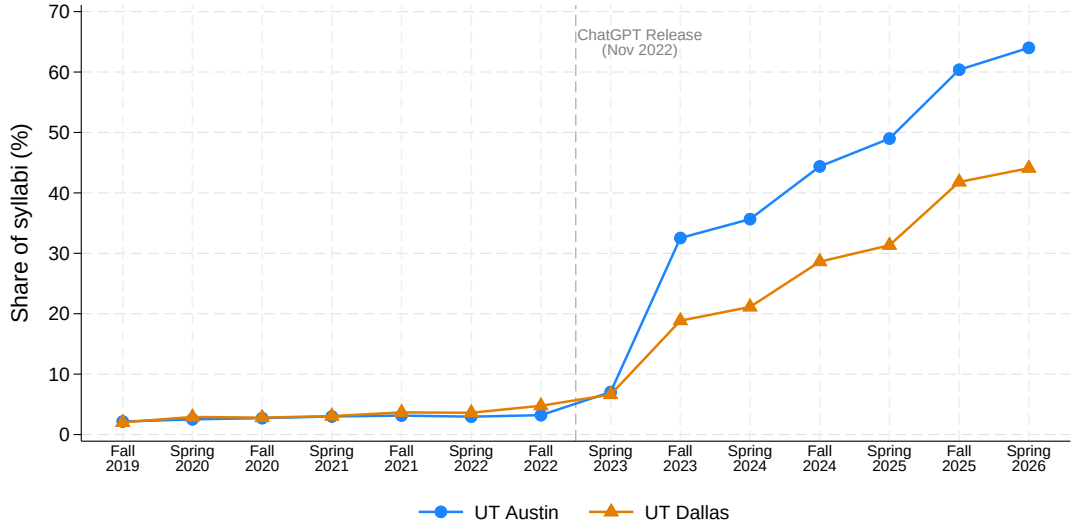
syllabi in STEM majors are the least likely to include explicit references to artificial intelligence.

Figure 1: AI prevalence across disciplines and over time

(a) By Major



(b) Time Series



Note: Figure 1 shows the prevalence of AI mentions in the syllabi. Panel (a) shows the Spring 2026 share of syllabi mentioning AI by harmonized major, ordered from highest to lowest prevalence and colored by macro-field. Panel (b) shows the term-by-term share of syllabi mentioning AI for UT Austin and UT Dallas from Fall 2019 through Spring 2026. The vertical dashed line marks the public release of ChatGPT in November 2022. AI mentions are identified using the keyword-based text analysis applied to the full syllabus text.

I draw on the longitudinal UT-Austin and UT-Dallas sample to measure the evolution of AI mentions in course syllabi dating back to 2019-2020. Figure 1 shows that a small share of syllabi already discussed artificial intelligence prior to the introduction of ChatGPT, typically in Computer Science and Business courses. At the same time, the share of syllabi mentioning AI increased sharply following the public release of ChatGPT in November 2022, with the first significant response emerging in the Fall 2023 semester at both universi-

ties. By Spring 2026, 64 percent of syllabi at UT Austin and 44 percent at UT Dallas mention artificial intelligence.

While this analysis shows that a significant share of syllabi now mention artificial intelligence, this approach combines fundamentally different approaches into a single measure. For instance, a syllabus stating that "AI-generated submissions constitute academic dishonesty" and one stating that "students are expected to use ChatGPT for the course project" both mention AI, but they capture diametrically opposed pedagogical stances.

3 Measuring AI Integration in the Classroom

In this section, I introduce an approach that captures how instructors are integrating artificial intelligence into higher education. I focus on instructors as they directly shape what students learn in college, and their decision to integrate AI into the classroom may shape students' familiarity with these tools. Integrating AI into a course is not the same as keeping course content up to date with recent advances in the field, which flows naturally from a professor's own research activity. AI integration instead requires deliberate pedagogical redesign around a general-purpose technology that cuts across fields (Bresnahan and Trajtenberg, 1995).

To understand what instructors are actually doing with AI in their classrooms, I develop measures that capture how they are integrating it into their teaching and evaluation. I focus on the teaching margin, as a professor who schedules a session on prompt engineering or designs a hands-on AI exercise is committing class time to engaging with artificial intelligence. A typical undergraduate semester includes 14-16 weeks of instructional time, so these decisions displace traditional content and reveal the instructor's judgment about what knowledge students need (Comin and Hobijn, 2010). Similarly, I analyze how professors integrate AI into student evaluation, as they may have also created and adapted their assignments in response to this new technology (Acemoglu and Restrepo, 2019).

3.1 AI Integration in Teaching and Evaluation

To understand the prevalence of AI classroom integration, I first worked with research assistants who read a large number of syllabi to identify the main ways in which professors incorporate AI into their teaching and evaluation (Haaland et al., 2025). Based on this analysis, I construct a teaching integration measure that captures whether AI is part of what the course teaches, through its class sessions, topics, or course materials. In particular, this measure identifies courses that dedicate a session or module to an AI topic, describe a class activity or demonstration involving AI, or assign readings or other materials specifically about AI, signaling that the instructor views AI literacy as knowledge worth developing (Javadian Sabet et al., 2024). This margin is central to understanding classroom adoption, as it reflects a deliberate change to the content that students are exposed to.

On the evaluation margin, I consider whether at least one graded or assessed task requires students to engage with AI. This measure identifies assignments in which students

must use an AI tool, submit a deliverable that includes or builds on AI output, or complete a required step that has them produce, compare, critique, or revise AI content. This margin captures instructors who bring AI directly into graded work, whether to help students develop skills in using these tools or simply to allow AI-assisted work in their assignments. Altogether, these measures capture how instructors have redesigned their courses to explicitly integrate artificial intelligence into their teaching and evaluation.

Beyond the classroom integration margin, I also examine whether professors explicitly address AI in their syllabi and what stance they take toward its use. In this context, I analyze whether the syllabus contains any explicit policy regarding student AI use, and if so, whether the instructor explicitly allows students to use artificial intelligence in their course. These two measures capture the difference between instructors who address AI and those who affirmatively welcome it into their courses.

To capture these dimensions from the syllabi, I construct the measures as binary questions, as this format tends to yield high inter-rater reliability rates (Dolnicar et al., 2011). Each question includes brief guidance clarifying the classification criteria, following standard practice in content analysis (Lombard et al., 2002; O'Connor and Joffe, 2020). Table 2 presents the four classification questions with their accompanying guidance.

Table 2: AI Integration Classification Questions and Guidance

AI Policy. Does the syllabus contain an explicit policy, rule, or guideline about student use of AI or generative AI tools (such as ChatGPT, Copilot, or similar tools) for coursework?
Guidance: True if the syllabus states rules, permissions, prohibitions, guidelines, or expectations about whether and how students may use AI tools for coursework. The policy may come from the instructor or the university, may appear anywhere in the syllabus, and may permit, prohibit, restrict, or conditionally allow AI use.
AI Permitted. Does the syllabus explicitly state that students are permitted, allowed, or welcome to use AI tools for coursework in this course?
Guidance: True if the syllabus contains an explicit statement permitting, allowing, welcoming, or approving student use of AI tools for coursework.
Teaching Integration. Is AI part of what this course teaches: its class sessions, topics, or course materials?
Guidance: True if the course dedicates class time or course content to AI: a schedule, module, or session entry on an AI topic, a class activity, lecture, discussion, or demonstration involving AI, or an assigned reading, video, tutorial, or other course material.
Evaluation Integration. Does at least one graded or assessed task require students to engage with AI?
Guidance: True if a specific assignment, project, exam, or task requires AI in the student's own work: the task requires students to use an AI tool, the deliverable must include or build on AI output, or a required step has students produce, compare, critique, or revise AI content.

Note: Table 2 presents the four binary classification questions used to measure AI integration in course syllabi. I classify each measure independently on the full syllabus text. Appendix C presents the full coding guidance with worked examples for each question.

Before applying these measures at scale, I validate these constructs against human raters. In particular, four students independently classified a sample of 60 syllabi across all four

measures.[6] Table 3 presents the inter-rater reliability results. The four raters agree on 95 percent of classifications on average across the four questions (Appendix C). While agreement rates are higher for questions with clear textual markers than for those requiring more interpretive judgment, these results provide a reliable benchmark for the LLM-based classification (Dell, 2025; Haaland et al., 2025).

3.2 LLM Classification Implementation

While the human raters described in Section 3.1 agree on 95 percent of classifications, coding 150,000 syllabi by hand would be prohibitively time-consuming. To classify these syllabi at scale, I take advantage of large language models, which process text by attending to the full context of each word, resolving ambiguities that keyword matching cannot (Ash and Hansen, 2023; Dell, 2025).

I implement the classification approach using open-weight models with fixed, published parameters, which I run locally to ensure the results are reproducible (Korinek, 2023). This avoids well-known problems with proprietary models, which are frequently updated or deprecated, and do not yield reproducible results even with fixed temperatures (Barrie et al., 2024). I classify each syllabus using three independent open-weight models and take the modal answer across them. I validate the resulting classifications against human raters, following the recommendation in Ludwig et al. (2025) to benchmark automated classifications against human judgments for the specific task at hand.

For each syllabus, I submit independent calls for each of the four classification questions, so that a model's response to one question does not shape its evaluation of the remaining questions. In addition to requiring a binary classification, I ask each model to provide a confidence rating, a brief justification, and direct evidence from the syllabus text when the answer is affirmative. This allows me to assess the quality of the model's reasoning and to recover detailed information on the ways in which AI appears in the classroom. In particular, I classify each question using three open-weight models, Gemma 4-31B, GPT-Oss 120B and Qwen 3.6-35B, which rank among the top-performing open-weight models on standard evaluation leaderboards (Chiang et al., 2024).[7] As such, each model independently classifies all syllabi, providing three complete sets of classifications.

For each question, I define the final classification as the modal answer across the three models. The evidence quotes that feed the qualitative analysis of how instructors integrate AI into the classroom come from a single default rater.

The three models generally agree in their answers across the four questions. Pairwise agreement rates range from 94 to 99 percent, highest for the policy question and lowest for the permission question, which requires more interpretive judgment (Table C4).

[6] All four raters classify 20 core syllabi, and the remaining 40 are rotated so that each rater classifies 50 in total, with stratification by AI mention status, institution, and classification profile (Krippendorff, 2004).

[7] I set the temperature, presence penalty, and frequency penalty parameters to zero, the top_p parameter and the repetition penalty to one, and fix the random seed, to maximize deterministic outputs (Appendix C).

3.3 Validation

I validate the LLM-based classification approach using the Alternative Annotator Test (AAT) of Calderon et al. (2025), which evaluates whether an automated classifier can serve as a credible substitute for human coders. At its core, the test asks whether the automated classifier agrees with the human consensus at least as often as the individual human raters do. The logic of this test is as follows: for each syllabus-question pair, I hold out one of the four human raters and compute the modal response of the remaining three as the benchmark. I then compare whether the modal LLM label or the held-out rater is more likely to agree with this benchmark. The LLM passes the test if it matches the consensus at least as well as the held-out rater, indicating that it can be reliably used to extend the human classifications to the full set of syllabi.

Table 3 presents the validation results. The average leave-one-out agreement rate among the four human raters equals 97 percent, and the modal LLM label agrees with the modal human response in 95 percent of syllabus-question pairs. The proposed classification approach thus comfortably passes the Alternative Annotator Test, as it performs as well as humans in 96 percent of questions, far exceeding the required 50 percent threshold.

Table 3: Agreement with Modal Human Response (%)

	Individual Raters						
Question	**R1** (1)	**R2** (2)	**R3** (3)	**R4** (4)	**Mean** (5)	**Model** (6)	**AAT** (7)
AI Policy	88	100	100	96	96	93	95
AI Permitted	100	96	98	94	97	93	95
Teaching Integration	93	95	95	96	94	92	93
Evaluation Integration	99	100	100	100	100	98	99
Average	95	98	98	96	97	94	96

Note: Table 3 compares the automated classification to human raters across the four headline measures. Each rater column gives leave-one-out agreement with the modal response of the remaining raters. Mean is the average across the four human raters. Model gives the agreement rate between the modal classification across the three models and the human modal response. AAT gives the win-plus-tie rate from the Alternative Annotator Test of Calderon et al. (2025), where rates above 50 percent indicate human-level performance. The validation sample consists of 60 syllabi drawn using stratified sampling by AI mention status, institution, and classification profile.

Moreover, each individual model performs similarly well against the human modal response (Appendix C). Altogether, these validation results establish that the LLM-based classification reliably measures AI policies and integration at scale, achieving agreement with human raters that is comparable to the agreement among the raters themselves. I therefore use the modal classification to characterize AI integration across the full sample of syllabi.

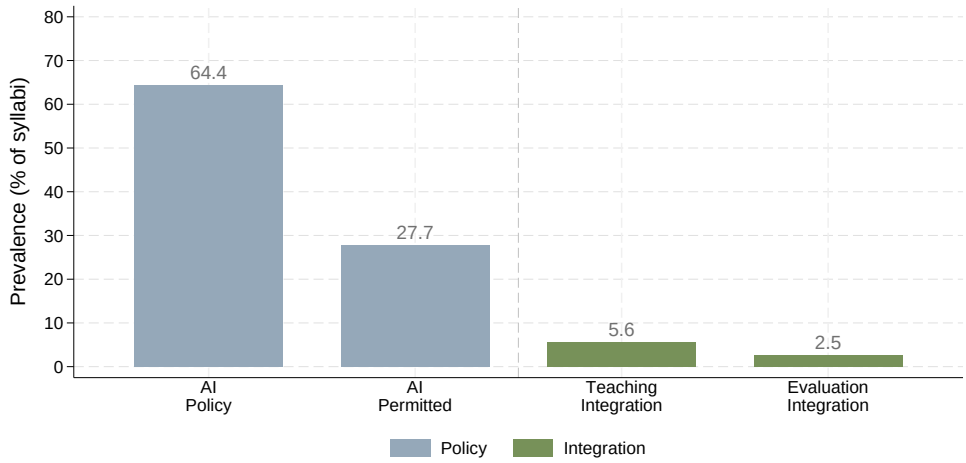
4 AI Integration in Higher Education

4.1 Prevalence of AI Integration

Figure 2 presents the extent to which artificial intelligence has been integrated into higher education in the Spring 2026 semester. First, close to 65 percent of courses in the

sample include an explicit AI policy in their syllabi.[8] This is largely in line with the share of syllabi that explicitly mention AI, as discussed in Table 1. At the same time, just 28 percent of courses in the Spring of 2026 allow students to use these tools for their classes. The majority of professors thus have AI policies in place but do not explicitly endorse students using AI in their classes.[9]

Figure 2: Spring 2026 Prevalence of the Four AI Measures



Note: Figure 2 shows the prevalence of the four AI measures in Spring 2026. The bars show the share of syllabi with each measure. In Spring 2026, 64.4% of syllabi include an AI policy, 27.7% explicitly permit AI use, 5.6% display teaching integration, and 2.5% display evaluation integration.

Figure 2 also shows that explicit integration of artificial intelligence into teaching is far less common than AI-focused policies. All in all, 6 percent of syllabi display concrete evidence of AI integration into teaching, whether through course content, class activities, or assigned readings. These results suggest that the main response to generative AI has taken the form of formal acknowledgment rather than direct integration into the classroom (McDonald et al., 2025).

AI integration into actual evaluation is even rarer in this context. By 2026, just 3 percent of courses include a graded task that requires students to engage with AI. Requiring students to engage with AI in graded work represents a distinct margin of integration. These assignments may help students develop skills in using these tools, though they may also capture coursework that simply builds on AI-assisted work.

Qualitative Evidence. To further understand what these integration measures capture, I sort every syllabus classified as integrating AI into teaching or evaluation into a small set of descriptive categories. Rather than relying on another model judgment, the sorting applies fixed lists of keywords to the passage of syllabus text that the classifier quotes as evidence, which isolates the exact language describing how AI enters each course (Appendix C de-

[8] In the full 2024–2026 sample, 57 percent of syllabi include an AI policy and close to 6 percent integrate it into their teaching. This reflects lower integration rates in the earlier semesters in the sample.

[9] Figure A1 shows the prevalence of these measures is broadly similar across the three classification models.

scribes the approach).

On the teaching side, the most common category involves AI applied to the course's own field, accounting for 27 percent of the syllabi with teaching integration, while courses in which students work with AI tools and courses that cover AI methods each account for 18 percent (Table A3). These categories also differ clearly in the language instructors use, as the terms most distinctive of AI Methods courses include machine learning and neural networks, while Using AI Tools courses are instead marked by references to ChatGPT and prompts (Figure A2), signaling that they capture distinct forms of integration. On the evaluation side, integration is dominated by graded tasks in which students produce their own work with AI assistance.[10]

AI Integration Over Time. Figure 3 shows how formal permission for AI use and classroom integration evolved at UT Austin and UT Dallas from Fall 2019 through Spring 2026. Reassuringly, I first find that no courses explicitly permit AI use prior to the rollout of ChatGPT. However, the share of syllabi with explicit permission increased rapidly over the following three years, so that close to one-quarter of courses at these universities now allow students to use generative AI. Classroom integration increased much more slowly, as teaching integration rose only to about 10 percent by Spring 2026.[11] Most of that integration still comes through the teaching margin, while evaluation remains comparatively rare throughout the period. Taken together, these patterns suggest that instructors adjusted formal permissions much faster than they changed what students were actually asked to do in class.

Table A4 presents the corresponding semester-by-semester changes for the four measures over the past four semesters. I find the same pattern in this broader sample, as the share of syllabi that explicitly permit AI use rises from 14 percent in Fall 2024 to 30 percent in Spring 2026. Classroom integration, in contrast, increases far more gradually over the same period, again driven by the teaching margin. Altogether, even in this shorter window, colleges appear to have moved faster in formalizing permission for AI than in incorporating it into classroom practice.

4.2 Heterogeneity Across Majors

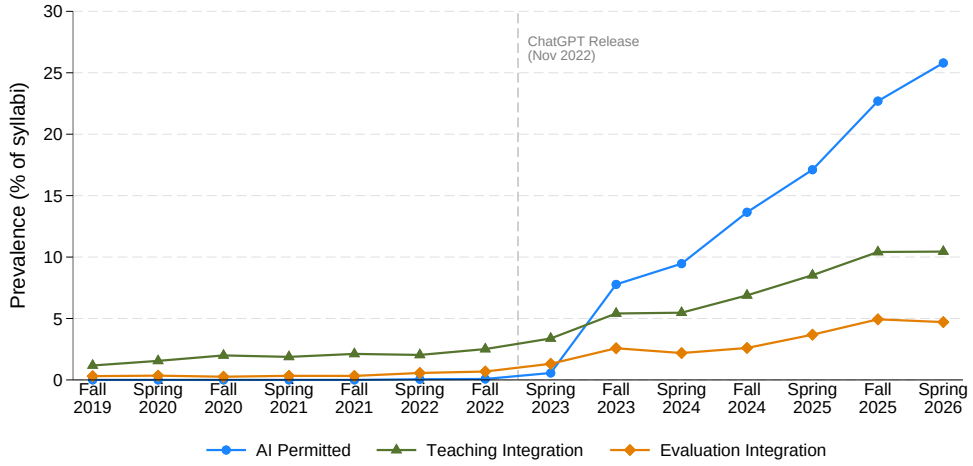
The results presented so far show that college professors have gradually started integrating AI into the classroom. In this context, as different fields prepare students for jobs with different exposure to generative AI (Eloundou et al., 2024), these patterns may also vary across majors (Contractor and Reyes, 2025; Handa et al., 2025).

Figure 4 presents evidence on the extent to which AI has been differentially integrated across college majors in Spring 2026. I find significant differences in adoption outcomes, as 16 percent of courses in Marketing have integrated AI into their teaching, compared

[10] This category accounts for 38 percent of the syllabi with evaluation integration, followed by tasks in which students build AI models and tasks in which students interact with or judge AI output.

[11] A small share of courses already integrated AI into teaching before ChatGPT, largely because some syllabi discussed AI as a course topic, especially in computer science and business fields. This baseline reflects real course content rather than early use of generative AI tools.

Figure 3: AI Policy and Classroom Integration Over Time at UT Austin and UT Dallas



Note: Figure 3 presents changes over time in AI policy and classroom integration in the balanced-course panel of UT Austin and UT Dallas from Fall 2019 through Spring 2026. Balance requires a course to appear in at least one term in both 2024–2025 and 2025–2026. The three lines track the share of syllabi that explicitly permit AI, teaching integration, and evaluation integration. The vertical dashed line marks the launch of ChatGPT in November 2022.

to just one percent of courses in Music, Dance, and Geology. Altogether, 10 percent of courses in Business majors have integrated AI into their teaching, which roughly doubles the adoption rate for majors in the Humanities, STEM, and Social Sciences.[12] While this integration is more likely to occur through the teaching margin than through AI-related assignments (Figure A4), the two margins are strongly correlated across majors. Majors that are quicker to incorporate AI into course activities are also more likely to integrate it into graded assignments.

On the other hand, courses in Social Sciences and in the Humanities are more likely to include an explicit AI policy in their syllabi (Table A5). This mirrors the differential prevalence of AI mentions across fields presented earlier. At the same time, this table shows that 32 percent of classes in Business majors explicitly allow students to use AI, which is far higher than in the other majors. While instructors in Business classes are less likely to include an explicit AI policy, they are much more likely to permit students to use it when they do. Majors in Social Sciences and the Humanities have thus moved rapidly to adopt policies without making corresponding changes to integrate AI into the classroom.[13]

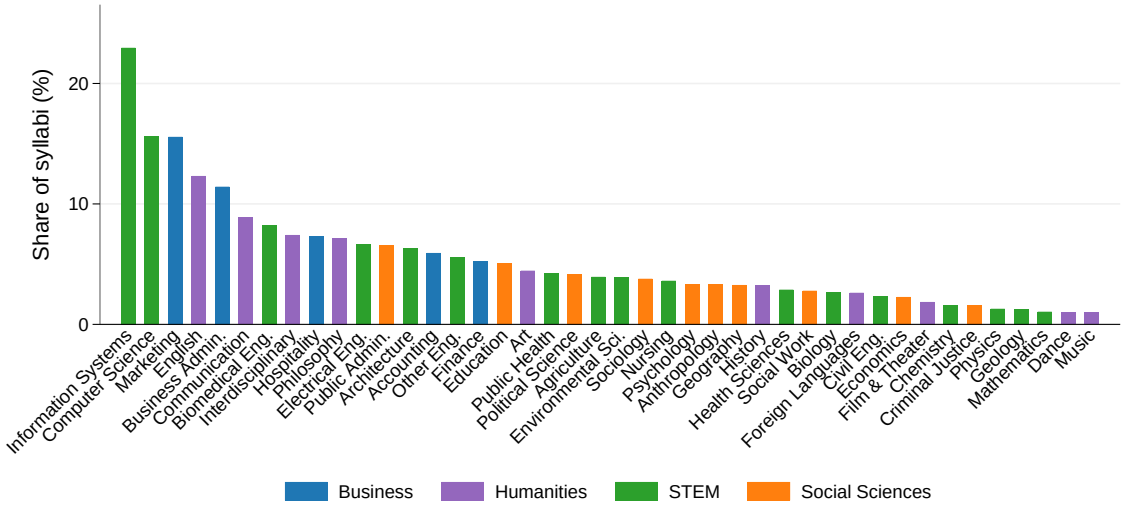
4.3 Heterogeneity Across Universities

I also examine whether the prevalence of these measures varies significantly across the universities in the sample. The first panel of Figure 5 shows that some universities have incorporated AI policies in almost all of their courses, and that this is more likely to have happened at less selective universities. While almost all courses at Florida Atlantic include

[12] While Business majors at UT-Austin and UT-Dallas were more likely to engage with AI content in the classroom prior to ChatGPT, the cross-field differences grow substantially after Fall 2022 (Figure A3).

[13] English classes stand out as an important exception, as this major is both more likely to have explicit AI policies and to have explicitly integrated it into the classroom.

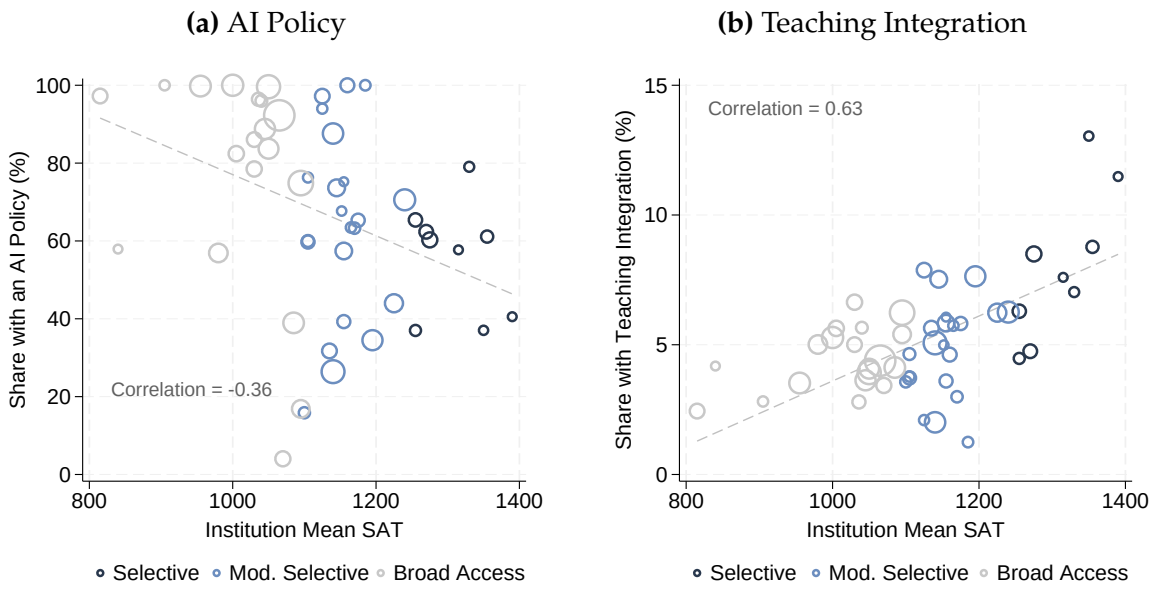
Figure 4: Teaching Integration Across Majors



Note: Figure 4 presents the share of syllabi with teaching integration in Spring 2026 across 42 harmonized majors, with bars colored by broad field (Business, Humanities, STEM, Social Sciences, Other).

an AI policy in the Spring 2026 semester, this is the case in less than half of courses at the University of Florida (Table A6).

Figure 5: AI Integration and Mean SAT Score



Note: Figure 5 presents university-level scatter plots of AI integration against mean SAT score in Spring 2026. Panel (a) plots AI policy prevalence against mean SAT; panel (b) plots teaching integration against mean SAT. Each point represents one of the 50 universities.

On the other hand, the second panel of Figure 5 shows that more selective institutions are more likely to have incorporated generative AI into their teaching. About 8 percent of courses at selective institutions have already integrated AI, compared to around five percent at moderately selective and broad-access universities. In particular, while more

than 12 percent of courses at UT-Austin and UT-Dallas have already integrated AI into their teaching, this share remains below 6 percent at West Texas A&M. Importantly, the selectivity gradient is mostly present on the teaching margin. This suggests that instructors' decisions around lecture materials and classroom activities may be shaping AI integration patterns across universities.

Altogether, these patterns are consistent with the growing evidence that larger, higher-paying, and better-resourced firms are more likely to adopt AI more deeply (McElheran et al., 2024; Yotzov et al., 2025). As such, a similar gradient is emerging in higher education as more selective institutions are far more likely to integrate it into the classroom.

5 Drivers of AI Integration

5.1 University and Major Characteristics

The results presented so far show that the integration of AI into the classroom varies significantly across universities and college majors. To understand which universities have taken the lead on this front, I draw on the extensive set of university-level characteristics described in Section 2. In particular, I analyze how university resources and selectivity shape the prevalence of the main outcomes by estimating the following equation:

$$Y_n^{(k)} = X'_{u(n)}\beta_u^{(k)} + \phi_{m(n),l(n),t(n)} + \varepsilon_n^{(k)}, \qquad k \in \mathcal{K}, \tag{1}$$

where $Y_n^{(k)}$ is a binary variable measuring the prevalence of one of the AI-integration measures (k) in syllabus n. Meanwhile, $X_{u(n)}$ captures the main set of university-level characteristics, including enrollment levels, graduate enrollment share, research expenditures, endowment levels, average faculty salary, mean SAT, and median parental income. Equation (1) includes major-by-course-level-by-term fixed effects ($\phi_{m(n),l(n),t(n)}$), allowing me to identify how AI adoption differs across universities within the same major, course level, and term. In practice, I estimate equation (1) separately for each outcome in $\mathcal{K}$, first using one covariate at a time and then including the full vector of university characteristics simultaneously. I cluster standard errors at the university level.

Moreover, given the sizable differences in adoption across majors documented above, I similarly examine which major characteristics are related to the main outcomes of interest. I draw on the main major-level covariates described in Section 2, which include the main measure of AI exposure, along with major-level mid-career earnings, unemployment rate, graduate-degree share, mean SAT, and the share of female students in the field. To understand the drivers of AI integration at the major level, I estimate a modified version of equation (1), replacing the university-level covariates with the vector of major characteristics ($X'_{m(n)}$), while including university fixed effects along with course-level-by-term fixed effects. This specification compares the prevalence of AI integration for different majors within the same university and semester, thus allowing me to understand how these characteristics shape differential adoption across fields.

University-Level Results. Figure 6 presents the results from equation (1), showing how university-level characteristics shape the prevalence of AI policy language in syllabi. This figure shows the results from a specification that includes one covariate at a time, as well as all covariates jointly. The pairwise results show that universities with lower average SAT scores are more likely to include AI policy language, confirming the evidence presented in Figure 4. Moreover, universities with lower average faculty salaries, and with lower graduate enrollment shares are similarly more likely to include policy language in their syllabi. However, these coefficients are not significant in the specification that includes all controls at the same time.

I also analyze how these characteristics shape the extent of teaching integration in the second panel of Figure 6. First, universities with larger graduate sectors and those with higher research expenditures per student are more likely to integrate AI into their teaching. This pattern also emerges for institutions with higher average faculty salaries, as a one-SD increase in salaries predicts a 1.2pp increase in teaching integration. While these coefficients become attenuated in the multivariate specification, these results suggest that universities with stronger research environments are more likely to integrate AI into their teaching.[14]

This evidence thus indicates that universities with stronger research environments are the most likely to have integrated AI into the classroom. This is consistent with Biasi and Ma (2022), who find that universities with more research-active faculty teach courses significantly closer to the knowledge frontier, even after controlling for other institutional characteristics. Moreover, the existing firm-level evidence similarly shows that more innovative companies are more likely to have integrated artificial intelligence (McElheran et al., 2024), as are firms that pay higher wages (Yotzov et al., 2025). These results suggest that a similar gradient may be emerging in higher education.

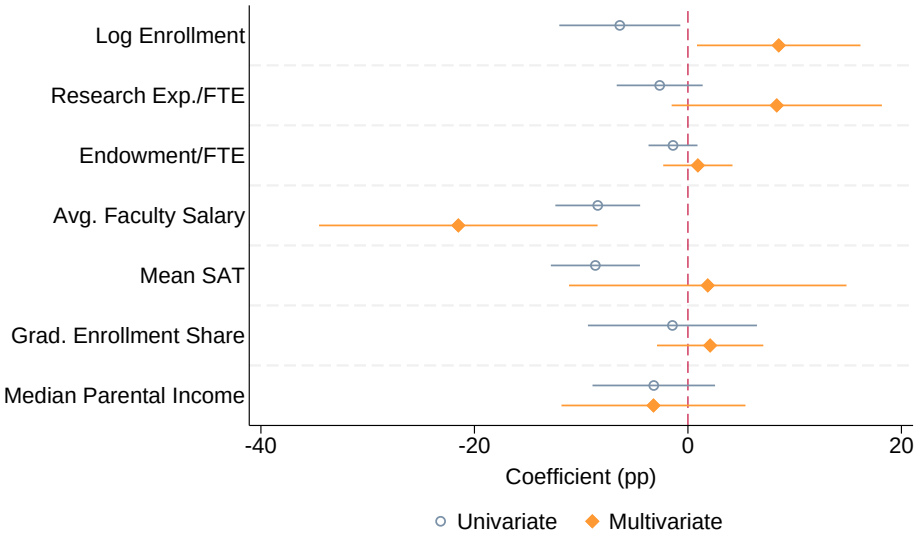
Major-Level Results. Figure 7 shows how major-level characteristics shape AI policy adoption conditional on university and course-level fixed effects. Majors with a larger share of female students, lower graduate-degree shares, and lower mid-career earnings are more likely to have an explicit AI policy in place. Importantly, I also find that majors with a one SD higher exposure to AI are 2 percentage points more likely to have put a policy in place. Majors with greater AI exposure are thus more likely to address it in their syllabi, once other differences across fields are accounted for.

Figure 7 also presents the corresponding results for teaching integration. First, majors with lower graduate-degree shares are more likely to integrate AI into their teaching, as are those with lower average SAT scores. I also find that majors with greater AI exposure are significantly more likely to have integrated these tools into their courses. AI exposure predicts AI-based teaching much more strongly than evaluation (Figure A6). This fits in with the evidence presented in Section 4.2, where the fields with the highest rates of classroom

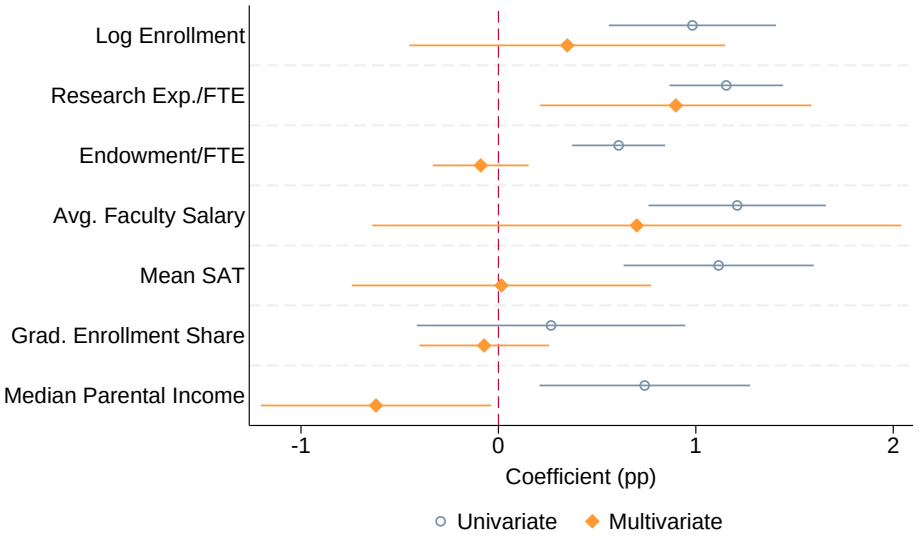
[14] Institutions with higher mean SAT scores and higher median parental income are also more likely to integrate AI into their teaching, yet these coefficients are insignificant in the multivariate specification. Figure A5 presents the corresponding results for AI permitted and evaluation integration.

Figure 6: University Covariates and AI Integration

(a) AI Policy



(b) Teaching Integration



Note: Figure 6 presents coefficient estimates from regressions of AI integration on seven university-level covariates. Panel (a) shows results for having an AI policy; panel (b) shows results for teaching integration. Circles show pairwise estimates; diamonds show multivariate estimates including all covariates simultaneously. Bars represent 90% confidence intervals. Fixed effects: major × course level × term. Standard errors clustered at the university level.

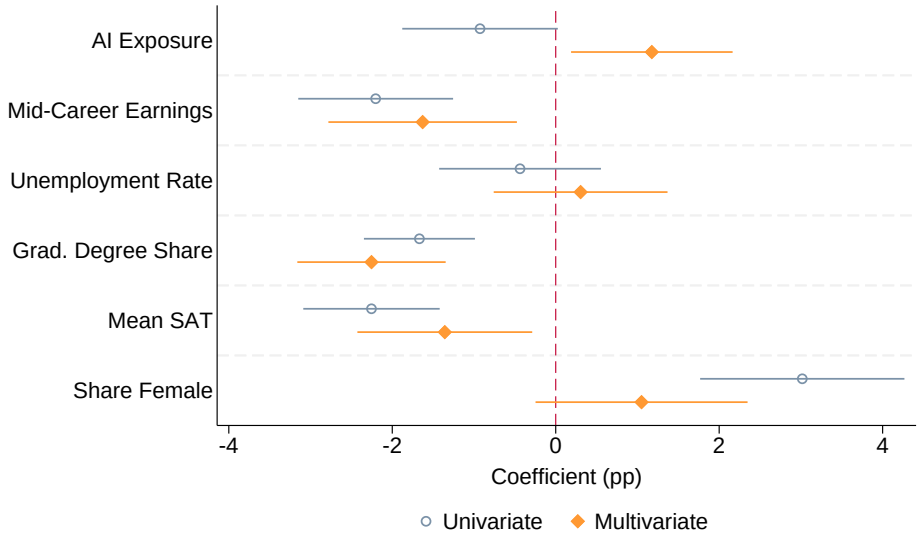
integration are also those where students and workers are most likely to use generative AI.

5.2 Course Characteristics

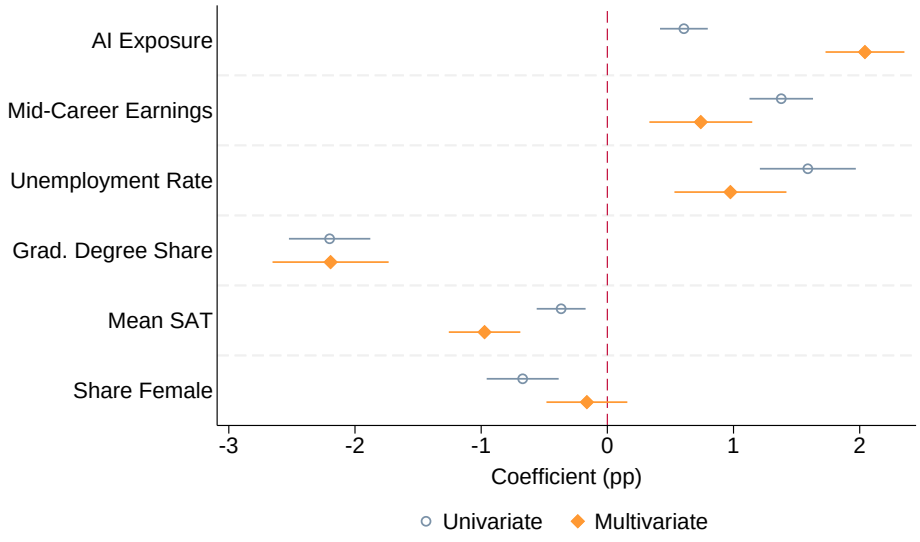
In Table A7, I further examine the extent to which AI has been differentially integrated across different course levels. Explicit policy language is more common in undergraduate

Figure 7: Major Covariates and AI Integration

(a) AI Policy



(b) Teaching Integration



Note: Figure 7 presents coefficient estimates from regressions of having an AI policy (panel a) and teaching integration (panel b) on six major-level covariates. Circles show pairwise estimates; diamonds show multivariate estimates including all covariates simultaneously. Bars represent 90% confidence intervals. Fixed effects: course level × term and university. Standard errors clustered at the university level.

courses than in graduate courses.[15] By contrast, graduate classes are most likely to have integrated AI into teaching. Altogether, 8 percent of graduate courses had incorporated AI into their teaching by the Spring of 2026. In this sense, graduate courses resemble the most selective universities in the sample, with less AI policy language but more actual classroom integration.

[15]More than 60 percent of both basic and advanced undergraduate syllabi contain an AI policy, compared with less than half of graduate syllabi.

To understand whether these patterns hold when accounting for differences across institutions, I estimate a modified version of equation (1). This specification regresses the main outcomes on observable course characteristics while including university-by-major-by-term and instructor fixed effects. I thus compare different courses taught by the same instructor while accounting for the broader teaching context in that term. As such, I analyze how enrollment levels, credit hours, and course levels shape AI integration outcomes.

Figure 8 shows that AI policy language is concentrated in introductory courses. Conditional on instructor and university-by-major-by-term fixed effects, freshman courses are about 5 percentage points more likely than graduate courses to include an AI policy, and this difference narrows at each successive undergraduate level. Larger courses and courses with teaching assistants are also more likely to include AI policy language, fitting in with the broader pattern that these high-enrollment settings have been the primary site of the institutional response to AI.

On the other hand, graduate-level courses are more likely than advanced undergraduate classes to have actually integrated AI into the classroom. This difference between the policy and integration results is especially clear on the teaching side, where advanced undergraduate courses are significantly less likely than graduate courses to dedicate class content to AI. In this context, graduate courses may have more natural scope for incorporating AI into teaching, precisely because their content engages with frontier knowledge (Biasi and Ma, 2026).[16]

5.3 Variance Decomposition

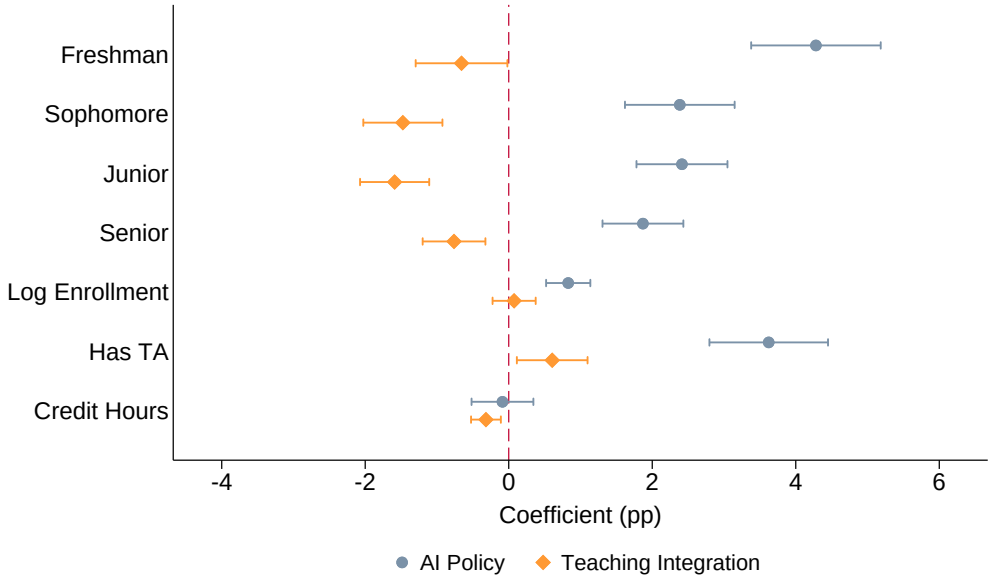
To understand the extent to which instructors, courses, and the broader academic environment contribute to shaping AI integration, I estimate the following equation:

$$Y_n^{(k)} = \alpha_{i(n)} + \lambda_{c(n)} + \phi_{u(n),m(n),l(n),t(n)} + \varepsilon_n^{(k)}, \qquad k \in \mathcal{K}, \tag{2}$$

where $\alpha_{i(n)}$ captures instructor fixed effects, whereas $\lambda_{c(n)}$ denotes local course fixed effects, representing university-specific courses. These course effects absorb stable features of a given course, such as its recurring content, structure, or assessment design, that may shape the scope for AI integration. Meanwhile, $\phi_{u(n),m(n),l(n),t(n)}$ denotes university-by-major-by-course-level-by-term fixed effects, which absorb the broader institutional and disciplinary context in which the syllabus is written. In practice, the decomposition is informed by comparisons across syllabi that share part of the same academic environment while differing along other dimensions. For instance, when the same local course is taught by different instructors, the remaining variation helps recover the instructor margin. Similarly, when instructors appear across different courses or terms within the same broader environment, the specification reveals how much of the variation is tied to the course itself versus the

[16] In the UT Austin and UT Dallas subsample, courses that historically relied more on take-home assignments are more likely to have integrated AI into teaching and evaluation (Table A8).

Figure 8: Course Characteristics and AI Integration



Note: Figure 8 shows the multivariate course-characteristic coefficients for AI Policy and Teaching Integration. The specification replaces local course fixed effects with observable course characteristics, includes instructor fixed effects and university-by-major-by-term fixed effects, and omits Graduate courses as the reference category. Missing indicators are included for Log Enrollment, Has TA, and Credit Hours. Bars represent 90% confidence intervals. Standard errors are two-way clustered by instructor and local course.

surrounding institutional and disciplinary context.[17]

To further understand whether AI integration is driven by individual instructors, by the courses they teach, or by the universities and majors in which they operate, I implement a Shapley–Owen decomposition (Audoly et al., 2025). In particular, this decomposition computes the average marginal contribution of each of these components to adjusted R^2 across all possible orderings of inclusion.[18] The Shapley–Owen decomposition thus recovers the partial contributions of instructors, courses, universities, and majors to AI integration in an additive and order-invariant approach.

Empirical Results. Table 4 presents the results from the variance decomposition of equation (2) for the main measures of AI integration. These results show that AI policy is shaped largely by instructors. That is, instructors explain about 37 percent of the variance in whether a syllabus includes AI policy language, while local course fixed effects account for an additional 25 percent of the variation. I find a similar pattern for whether AI is explicitly permitted. There too, instructors explain the largest share of the variance, local courses explain a smaller but still important share, and the broader environment accounts for relatively little.

[17] I also estimate a more stringent version that replaces $\lambda_{c(n)}$ with local-course-by-year fixed effects (λ_{cy}), tightening the comparison to instructors teaching the same course in the same academic year.

[18] Formally, for a given block j, the Shapley contribution equals $\sum_{T \subseteq V \setminus \{j\}} \frac{|T|!(K-|T|-1)!}{K!} \left[R^2(T \cup \{j\}) - R^2(T)\right]$, where $R^2(S)$ denotes the adjusted R^2 from the specification that includes the subset of blocks S, V is the full set of blocks, and $K = |V|$.

Table 4: Variance Decomposition of AI Integration

	AI Policy (1)	AI Permitted (2)	Teaching Integration (3)	Evaluation Integration (4)
Instructor	36.9	41.5	31.2	31.3
Course	24.6	23.2	30.5	28.6
University	7.8	1.8	0.3	0.2
Major	1.3	1.0	1.6	0.5
Term	3.1	1.9	0.1	0.1
Unexplained	26.4	30.5	36.3	39.3
Adjusted R^2	73.6	69.5	63.7	60.7
Observations	109,770	109,773	109,773	109,773

Note: Table 4 shows the Shapley–Owen variance decomposition (Israeli, 2007; Shorrocks, 2013) for the four AI outcomes. The five components are instructor, local course, university, major, and term fixed effects. Each entry is the share of total outcome variance, in percentage points. Adjusted R^2 is the total explained share.

I find similar patterns on the integration margins. Instructors explain around one-fourth of the variation across the integration measures, with local course fixed effects accounting for a comparable share.[19] By contrast, the university-by-major-by-level-by-term component explains a very small share of the variation in AI integration.[20] This indicates that professors and the courses they teach drive most of the systematic variation in whether AI is integrated into higher education.[21]

Table 4 presents the estimates from the full Shapley–Owen decomposition. Altogether, instructors and local courses remain the primary drivers of AI policy and actual integration. At the same time, universities account for 7 percent of the variation in AI policy, but a negligible amount of the variation in whether it is actually integrated into teaching or evaluation. Universities thus shape whether AI is formally regulated in the syllabus but do not contribute to whether it is actually incorporated into teaching. This pattern is consistent with Biasi and Ma (2022), who similarly find that instructors and courses are the primary drivers of the integration of frontier research into teaching.

6 Who Integrates AI into the Classroom?

6.1 Instructor Characteristics

The results presented in Section 5 show that instructors play a key role in shaping how AI is incorporated into the classroom. Table 5 presents descriptive evidence showing which types of instructors are more likely to integrate AI into their courses. First, the prevalence of explicit policies mirrors the results on AI mentions presented in Section 2. That is, women

[19] When including course-year fixed effects, instructors explain the largest share of the variance for all four main outcomes.

[20] Figure A7 shows that the relative magnitudes of these components are similar across the five broad major groups and university selectivity tiers.

[21] Table A9 shows that part of the course component reflects broader course characteristics that extend across institutions, rather than university-specific implementation.

are more likely to include explicit policies in their syllabi than men, and there are small differences in policy adoption across academic rank and publication activity.

At the same time, there are important differences in which instructors actually integrate AI into the classroom. On this margin, the gender difference largely disappears. Moreover, actual integration is far more common among assistant professors, as 13 percent of them integrate AI into their teaching, compared with around 10 percent of faculty members in other positions. Similarly, professors who have published papers are more likely to have AI in their teaching than their peers with no publications. I lastly note that professors with more publications are also more likely to incorporate AI into their classes. These differences arise mainly through the teaching margin rather than through the inclusion of AI-focused assessments.[22] In that sense, the instructors who are quickest to formalize AI in the syllabus are not the same as those most likely to incorporate it into their courses.

Table 5: Instructor Characteristics and AI Integration

	AI Policy (1)	**Teaching Integration** (2)
Gender		
Female	69.9	9.5
Male	60.8	10.4
Academic Rank		
Professor	61.8	10.1
Associate Professor	66.4	10.5
Assistant Professor	69.2	13.1
Non-Tenure-Track	65.4	10.0
Research Activity		
No Publications	64.1	9.1
Any Publications	65.2	11.4
Top Publications Quartile	57.8	12.5
Any ML Publications	61.8	15.5
Instructors	65,750	

Note: Table 5 shows the prevalence of the main AI outcomes by instructor characteristics. Each cell gives the share of instructors for whom at least one syllabus exhibits the column outcome, in percent. Unit of observation is the instructor. Top Publications Quartile marks instructors in the top quartile of publication counts, among those with any publications. Any ML Publications marks instructors with at least one machine-learning-related publication.

I also measure each instructor's proximity to machine learning research using the titles and abstracts of their publications. I represent each paper with SPECTER2, a language model trained on scientific text that places papers with similar content near one another (Singh et al., 2023), and score its proximity to a benchmark corpus of machine learning research. Specifically, I rank each paper among all papers published in the same subfield and year and average these ranks at the instructor level.[23] The measure thus captures the

[22] Table A10 shows that similar patterns appear across fields and across the selectivity spectrum.

[23] Figure A8 shows the most distinctive terms of machine learning research by field. This covers, for instance, spatial, statistical, and Monte Carlo approaches in economics and terms such as stochastic and gradient in computer science.

machine learning content of an instructor's research relative to researchers in their own field, though I also construct a version that ranks papers across all fields.[24] In line with this, Table 5 also shows that instructors with at least one machine learning publication have the highest rate of AI integration into their courses.

6.2 Main Analysis

Empirical Framework. To assess the extent to which instructor characteristics shape AI outcomes conditional on the courses they teach and the institutional context in which they operate, I estimate the following regression:

$$Y_n^{(k)} = X'_{i(n)}\beta_i^{(k)} + \phi_{c(n)} + \lambda_{u(n),m(n),l(n),t(n)} + \varepsilon_n^{(k)}, \qquad k \in \mathcal{K}, \tag{3}$$

where $X_{i(n)}$ includes instructor gender, academic rank relative to full professors, an indicator for any publications, the instructor's citation count, and the ML Research measure described above. The fixed effects $\phi_{c(n)}$ and $\lambda_{u(n),m(n),l(n),t(n)}$ absorb local-course and university-by-major-by-course-level-by-term variation, respectively, so that the estimates compare instructors teaching the same course in the same context. I estimate a version of Equation (3) that includes one characteristic at a time and a version that includes all of the main variables at the same time. In both cases, I control directly for the text-based characteristics described in Section 2, which accounts for differences in how instructors structure their syllabi.[25]

Empirical Results. Figure 9 presents the estimates for teaching integration. I find that assistant professors are more likely than full professors to integrate AI into their teaching, even conditional on the courses they teach, though this difference is less precisely estimated. Instructors with published papers are also more likely to have incorporated AI into their courses. These results thus suggest that young tenure-track faculty with active research agendas are leading the integration of AI in the classroom, fitting in with evidence that younger and more educated workers tend to lead the adoption of artificial intelligence in the workplace (McElheran et al., 2024; Bick et al., 2025).[26]

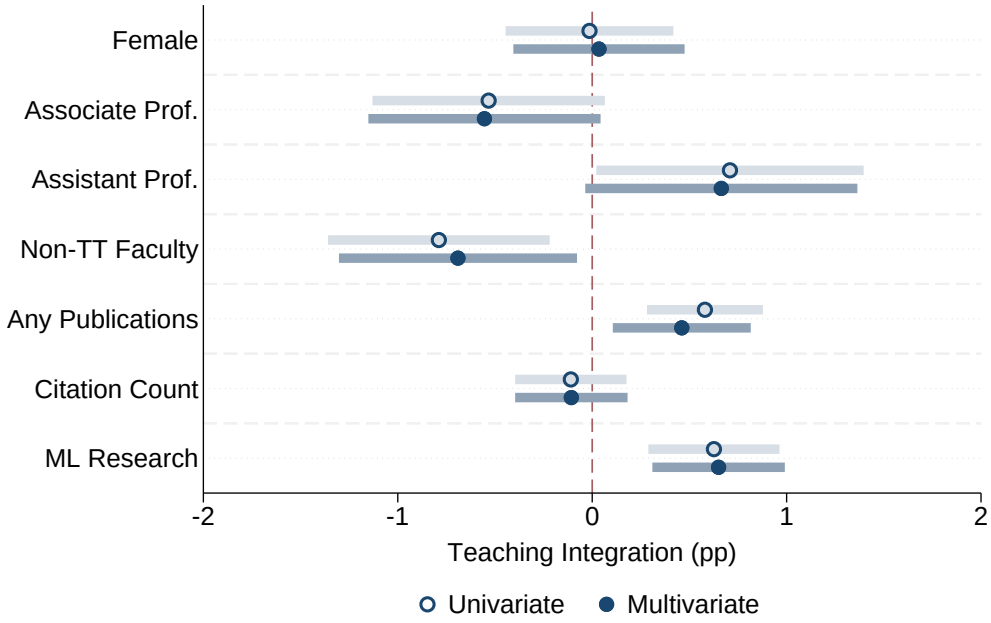
Importantly, the nature of an instructor's own research plays an important role in driving the integration process. That is, I find that professors whose published work is closer to machine learning are far more likely to bring AI into their teaching. Since the specification includes course fixed effects, this difference does not reflect machine-learning researchers teaching courses about AI. Instead, this emerges among instructors teaching the

[24] Table A11 presents both versions by field. Instructors in computer science rank closest to machine learning research overall, while the within-field version compares each instructor to their own peers.

[25] While these text-based characteristics are themselves outcomes of the syllabus-writing process, I include them to assess whether the instructor coefficients reflect genuine differences in AI adoption rather than differences in writing style.

[26] Assistant professors may be more likely to integrate AI because they are more often teaching a course for the first time, which requires designing new materials rather than updating existing ones. In the UT Austin and UT Dallas sample, where I observe whether an instructor has previously taught each course, the estimates are unchanged upon controlling for first-time teaching (Figure A9).

Figure 9: Instructor Characteristics and Teaching Integration



Note: Figure 9 shows coefficient estimates from regressions of teaching integration, measured in percentage points, on instructor characteristics. Hollow circles show univariate estimates, entering one characteristic at a time with all rank groups entering together. Solid circles show multivariate estimates entering all characteristics jointly. Academic rank uses the four-group ladder, with full professors as the omitted group. ML Research measures how close an instructor's published work is to machine-learning research, relative to other work in the instructor's own field. All specifications include syllabus text controls and absorb course and university-by-major-by-level-by-term fixed effects. Standard errors are clustered at the university level. Bars represent 90% confidence intervals.

same course, indicating that professors whose research brings them close to this technology are more likely to incorporate it into what their students are taught.

On the other hand, professors with higher citation counts are no more likely to integrate AI into their teaching. Moreover, Figure 9 shows that these estimates are similar when I include one characteristic at a time, so these patterns do not reflect overlap across the main covariates. In that sense, what matters for AI integration is what an instructor researches rather than how widely that work is cited.

I also consider the specific ways in which instructors integrate AI into their teaching, drawing on the qualitative classification described earlier. Table A12 shows that the results on integration are concentrated in course content about AI methods themselves. That is, instructors whose research is closer to machine learning are particularly likely to teach students how these tools work, rather than simply referencing AI elsewhere in their course materials.

Table 6 presents the corresponding estimates for the remaining outcomes. On the evaluation margin, instructors whose research is closer to machine learning are also more likely to incorporate AI into graded coursework, a difference amounting to roughly a sixth of the sample mean.[27] At the same time, assistant professors are no more likely than full profes-

[27]The integration results are also robust to excluding the text-based syllabus controls (Figure A10).

sors to have done so. Table A13 examines the specific types of AI-related assessments, and I find that assistant professors stand out in assignments in which students build applications with AI.

On the policy margin, I find that in contrast with the integration margins, neither instructors with publications nor those researching machine-learning-related topics are more likely to have an explicit policy in place. While assistant professors are more likely than full professors to include a policy in their syllabi, this difference is modest relative to the widespread adoption of AI policies. Lastly, I find similar patterns on whether the course explicitly allows students to use AI.

The main results remain consistent across various robustness checks. First, I re-estimate Equation (3) and additionally control for how many papers each instructor has published, the novelty of their research agenda, and whether they earned their PhD at a highly selective institution. Table A14 shows that none of these measures shapes AI integration, while the main estimates remain unchanged. In this context, Biasi and Ma (2022) show that instructors who publish more and are more cited teach content closer to the research frontier in their own fields. My results sharpen this pattern for the arrival of a new technology: what matters for integrating AI is not how much an instructor publishes but whether their own work has brought them close to the technology itself. The results are also robust to alternative definitions of machine-learning research. For instance, I find similar estimates when measuring research on machine-learning-related topics through the vocabulary of an instructor's publications, through the version that ranks papers across all fields, or through an indicator for having published any AI-related work (Table A15).

Taken together, these results show that AI is entering the classroom through the instructors whose own research places them closest to the technology. The same characteristics play no role in whether instructors adopt a policy, a margin that requires no engagement with the technology itself. The arrival of generative AI thus reveals the diffusion of a new technology in real time, as the professors already equipped to teach with it are the first to bring it into classrooms where it barely existed three years ago.

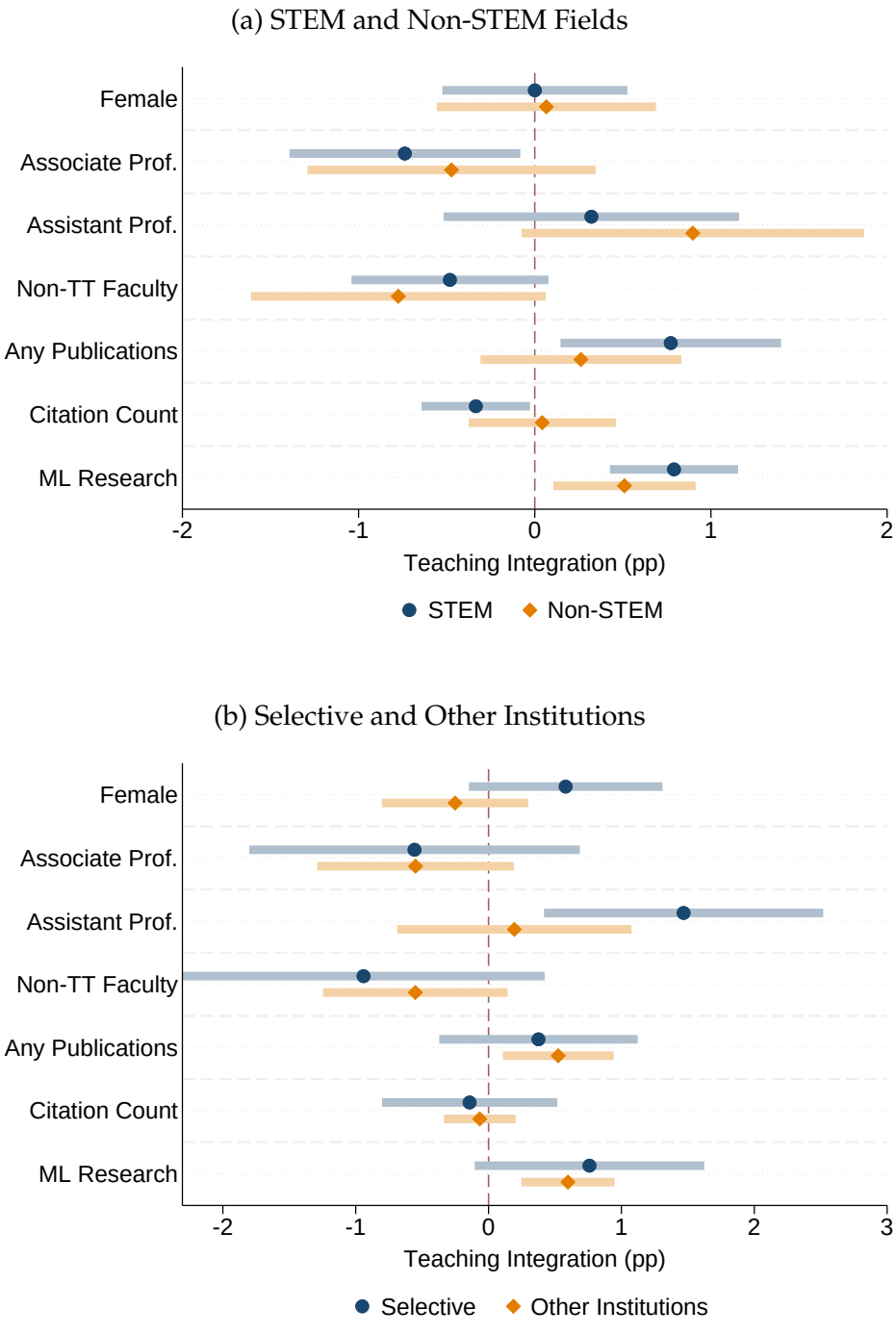
6.3 Heterogeneity Across Universities and Majors

The results presented above show that assistant professors with active research agendas are the instructors most likely to integrate AI into the classroom, yet the extent of AI integration varies across disciplinary and institutional contexts. I thus split the sample into STEM and non-STEM fields and re-estimate Equation (3) for teaching integration within each group. Figure 10 shows that the main results hold in both groups. In particular, assistant professors and instructors with published papers remain more likely to integrate AI into their teaching in STEM and non-STEM fields alike, though these estimates are less precise within the smaller samples.[28]

Interestingly, machine learning research also shapes teaching integration within non-

[28]Table A16 presents these results across five broad fields, and Table A17 presents the corresponding estimates for evaluation integration.

Figure 10: Heterogeneity in Teaching Integration



Note: Figure 10 shows multivariate coefficient estimates from regressions of teaching integration, measured in percentage points, on instructor characteristics. Panel (a) estimates the specification separately within STEM and non-STEM fields. Panel (b) estimates it separately within the nine Selective institutions and all other institutions, where tiers reflect the mean SAT score of the 2019 entering class. All specifications include syllabus text controls and absorb course and university-by-major-by-level-by-term fixed effects. Standard errors are clustered at the university level. Bars represent 90% confidence intervals.

Table 6: Instructor Characteristics: Evaluation Integration, AI Policy, and AI Permitted

	Evaluation Integration (1)	AI Policy (2)	AI Permitted (3)
Female	0.12	2.63***	0.90*
	(0.15)	(0.60)	(0.53)
Associate Prof.	-0.52***	0.10	-0.21
	(0.19)	(0.95)	(0.72)
Assistant Prof.	-0.29	1.94**	1.37**
	(0.22)	(0.88)	(0.67)
Non-TT Faculty	-0.59***	1.10*	0.48
	(0.19)	(0.58)	(0.43)
Any Publications	0.18	0.55	0.60
	(0.15)	(0.41)	(0.46)
Citation Count	0.02	-0.00	-0.59
	(0.12)	(0.34)	(0.36)
ML Research	0.40***	-0.18	0.43
	(0.10)	(0.24)	(0.31)
Text Controls	Yes	Yes	Yes
Course FE	Yes	Yes	Yes
Univ.-Major-Level-Term FE	Yes	Yes	Yes
Outcome mean (%)	2.56	57.42	23.67
Observations	134,766	134,761	134,766
Instructors	50,719	50,717	50,719
Institutions	50	50	50

Note: Table 6 presents coefficients from separate regressions of each column outcome, measured in percentage points, on instructor characteristics. Academic rank uses the harmonized four-group ladder measure, with full professors as the omitted group. Indicators for missing gender and rank are included. ML Research measures how close an instructor's published work is to machine-learning research, relative to other work in the instructor's own field. Citation Count is the log of an instructor's citations. All specifications include syllabus text controls and absorb course and university-by-major-by-level-by-term fixed effects. Standard errors, clustered at the university level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

STEM fields. In these courses, AI is rarely the subject being taught, so this pattern does not reflect specialists covering their own specialty. Rather, instructors whose research has brought them close to machine learning are bringing the technology into courses across the curriculum.

Since AI has been differentially integrated into the classroom at selective institutions compared with the rest of the sample, I further examine whether these results differ across selectivity levels. Assistant professors at selective universities are significantly more likely than full professors to have integrated AI into their teaching, while this career-stage difference is largely absent elsewhere. Research activity, on the other hand, matters across the selectivity spectrum, as published instructors and those with machine-learning-related research remain more likely to integrate AI outside the most selective tier (Table A18). As such, selective universities are not just more likely to integrate AI into the classroom, as it is early-career instructors at these institutions who are experimenting most actively with artificial intelligence.

Altogether, these results show that young, research-active faculty members lead the in-

tegration of AI in the classroom across fields and across the selectivity spectrum, with the career-stage differences concentrated at selective universities. These results thus align with growing evidence showing that AI adoption is more prevalent in firms with better organizational resources and more educated workers (McElheran et al., 2024; Bick et al., 2026). Early AI integration in higher education follows the same broad patterns as in the rest of the economy.

7 Conclusion

In this paper, I have examined how AI is being integrated in higher education, using a dataset of 185,000 course syllabi across 50 public universities linked to detailed instructor, course, and institutional characteristics. I have shown that explicit AI policy language has diffused widely, appearing in close to two-thirds of Spring 2026 syllabi. Yet this diffusion has not translated into comparable changes to what instructors teach or how they assess students, and only about six percent of syllabi show teaching integration.

To understand what drives these patterns, I have decomposed the variation in AI outcomes across instructors, courses, and institutions, and examined which observable instructor characteristics predict classroom integration. I find that instructors and the specific courses they teach account for the large majority of explained variation, while the university component is negligible. Institutions shape the conditions for formal policy diffusion, but they do not determine whether instructors actually redesign their courses around AI. Moreover, assistant professors and research-active instructors are the most likely to adopt AI in the classroom, and they do so primarily on the teaching margin rather than through evaluation. What matters is not the volume of an instructor's research but rather proximity to active engagement with new methods. Altogether, young, research-active faculty at the most research-intensive universities are leading the diffusion of AI into the classroom.

In this context, future work could link the syllabus-based measures developed here to student-level administrative records to estimate the returns to learning from a professor who was an early adopter of AI in the classroom (Biasi and Ma, 2026). Whether these patterns persist as the technology matures, or whether they spread more broadly across institutions and fields, will determine how AI reshapes higher education in the years ahead.

References

Abel, J. R., Deitz, R., and Su, Y. (2014). Are recent college graduates finding good jobs? *Current Issues in Economics and Finance*, 20(1):1–8.

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2022). AI and jobs: Evidence from online vacancies. *Journal of Labor Economics*, 40(S1):S293–S340.

Acemoglu, D. and Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30.

Ash, E., Calderon, S., Chatzivasileiadis, M., and Hansen, S. (2025). Using LLMs to analyze interview transcripts at scale. Working paper, ETH Zürich / University College London.

Ash, E. and Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, 15:659–688.

Audoly, R., McGee, R., Ocampo, S., and Paz-Pardo, G. (2025). A practitioner's note on the Shapley-Owen-Shorrocks decomposition. Staff Reports 1163, Federal Reserve Bank of New York.

Babina, T., Fedyk, A., He, A., and Hodson, J. (2024). Firm investments in artificial intelligence technologies and changes in workforce composition. Working Paper 31325, NBER.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131(4):1593–1636.

Barrie, C., Palmer, A., and Spirling, A. (2024). Replication for language models: Problems, principles, and best practice for political science. Working Paper.

Bartik, A. W., Gupta, A., and Milo, D. (2024). The costs of housing regulation: Evidence from generative regulatory measurement. Working paper.

Biasi, B. and Ma, S. (2022). The education-innovation gap. Working Paper 29853, NBER.

Biasi, B. and Ma, S. (2026). Frontier knowledge in college and student success. Working paper, Yale SOM / NBER affiliation version. February 26, 2026 version.

Bick, A., Blandin, A., and Deming, D. J. (2025). The rapid adoption of generative AI. Working Paper 32337, NBER.

Bick, A., Blandin, A., Deming, D. J., Fuchs-Schündeln, N., and Jessen, J. (2026). Mind the gap: AI adoption in Europe and the U.S. Working Paper 34995, National Bureau of Economic Research.

Bonney, K., Breaux, C., Buffington, C., Dinlersoz, E., Foster, L., Goldschlag, N., Haltiwanger, J., Kroff, Z., and Savage, K. (2024). Tracking firm use of AI in real time: A snapshot from the business trends and outlook survey. U.S. Census Bureau working paper / report.

Braga, M., Paccagnella, M., and Pellizzari, M. (2016). The impact of college teaching on students' academic and labor market outcomes. *Journal of Labor Economics*, 34(3):781–822.

Braghieri, L., Eichmeyer, S., Levy, R., Mobius, M., Steinhardt, J., and Zhong, R. (2024). Article-level slant and polarization of news consumption on social media. CEPR Discussion Paper 19807.

Bresnahan, T. F. and Trajtenberg, M. (1995). General purpose technologies: "engines of growth"? *Journal of Econometrics*, 65(1):83–108.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2025). Generative AI at work. *Quarterly Journal of Economics*, 140(2):889–942.

Bursztyn, L., Handel, B. R., Jiménez-Durán, R., and Roth, C. (2025). When product markets become collective traps: The case of social media. *American Economic Review*, 115(12):4105–4136.

Calderon, N., Reichart, R., and Dror, R. (2025). The alternative annotator test for LLM-as-a-judge: How to statistically justify replacing human annotators with LLMs. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 16051–16081, Vienna, Austria.

Carrell, S. E. and West, J. E. (2010). Does professor quality matter? evidence from random assignment of students to professors. *Journal of Political Economy*, 118(3):409–432.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., and Yagan, D. (2020). Income segregation and intergenerational mobility across colleges in the united states. *Quarterly Journal of Economics*, 135(3):1567–1633. Previously circulated as "Mobility Report Cards" (NBER Working Paper No. 23618).

Chiang, W.-L., Zheng, L., Sheng, Y., Angelopoulos, A. N., Li, T., Li, D., Zhang, H., Zhu, B., Jordan, M., Gonzalez, J. E., and Stoica, I. (2024). Chatbot arena: An open platform for evaluating LLMs by human preference. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*.

Chirikov, I. (2026). How instructors regulate AI in college: Evidence from 31,000 course syllabi. Higher Education Working Paper 26-1, UC Berkeley Center for Studies in Higher Education.

Chyn, E., Haggag, K., and Maruthiah, C. (2025). Ideology in government: Evidence from the office of Indian Affairs and the assimilation era. Working Paper 34415, National Bureau of Economic Research.

Clayton, C., Coppola, A., Maggiori, M., and Schreger, J. (2025). Geoeconomic pressure. Working Paper 34020, National Bureau of Economic Research.

Comin, D. and Hobijn, B. (2010). An exploration of technology diffusion. *American Economic Review*, 100(5):2031–2059.

Contractor, Z. and Reyes, G. (2025). Generative AI in higher education: Evidence from an elite college. Working paper, Middlebury College.

De Vlieger, P., Jacob, B., and Stange, K. (2016). Measuring instructor effectiveness in higher education. Working Paper 22998, NBER.

Dell, M. (2025). Deep learning for economists. *Journal of Economic Literature*, 63(1):5–58.

Dillon, E., Guo, N., and Saurer, R. (2025). Shifting work patterns in the age of generative AI. Working Paper 33845, NBER.

Dolnicar, S., Grün, B., and Leisch, F. (2011). Quick, simple and reliable: Forced binary survey questions. *International Journal of Market Research*, 53(2):231–252.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2024). GPTs are GPTs: Labor market impact potential of large language models. *Science*, 384(6702):1306–1308.

Feld, J., Salamanca, N., and Zölitz, U. (2020). Are professors worth it? the value-added and costs of tutorial instructors. *Journal of Human Resources*, 55(3):836–863.

Felten, E. W., Raj, M., and Seamans, R. (2023). How will language modelers like ChatGPT affect occupations and industries? arXiv preprint arXiv:2303.01157.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–574.

Haaland, I., Roth, C., Stantcheva, S., and Wohlfart, J. (2025). Understanding economic behavior using open-ended survey data. *Journal of Economic Literature*, 63(4):1244–1280.

Handa, K., Bent, D., Tamkin, A., McCain, M., Durmus, E., Stern, M., Schiraldi, M., Huang, S., Ritchie, S., Syverud, S., Jagadish, K., Vo, M., Bell, M., and Ganguli, D. (2025). How university students use Claude. Technical report, Anthropic.

Hartley, J., Jolevski, F., Melo, V., and Moore, B. (2025). The labor market effects of generative artificial intelligence. Working Paper 5136877, SSRN.

Hoffmann, F. and Oreopoulos, P. (2009). Professor qualities and student achievement. *Review of Economics and Statistics*, 91(1):83–92.

Humlum, A. and Vestergaard, E. (2025). Large language models, small labor market effects. Working Paper 32424, NBER.

Israeli, O. (2007). A Shapley-based decomposition of the R-square of a linear regression. *Journal of Economic Inequality*, 5(2):199–212.

Javadian Sabet, A., Bana, S. H., Yu, R., and Frank, M. R. (2024). Course-skill atlas: A national longitudinal dataset of skills taught in u.s. higher education curricula. *Scientific Data*, 11:1410.

Korinek, A. (2023). Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature*, 61(4):1281–1317.

Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology*. Sage Publications, Thousand Oaks, CA, 2nd edition.

Lagakos, D., Michalopoulos, S., and Voth, H.-J. (2025). American life histories. Working Paper 33373, National Bureau of Economic Research.

Link, S., Peichl, A., Roth, C., and Wohlfart, J. (2024). Attention to the macroeconomy. Working Paper 32988, NBER.

Lombard, M., Snyder-Duch, J., and Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28(4):587–604.

Ludwig, J., Mullainathan, S., and Rambachan, A. (2025). Large language models: An applied econometric framework. Working Paper 33344, National Bureau of Economic Research.

McDonald, N., Johri, A., Ali, A., and Collier, A. H. (2025). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans*, 3:100121.

McElheran, K., Li, J. F., Brynjolfsson, E., Kroff, Z., Dinlersoz, E., Foster, L., and Zolas, N. (2024). AI adoption in america: Who, what, and where. *Journal of Economics & Management Strategy*, 33(2). NBER Working Paper 31788.

O'Connor, C. and Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19:1–13.

Priem, J., Piwowar, H., and Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833*.

Saltiel, F. (2026). Imagined futures and early retirement outcomes. Working Paper, McGill University.

Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: A unified framework based on the Shapley value. *Journal of Economic Inequality*, 11(1):99–126.

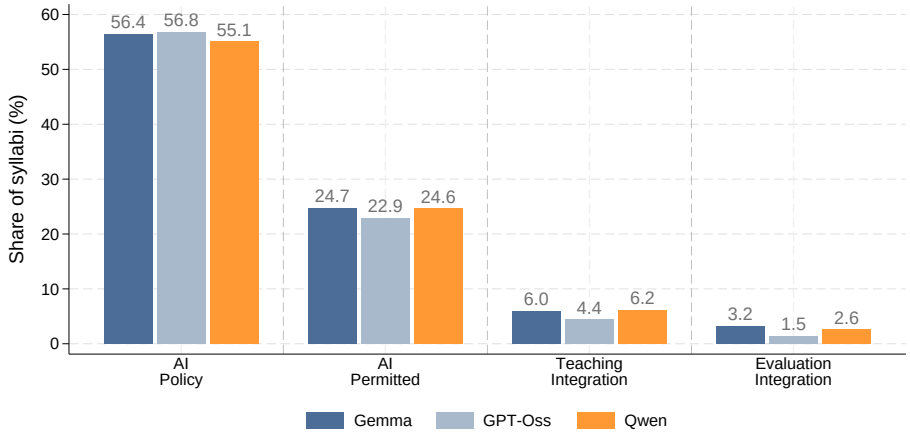
Singh, A., D'Arcy, M., Cohan, A., Downey, D., and Feldman, S. (2023). SciRepEval: A multi-format benchmark for scientific document representations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*.

Yotzov, I., Barrero, J. M., Bloom, N., and Davis, S. J. (2025). Firm data on AI. Working Paper 34836, NBER.

- SUPPLEMENTARY APPENDICES -
For Online Publication

Appendix A: Additional Figures and Tables

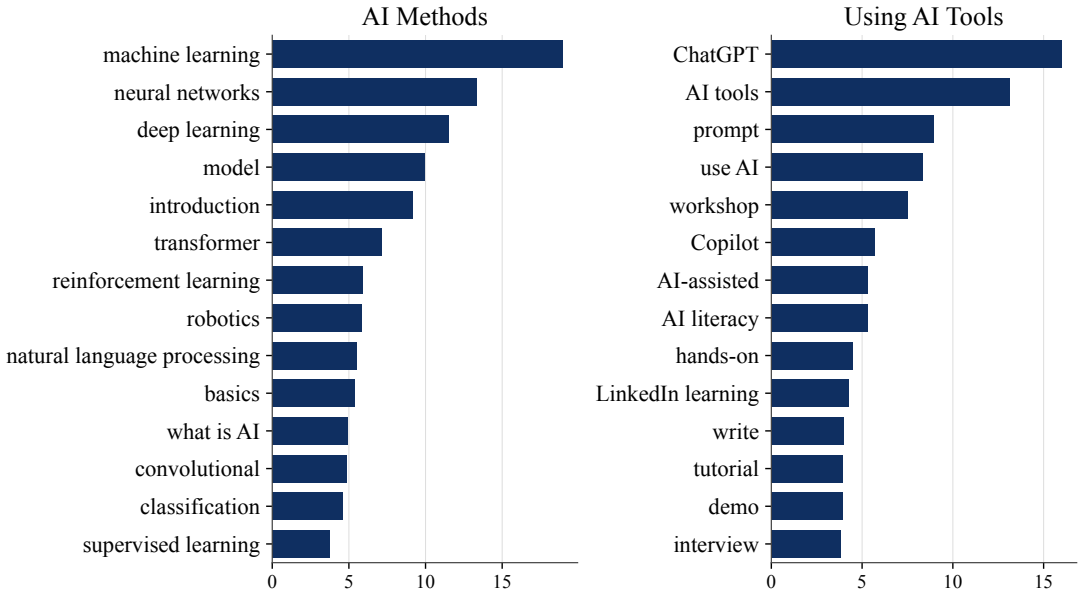
Figure A1: Prevalence Across the Three Classification Models



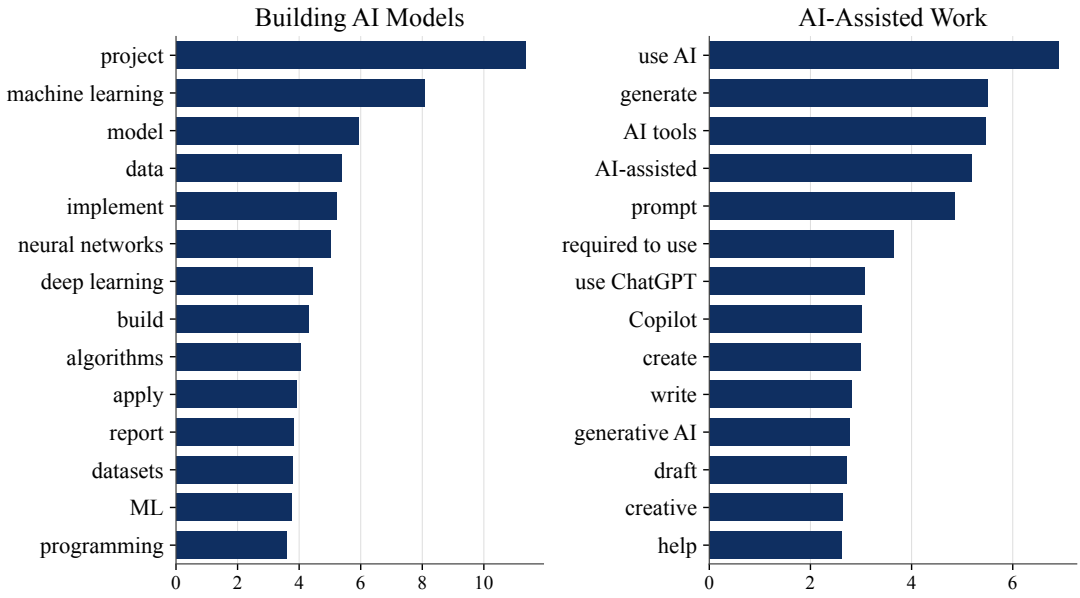
Note: Figure A1 shows the pooled prevalence of each AI integration measure under the three classification models, Gemma, GPT-Oss, and Qwen, for the 50 universities across Fall 2024 through Spring 2026. Each bar covers the syllabi rated by that model, with syllabi that never mention AI coded zero. The sample covers 185,154 syllabi.

Figure A2: Distinctive Language of the Main Integration Categories

(a) Teaching integration

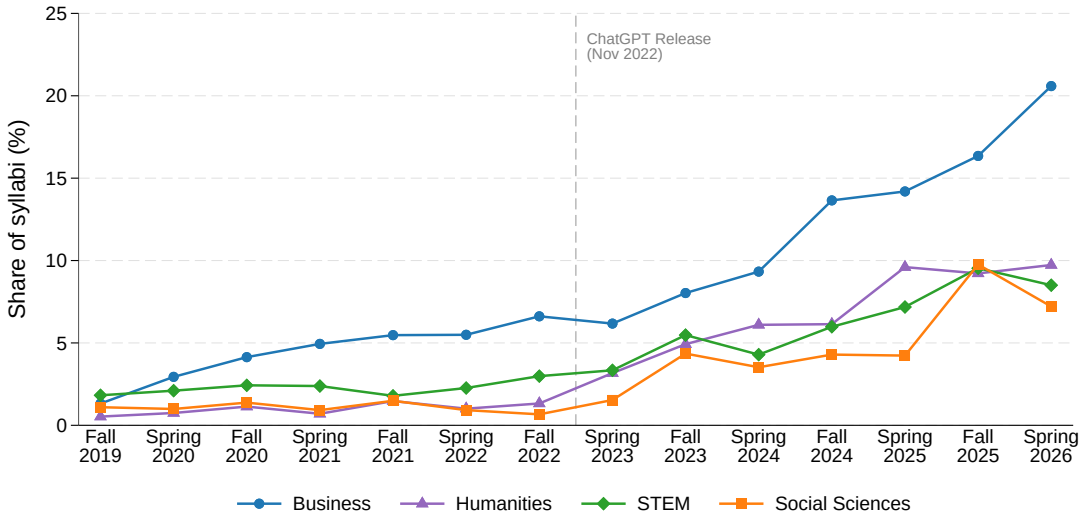


(b) Evaluation integration



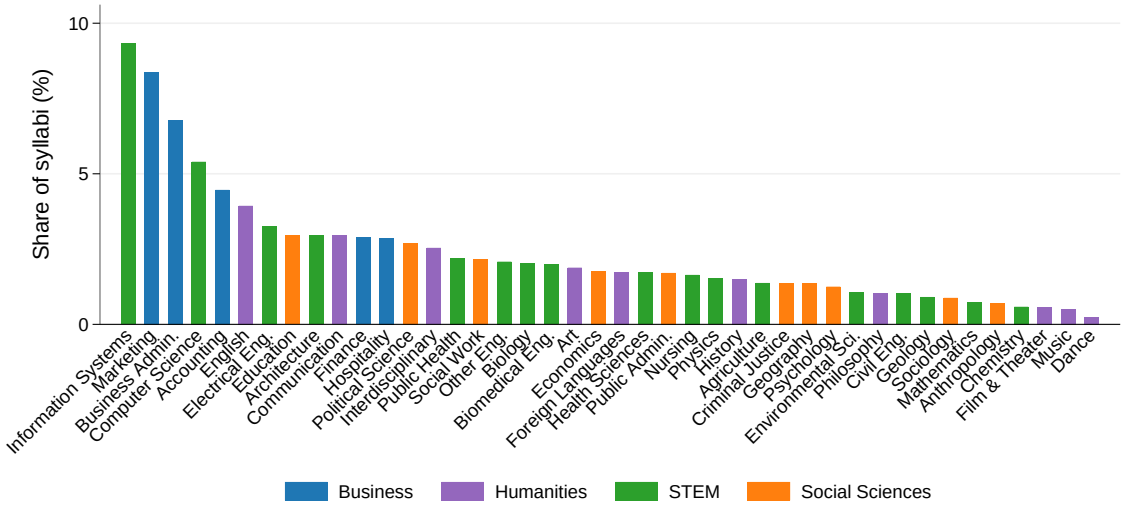
Note: Figure A2 shows the terms most distinctive of two main teaching categories (panel a) and two main evaluation categories (panel b). Bars show weighted log-odds scores comparing the language of each category's evidence passages against all other passages on the same margin.

Figure A3: Teaching Integration by Broad Field, Longitudinal



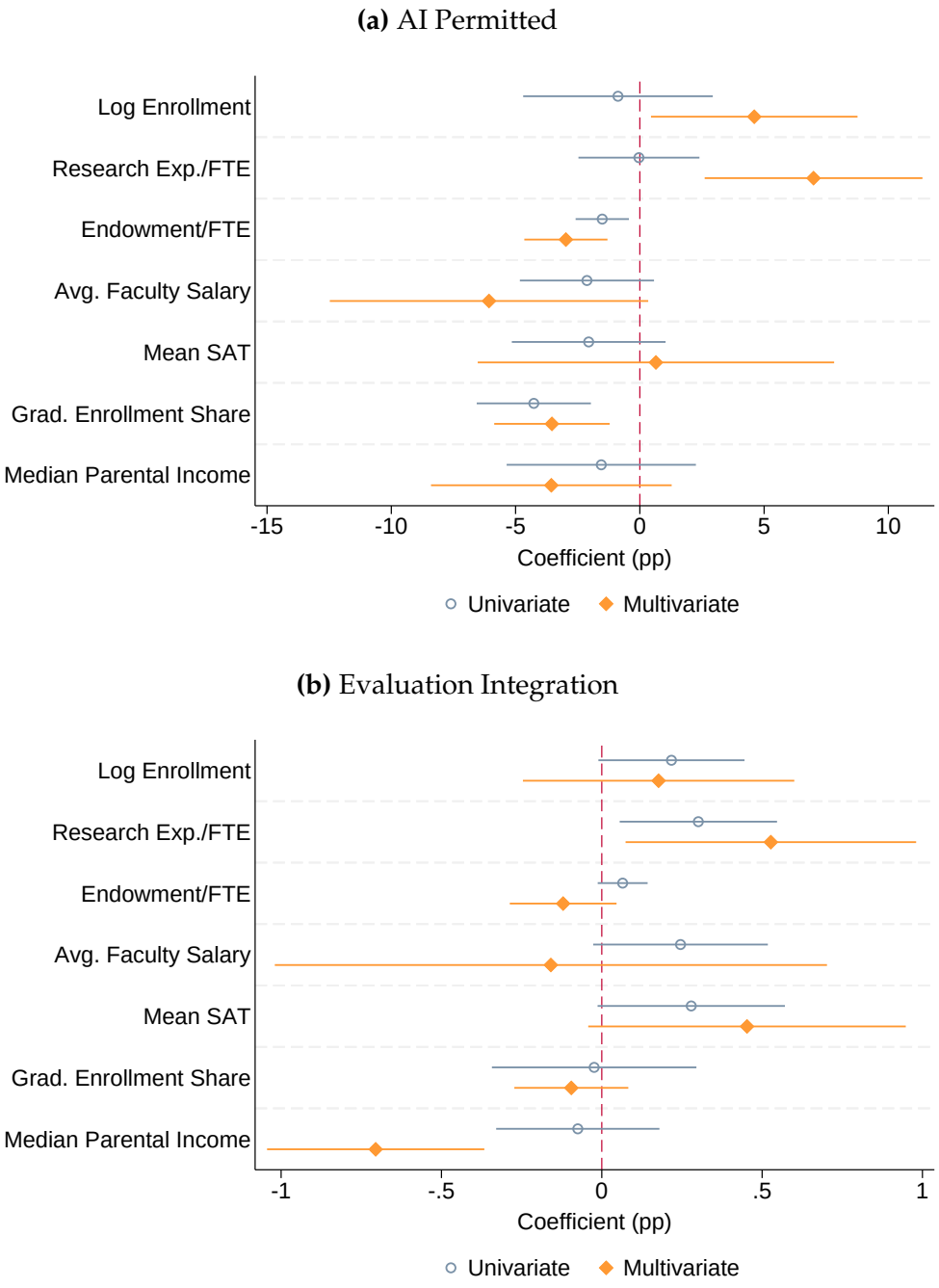
Note: Figure A3 presents the evolution of teaching integration across four broad field groups (Business, Humanities, STEM, and Social Sciences), excluding Other. Each line connects term-level means for courses in the balanced panel. The vertical dashed line marks the launch of ChatGPT in November 2022. The longitudinal sample covers UT Austin and UT Dallas from Fall 2019 through Spring 2026.

Figure A4: Evaluation Integration by Major



Note: Figure A4 presents the share of syllabi with AI evaluation integration across 42 harmonized majors in Spring 2026. Bars are colored by broad field (Business, Humanities, STEM, Social Sciences, Other).

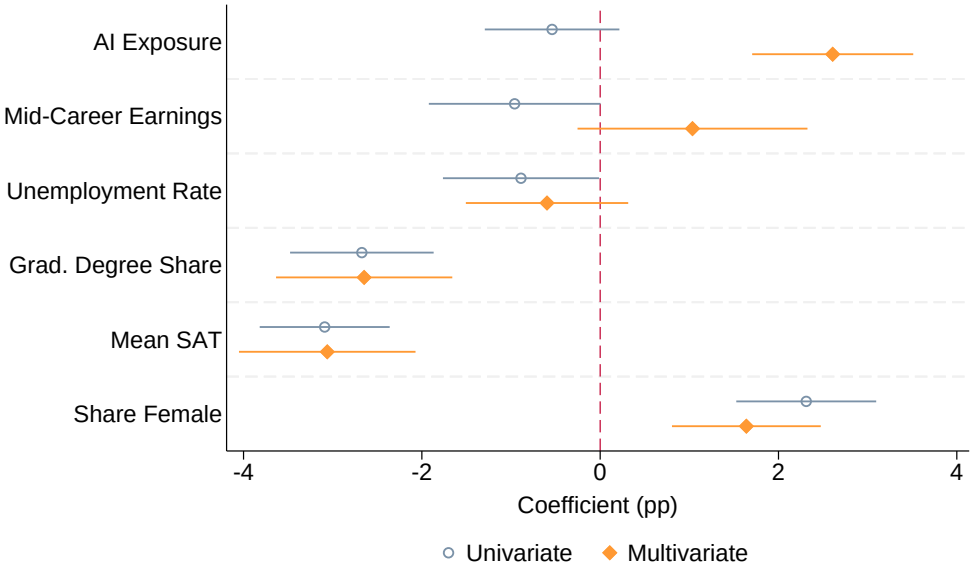
Figure A5: University Covariates: AI Permitted and Evaluation Integration



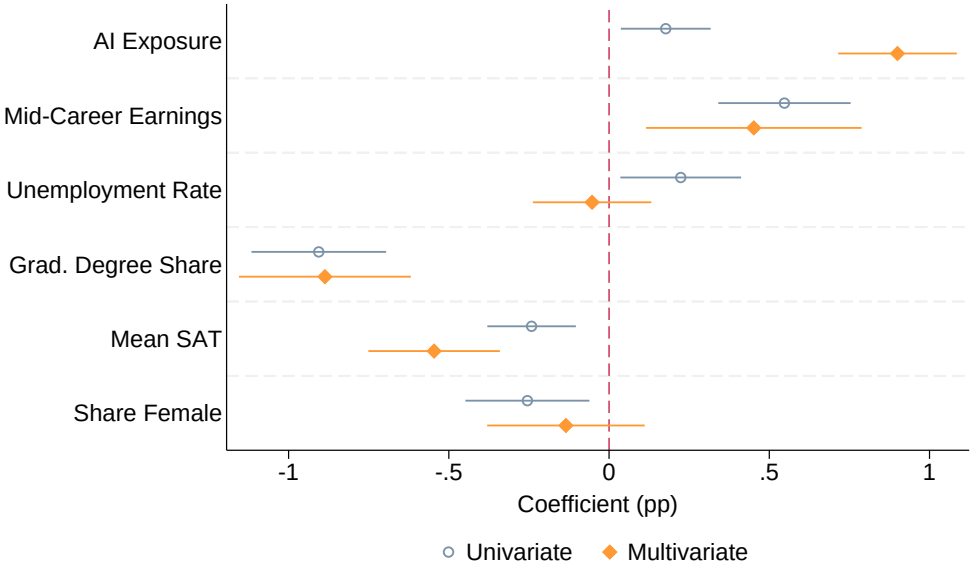
Note: Figure A5 presents coefficient estimates from regressions of AI being permitted (panel a) and AI evaluation integration (panel b) on seven university-level covariates. Circles show pairwise estimates; diamonds show multivariate estimates including all covariates simultaneously. Bars represent 90% confidence intervals. Fixed effects: major × course level × term. Standard errors clustered at the university level.

Figure A6: Major Covariates: AI Permitted and Evaluation Integration

(a) AI Permitted



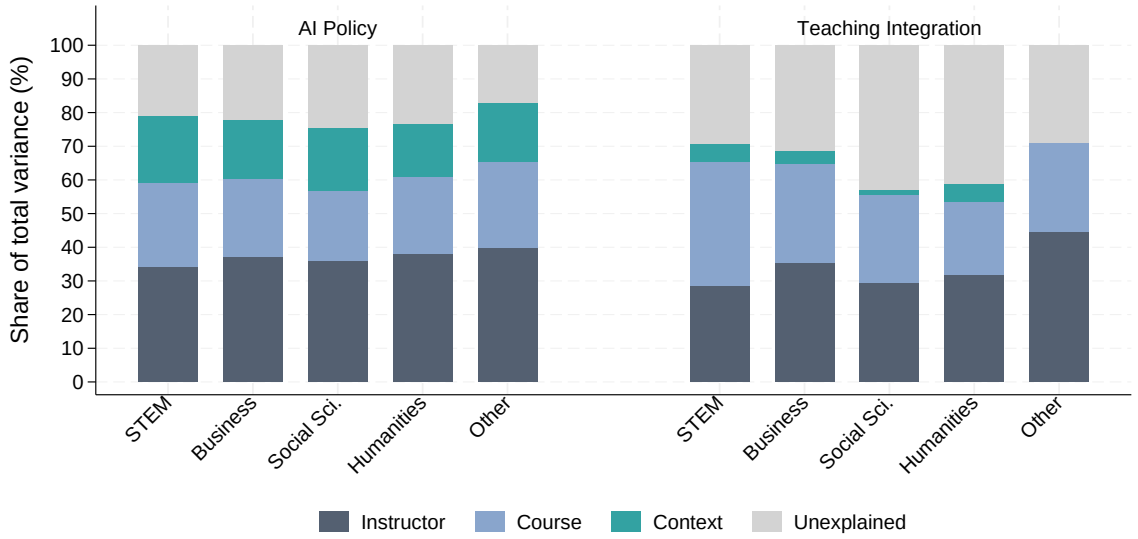
(b) Evaluation Integration



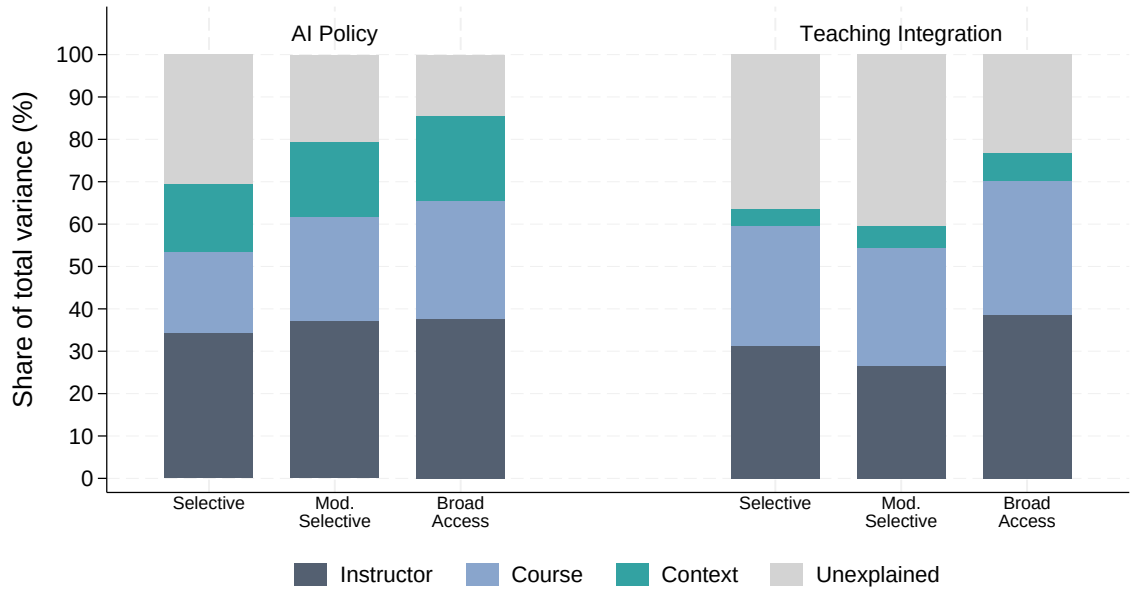
Note: Figure A6 presents coefficient estimates from regressions of AI being permitted (panel a) and AI evaluation integration (panel b) on six major-level covariates. Circles show pairwise estimates; diamonds show multivariate estimates including all covariates simultaneously. Bars represent 90% confidence intervals. Fixed effects: course level × term and university. Standard errors clustered at the university level.

Figure A7: Variance Decomposition by Broad Field and Selectivity

(a) By Broad Major Group

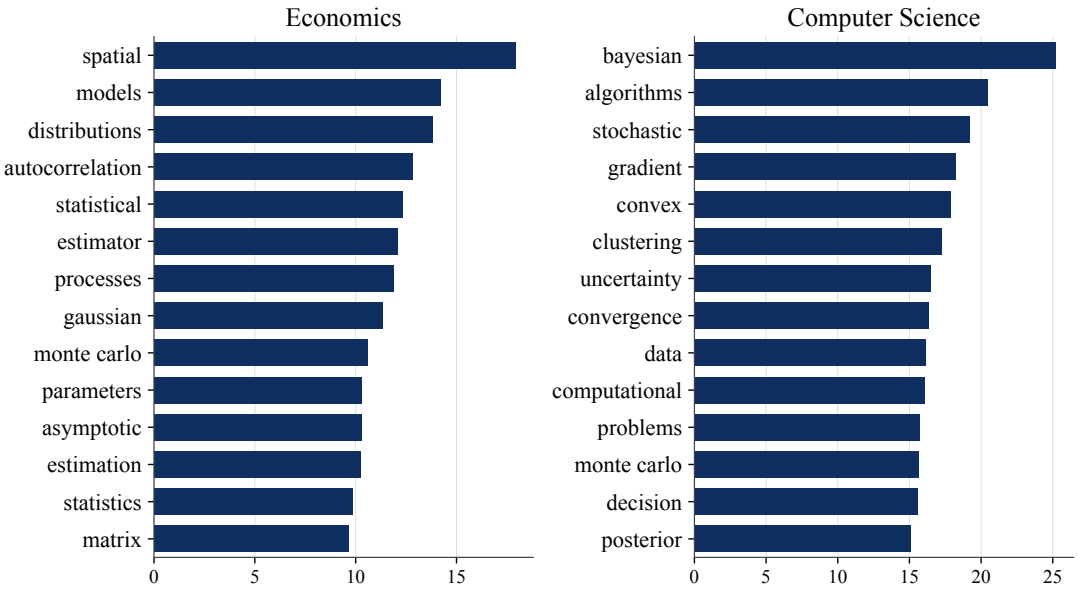


(b) By University Selectivity



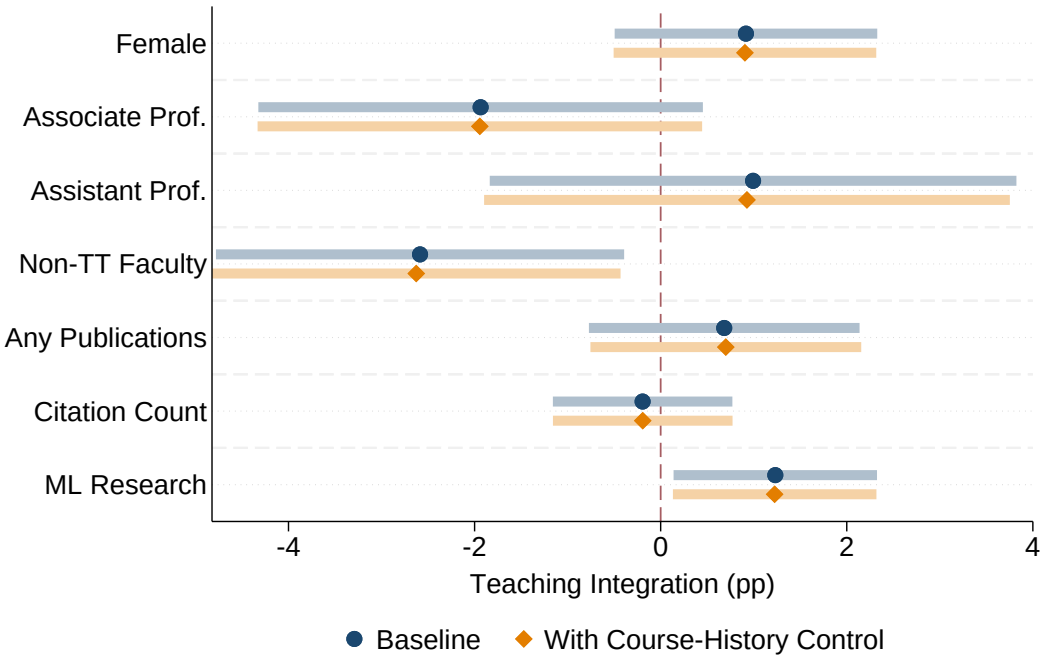
Note: Figure A7 shows the baseline three-way variance decomposition for AI Policy and Teaching Integration across subsamples. Panel (a) shows the decomposition separately across the five broad fields. Panel (b) shows the decomposition separately across the three selectivity tiers. Each bar shows the shares of total variance attributed to instructor fixed effects, local course fixed effects, the broader context component, and the unexplained residual.

Figure A8: Distinctive Language of the ML Research Measure Within Field



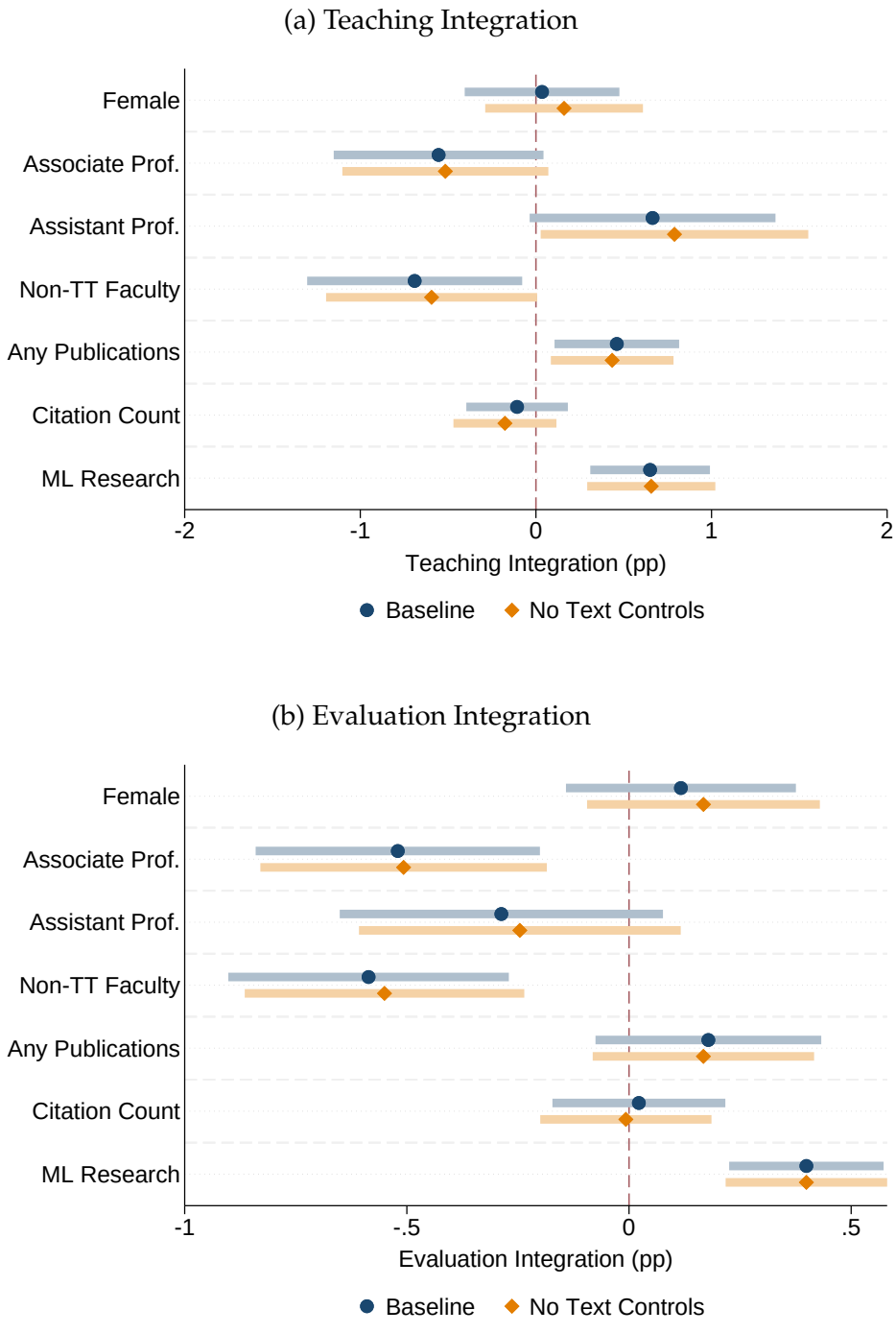
Note: Figure A8 shows the terms most distinctive of instructors in the top quartile of the within-field ML Research measure in Economics (left panel) and Computer Science (right panel). Bars show weighted log-odds scores comparing the language of these instructors' paper titles and abstracts against those of all other instructors in the same field.

Figure A9: Instructor Characteristics and Teaching Integration: UT Austin and UT Dallas



Note: Figure A9 shows multivariate coefficient estimates from regressions of teaching integration, measured in percentage points, on instructor characteristics, restricted to UT Austin and UT Dallas, the two universities whose teaching records reach back to 2019. The second series adds a course-history control equal to one when the instructor had not taught that course at the same university in any earlier term on record. Both series use the estimation sample of instructors whose earlier teaching record is observed. All specifications include syllabus text controls and absorb course and university-by-major-by-level-by-term fixed effects. Standard errors are clustered at the instructor level. Bars represent 90% confidence intervals.

Figure A10: Robustness to Text Controls



Note: Figure A10 shows multivariate coefficient estimates from regressions of the panel outcome, measured in percentage points, on instructor characteristics, with and without the nine syllabus text controls. All specifications absorb course and university-by-major-by-level-by-term fixed effects. Standard errors are clustered at the university level. Bars represent 90% confidence intervals.

Table A1: Number of Syllabi by University and Academic Term

University	Fall 2024	Spring 2025	Fall 2025	Spring 2026	Total
Texas					
UT Austin	3,231	3,446	3,443	3,273	13,393
UT San Antonio	2,164	2,197	2,203	2,146	8,710
UT Dallas	1,957	1,893	2,064	1,871	7,785
Houston	0	0	3,389	3,287	6,676
Texas A&M	0	0	1,169	5,012	6,181
Texas State	0	0	2,777	2,614	5,391
Texas Tech	1,382	1,174	1,342	1,081	4,979
Sam Houston State	0	0	449	4,035	4,484
Stephen F. Austin	1,316	512	518	1,844	4,190
UT Arlington	0	0	0	2,834	2,834
North Texas	0	0	0	2,780	2,780
West Texas A&M	599	597	615	601	2,412
Lamar	0	0	0	2,343	2,343
Houston Downtown	0	0	0	1,951	1,951
Prairie View A&M	0	0	0	1,872	1,872
Tarleton State	0	0	0	1,766	1,766
Texas A&M Commerce	0	0	0	1,717	1,717
UT El Paso	0	0	0	1,498	1,498
Houston Clear Lake	0	0	0	1,424	1,424
Texas A&M International	0	0	0	1,030	1,030
Texas A&M San Antonio	0	0	0	1,009	1,009
Texas A&M Corpus Christi	0	0	0	1,008	1,008
Texas A&M Kingsville	0	0	0	939	939
Texas Southern	0	0	0	882	882
UT Tyler	0	0	0	775	775
UT Permian Basin	0	0	0	751	751
North Texas at Dallas	0	0	0	335	335
Florida					
Central Florida	3,820	3,719	4,034	3,957	15,530
Florida International	1,340	1,446	3,803	4,072	10,661
Florida Atlantic	2,179	1,810	2,163	1,948	8,100
Florida	1,542	1,509	1,766	1,708	6,525
Florida State	0	0	0	4,147	4,147
Florida A&M	0	0	1,320	1,352	2,672
Florida Gulf Coast	0	0	0	2,188	2,188
North Florida	0	0	85	1,753	1,838
West Florida	0	0	0	1,304	1,304
Georgia					
Kennesaw State	0	0	2,306	4,178	6,484
Georgia	1,066	1,098	1,416	2,205	5,785
Georgia State	0	0	160	1,721	1,881
Georgia College	0	0	578	1,122	1,700
Georgia Southern	0	0	191	1,049	1,240
Albany State	0	0	0	941	941
Fort Valley State	0	0	0	462	462
Indiana					
IU Bloomington	0	0	2,725	2,705	5,430
Purdue	157	179	2,145	1,698	4,179
IU Indianapolis	0	0	1,297	1,261	2,558
Purdue Northwest	0	0	0	991	991
Other States					
San Jose State	0	0	0	5,288	5,288
Arizona State	0	0	0	4,430	4,430
Western Kentucky	0	0	0	1,747	1,747
Total	**20,753**	**19,580**	**41,958**	**102,905**	**185,196**

Table A1 reports the number of deduplicated syllabi in the analysis sample for each term, separately by university, with one representative syllabus retained per course-instructor-term group. Universities are grouped by state, with Texas, Florida, Georgia and Indiana shown on their own and the remaining states pooled, and ordered within group by total syllabi. Term coverage differs across universities: 12 of the 50 institutions are observed in every term, while all 50 contribute syllabi in Spring 2026. A zero means no syllabi were collected for that university in that term.

Table A2: Instructor Characteristics by University

University	Instructors (1)	Instructor Coverage (%) Gender (2)	Rank (3)	Any Pub. (4)
Texas				
Houston	2,358	94	77	39
Houston Clear Lake	457	98	70	45
Houston Downtown	660	98	74	35
Lamar	576	98	75	44
North Texas	1,688	93	90	42
North Texas at Dallas	135	92	78	29
Prairie View A&M	526	98	90	46
Sam Houston State	1,735	95	74	40
Stephen F. Austin	775	95	78	38
Tarleton State	635	95	69	34
Texas A&M	3,561	92	73	53
Texas A&M Commerce	771	90	41	25
Texas A&M Corpus Christi	464	96	79	49
Texas A&M International	410	89	92	37
Texas A&M Kingsville	391	87	74	49
Texas A&M San Antonio	379	92	84	38
Texas Southern	399	85	42	33
Texas State	1,922	93	86	32
Texas Tech	1,602	90	72	42
UT Arlington	1,106	93	96	54
UT Austin	3,730	94	85	46
UT Dallas	2,077	81	68	34
UT El Paso	842	98	87	55
UT Permian Basin	206	97	65	36
UT San Antonio	2,472	92	84	32
UT Tyler	396	61	28	18
West Texas A&M	370	95	84	39
Florida				
Central Florida	3,545	87	75	38
Florida	2,768	74	61	36
Florida A&M	652	85	84	36
Florida Atlantic	1,686	93	77	40
Florida Gulf Coast	907	91	86	39
Florida International	3,048	85	71	31
Florida State	2,420	90	77	40
North Florida	1,018	94	94	40
West Florida	609	67	70	36
Georgia				
Albany State	337	87	76	21
Fort Valley State	133	84	74	38
Georgia	2,047	91	84	51
Georgia College	605	95	86	43
Georgia Southern	635	92	79	40
Georgia State	1,158	90	85	35
Kennesaw State	2,293	55	72	27
Indiana				
IU Bloomington	2,686	81	80	36
IU Indianapolis	1,240	83	77	23
Purdue	1,914	89	90	57
Purdue Northwest	400	89	70	40
Other States				
Arizona State	2,234	91	81	47
San Jose State	2,105	92	45	34
Western Kentucky	760	95	61	45
All 50 universities	**65,843**	**88**	**76**	**40**

Note: Table A2 presents instructor coverage by university across the 50 universities in the sample, pooling all terms from Fall 2024 through Spring 2026. Instructors is the number of distinct instructors teaching at the university over this period. The Gender, Rank, and Any Pub. columns report the share of distinct instructors, in percent, with gender, academic rank on the four-group ladder, and at least one publication recorded. Gender combines direct records with a name-based proxy, and Academic Rank reflects any available source (salary database, faculty directory, ORCID employment, or syllabus text). Universities are grouped by state, with Texas, Florida, Georgia and Indiana shown on their own and the remaining states pooled, and ordered alphabetically within group. The bottom row pools all distinct instructors, so it is not the average of the rows above.

Table A3: Types of AI Integration in the Classroom

	Share (%)
Panel A. Teaching Integration	
AI in the Field	27
Using AI Tools	18
AI Methods	18
Ethics and Society	11
Other	40
Observations	11,036
Panel B. Evaluation Integration	
AI-Assisted Work	38
Building AI Models	17
Interacting with AI	11
Other	41
Observations	5,968

Note: Table A3 groups the syllabi classified as integrating AI into teaching (Panel A) and into evaluation (Panel B) into descriptive categories, based on the evidence passages behind each classification. Each entry gives the share of classified syllabi in the category, in percent. A syllabus can fall in more than one category, so shares do not sum to one hundred. The sample covers the 50 universities from Fall 2024 through Spring 2026.

Table A4: Prevalence of the Four Measures by Semester

	Fall 2024	Spring 2025	Fall 2025	Spring 2026
Policy				
AI Policy	39.9	44.0	57.0	61.0
AI Permitted	13.7	16.7	27.0	30.3
Classroom Integration				
Teaching Integration	4.5	5.2	7.2	7.6
Evaluation Integration	1.8	2.0	3.1	3.2
Syllabi	20,748	19,580	25,509	26,401
Institutions	12	12	12	12

Note: Table A4 shows the prevalence of the four measures semester by semester. Each entry gives the share of syllabi exhibiting the row measure, in percent. The table is restricted to the 12 institutions observed in all four terms, so that the change across semesters is not driven by which campuses enter the data in later terms.

Table A5: AI Prevalence by Broad Major Field

	Business (1)	Humanities (2)	STEM (3)	Social Sci. (4)	Other (5)
AI Mentioned	55.4	57.0	48.2	55.8	49.1
Policy					
AI Policy	57.5	60.0	52.6	61.8	57.8
AI Permitted	31.7	23.3	22.0	25.9	26.2
Classroom Integration					
Teaching Integration	10.1	6.1	5.0	3.8	7.2
Evaluation Integration	5.6	2.2	2.1	1.9	3.7
Number of Syllabi	17,835	56,441	69,390	37,638	3,850

Note: Table A5 presents the share of syllabi with each measure, in percent, across the five broad fields into which harmonized majors are grouped, pooling Fall 2024 to Spring 2026. AI Mentioned is a keyword search over the full syllabus text. AI Policy and AI Permitted describe what the syllabus says about student use of AI. Teaching Integration and Evaluation Integration describe AI used in course activities and in graded work.

Table A6: AI Prevalence by Institution, Spring 2026

Institution	**Syllabi**	**AI Policy**	**AI Permitted**	**Teaching Integration**	**Evaluation Integration**
	(1)	(2)	(3)	(4)	(5)
Panel A. Selective					
Central Florida	3,957	65	38	6.3	2.9
Florida	1,707	41	21	11.5	5.4
Florida State	4,147	62	31	4.8	1.5
Georgia	2,205	79	36	7.0	2.6
IU Bloomington	2,705	37	15	4.5	2.3
Purdue	1,698	58	38	7.6	3.3
Texas A&M	5,012	60	18	8.5	2.7
UT Austin	3,272	61	32	8.8	2.6
UT Dallas	1,871	37	18	13.0	5.8
Tier total and mean	*26,574*	*56*	*28*	*8.0*	*3.2*
Panel B. Moderately Selective					
Arizona State	4,426	71	34	6.3	3.2
Florida Atlantic	1,948	100	29	4.6	2.3
Florida Gulf Coast	2,181	32	15	5.6	4.4
Florida International	4,072	35	14	7.6	3.1
Georgia College	1,122	100	2	1.2	2.0
Georgia Southern	1,049	94	16	2.1	1.3
Georgia State	1,721	39	18	3.6	2.3
Houston	3,287	44	22	6.2	2.5
Houston Clear Lake	1,422	60	25	4.6	2.7
IU Indianapolis	1,261	16	9	3.6	1.9
Kennesaw State	4,178	88	35	2.0	1.1
North Florida	1,753	65	34	5.8	2.2
North Texas	2,777	57	24	5.8	2.4
San Jose State	5,284	26	11	5.1	1.8
Texas A&M Corpus Christi	1,008	76	38	3.8	1.9
Texas Southern	882	68	47	5.0	2.7
Texas Tech	1,081	63	33	5.7	3.1
UT Arlington	2,831	74	34	7.5	2.8
UT San Antonio	2,146	97	54	7.9	3.3
UT Tyler	775	75	40	6.1	2.1
West Florida	1,302	63	36	3.0	2.2
Western Kentucky	1,746	60	27	3.7	2.2
Tier total and mean	*48,252*	*64*	*27*	*4.9*	*2.4*
Panel C. Broad Access					
Albany State	941	97	57	2.4	1.1
Florida A&M	1,352	17	10	5.4	3.7
Fort Valley State	462	100	76	2.8	1.1
Houston Downtown	1,951	100	40	5.3	1.5
Lamar	2,342	100	15	3.9	1.7
North Texas at Dallas	335	58	24	4.2	1.5
Prairie View A&M	1,872	100	11	3.5	4.1
Purdue Northwest	991	4	2	3.4	2.4
Sam Houston State	4,035	92	44	4.4	2.3
Stephen F. Austin	1,843	39	20	4.1	1.9
Tarleton State	1,763	89	44	3.6	1.1
Texas A&M Commerce	1,717	84	9	4.1	1.3
Texas A&M International	1,030	82	35	5.6	3.2
Texas A&M Kingsville	939	86	60	5.0	1.3
Texas A&M San Antonio	1,009	78	33	6.6	2.6
Texas State	2,614	75	40	6.2	3.4
UT El Paso	1,497	57	32	5.0	3.5
UT Permian Basin	751	96	22	2.8	2.0
West Texas A&M	601	96	45	5.7	3.3
Tier total and mean	*28,045*	*76*	*33*	*4.4*	*2.3*
All institutions	**102,871**	**67**	**29**	**5.3**	**2.5**

Note: Table A6 presents Spring 2026 prevalence for each of the 50 institutions, grouped by institutional selectivity tier: 9 Selective, 22 Moderately Selective and 19 Broad Access. Column (1) is the number of syllabi the institution contributes and columns (2) to (5) are shares in percent. In the tier rows and the final row, column (1) is a total and columns (2) to (5) are unweighted averages across institutions.

Table A7: AI Prevalence by Course Level

	All	Basic Undergrad.	Advanced Undergrad.	Graduate
	(1)	(2)	(3)	(4)
Panel A. Fall 2024 to Spring 2026				
AI Mentioned	53.2	54.0	54.2	49.4
Policy				
AI Policy	57.3	59.8	57.7	52.2
AI Permitted	24.2	21.0	26.1	25.4
Classroom Integration				
Teaching Integration	5.6	4.6	5.6	7.6
Evaluation Integration	2.5	2.1	2.6	2.8
Number of Syllabi	185,154	62,867	85,339	36,654
Panel B. Spring 2026				
AI Mentioned	54.3	53.2	56.6	51.6
Policy				
AI Policy	64.4	64.7	65.7	60.6
AI Permitted	27.7	23.6	30.2	29.6
Classroom Integration				
Teaching Integration	5.6	4.4	5.7	7.8
Evaluation Integration	2.5	2.1	2.7	2.9
Number of Syllabi	102,871	36,981	45,887	19,768

Note: Table A7 presents the share of syllabi with each measure, in percent, by course level. Panel A pools Fall 2024 to Spring 2026 and Panel B is the Spring 2026 cross-section. Column (1) pools all syllabi, including the small number whose level could not be assigned, so it is not a weighted average of columns (2) to (4).

Table A8: Course Characteristics with Pre-2024 Course Information

	AI Policy (1)	AI Permitted (2)	Teaching Integration (3)	Evaluation Integration (4)
Panel A. Course Level				
Freshman	9.20***	4.10	1.29	1.04
	(2.94)	(2.50)	(2.45)	(2.96)
Sophomore	3.69**	-0.07	-0.52	-1.21*
	(1.60)	(1.34)	(1.14)	(0.72)
Junior	3.36**	0.72	-1.06	-1.86*
	(1.57)	(1.24)	(1.43)	(1.12)
Senior	4.65***	3.09***	-0.14	-0.16
	(1.19)	(0.97)	(0.88)	(0.52)
Panel B. Other Course Characteristics				
Log Enrollment	1.05***	0.33	0.19	0.61***
	(0.39)	(0.34)	(0.33)	(0.22)
Has TA	5.72***	3.23***	1.24*	0.59
	(1.06)	(0.92)	(0.71)	(0.40)
Credit Hours	1.02	0.65	0.07	-0.21
	(0.72)	(0.61)	(0.35)	(0.20)
Average GPA	3.42***	1.18	1.74*	-0.24
	(1.28)	(1.20)	(0.91)	(0.62)
Take-Home Share	0.02	0.05**	0.07***	0.02
	(0.03)	(0.02)	(0.02)	(0.01)
Observations	19,590	19,590	19,590	19,590
Adjusted R^2	0.636	0.599	0.469	0.490

Note: Table A8 shows the multivariate course-characteristic coefficients for the UT Austin and UT Dallas subsample, where additional pre-2024 course information is available. The specification adds Average GPA, the pre-2024 average course grade, and Take-Home Share, the pre-2024 share of course points allocated to homework, papers, and projects, to the main course-characteristic bundle, while continuing to absorb instructor fixed effects and university-by-major-by-term fixed effects. Graduate courses are the omitted category and missing indicators are included for the continuous covariates. Standard errors, two-way clustered by instructor and local course, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Variance Decomposition with Statewide Course Codes

	AI Policy (1)	AI Permitted (2)	Teaching Integration (3)	Evaluation Integration (4)
Instructor	30.3	33.5	24.0	25.3
Local Course	17.3	11.5	15.4	13.7
Statewide Course by Term	13.6	12.0	13.2	8.9
Context	17.4	9.4	7.3	2.5
Unexplained	21.4	33.6	40.0	49.5
Adjusted R^2	78.6	66.4	60.0	50.5
Observations	20,735	20,735	20,735	20,735

Note: Table A9 shows the harmonized-course decomposition for the four AI measures, estimated on the 9 Florida institutions that share a common state course numbering system. The decomposition separates local course variation from statewide-course-by-term variation within a specification that also includes instructor fixed effects and the broader context component. Each entry gives the share of total outcome variance, in percentage points, attributed to the corresponding component.

Table A10: Instructor Characteristics: Heterogeneity by Field and Selectivity

Panel A: STEM vs. Non-STEM				
	STEM		**Non-STEM**	
	AI Policy (1)	Teaching Integration (2)	AI Policy (3)	Teaching Integration (4)
Gender				
Female	65.9	7.7	71.3	10.1
Male	55.5	9.1	63.7	11.0
Academic Rank				
Professor	54.8	8.4	66.3	11.1
Associate Professor	59.9	9.1	69.7	11.2
Assistant Professor	64.2	12.1	72.6	13.6
Non-Tenure-Track	62.1	8.7	66.5	10.4
Research Activity				
Any Publications	59.6	9.8	68.9	12.3
No Publications	59.5	7.6	65.8	9.7
ML Research Tercile				
Low	60.1	5.2	67.1	11.1
Medium	61.4	8.9	72.8	9.6
High	57.4	16.0	67.0	16.8
Instructors	25,580		41,996	
Panel B: Selective vs. Other Institutions				
	Selective		**Other**	
	AI Policy (1)	Teaching Integration (2)	AI Policy (3)	Teaching Integration (4)
Gender				
Female	63.7	11.9	73.1	8.2
Male	54.4	11.9	64.9	9.3
Academic Rank				
Professor	54.0	11.7	67.9	8.9
Associate Professor	60.8	12.6	69.7	9.3
Assistant Professor	61.0	15.7	74.2	11.5
Non-Tenure-Track	58.9	12.2	69.0	8.7
Research Activity				
Any Publications	58.8	12.8	69.7	10.4
No Publications	56.1	11.6	68.4	7.8
ML Research Tercile				
Low	57.7	8.9	67.2	7.8
Medium	64.2	11.7	74.3	8.3
High	54.9	18.0	67.6	15.4
Instructors	24,726		41,024	

Note: Each cell gives the share of instructors, in percent, for whom at least one syllabus exhibits the column outcome. The unit of observation is the instructor within the panel group, so an instructor who teaches in both groups appears in both. Panel A splits by STEM versus non-STEM fields. Panel B splits by selectivity tier, with Selective the top tier and Other the remaining two tiers, together covering all 50 institutions in the sample. Academic rank uses the harmonized four-group ladder measure, with full professors as the omitted group. ML Research measures the machine learning content of an instructor's published research from the titles and abstracts of their papers, ranked among papers published in the same subfield and year, standardized among instructors with a measured research record and set to zero for other linked instructors. Its terciles here are formed from the underlying density before that rescaling, within each panel group, among the instructors with a measured research record.

Table A11: Research Measures by Field

Field	Instructors	Within-Field ML Research	Global ML Research
	(1)	(2)	(3)
Computer Science	1,168	58	88
Decision Sciences	91	58	82
Mathematics	362	51	64
Neuroscience	388	50	63
Economics, Econometrics and Finance	435	48	52
Business, Management and Accounting	872	47	51
Engineering	2,007	51	51
Arts and Humanities	665	47	50
Biochemistry, Genetics and Molecular Biology	865	51	50
Health Professions	205	47	50
Environmental Science	826	50	49
Medicine	1,829	49	47
Psychology	932	45	46
Social Sciences	3,465	45	44
Immunology and Microbiology	45	49	41
Earth and Planetary Sciences	331	46	39
Agricultural and Biological Sciences	588	47	37
Physics and Astronomy	464	49	36
Chemistry	187	53	35
Veterinary	27	41	33
Materials Science	372	44	26
All fields	16,202	49	50

Note: Table A11 presents instructor-level means of the ML Research measure by the field of instructors' publications, expressed as percentiles. The within-field version ranks each paper among papers published in the same subfield and year, while the global version ranks papers across all fields.

Table A12: Instructor Characteristics and Teaching Integration by Category

	Teaching Integration (1)	AI Methods (2)	Using AI Tools (3)	AI in the Field (4)	Ethics and Society (5)	Other (6)
Female	0.03	-0.13*	0.07	0.04	-0.03	0.07
	(0.26)	(0.07)	(0.06)	(0.14)	(0.10)	(0.15)
Associate Prof.	-0.55	-0.03	-0.47***	-0.07	-0.22	-0.28
	(0.36)	(0.15)	(0.15)	(0.24)	(0.17)	(0.19)
Assistant Prof.	0.66	0.39*	-0.03	0.02	-0.10	0.37*
	(0.42)	(0.23)	(0.19)	(0.22)	(0.12)	(0.20)
Non-TT Faculty	-0.69*	0.01	-0.18	-0.04	-0.22	-0.52**
	(0.37)	(0.14)	(0.15)	(0.18)	(0.16)	(0.21)
Any Publications	0.46**	0.13	0.09	-0.02	-0.00	0.23
	(0.21)	(0.09)	(0.12)	(0.11)	(0.08)	(0.15)
Citation Count	-0.11	0.03	0.02	-0.07	-0.10**	-0.00
	(0.17)	(0.06)	(0.07)	(0.11)	(0.05)	(0.11)
ML Research	0.65***	0.24***	0.07	-0.05	0.09***	0.34**
	(0.20)	(0.07)	(0.07)	(0.08)	(0.03)	(0.17)
Text Controls	Yes	Yes	Yes	Yes	Yes	Yes
Course FE	Yes	Yes	Yes	Yes	Yes	Yes
Univ.-Major-Level-Term FE	Yes	Yes	Yes	Yes	Yes	Yes
Outcome mean (%)	5.59	0.93	1.15	1.62	0.65	2.35
Observations	134,766	134,458	134,458	134,458	134,458	134,458
Instructors	50,719	50,641	50,641	50,641	50,641	50,641
Institutions	50	50	50	50	50	50

Note: Table A12 presents coefficients from separate regressions on instructor characteristics. Column (1) is the teaching integration measure itself. The remaining columns equal one hundred when the syllabus is teaching-positive in the named category and zero otherwise. Academic rank uses the harmonized four-group ladder measure, with full professors as the omitted group. Indicators for missing gender and rank are included. ML Research measures how close an instructor's published work is to machine-learning research, relative to other work in the instructor's own field. Citation Count is the log of an instructor's citations. All specifications include syllabus text controls and absorb course and university-by-major-by-level-by-term fixed effects. Standard errors, clustered at the university level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: Instructor Characteristics and Evaluation Integration by Category

	Evaluation Integration (1)	Building AI Models (2)	Interacting with AI (3)	AI-Assisted Work (4)	Other (5)
Female	0.12	-0.04	0.01	0.19**	0.17**
	(0.15)	(0.04)	(0.07)	(0.09)	(0.08)
Associate Prof.	-0.52***	0.06	-0.00	-0.37***	-0.22
	(0.19)	(0.13)	(0.10)	(0.13)	(0.16)
Assistant Prof.	-0.29	0.40**	0.14	-0.74***	0.10
	(0.22)	(0.19)	(0.11)	(0.19)	(0.19)
Non-TT Faculty	-0.59***	0.03	0.01	-0.41***	-0.25
	(0.19)	(0.10)	(0.07)	(0.15)	(0.15)
Any Publications	0.18	0.05	0.01	0.08	0.19
	(0.15)	(0.07)	(0.05)	(0.09)	(0.14)
Citation Count	0.02	-0.02	-0.01	-0.13**	0.12
	(0.12)	(0.04)	(0.03)	(0.06)	(0.10)
ML Research	0.40***	0.16***	-0.03	0.14	0.26***
	(0.10)	(0.06)	(0.04)	(0.08)	(0.08)
Text Controls	Yes	Yes	Yes	Yes	Yes
Course FE	Yes	Yes	Yes	Yes	Yes
Univ.-Major-Level-Term FE	Yes	Yes	Yes	Yes	Yes
Outcome mean (%)	2.56	0.46	0.38	1.26	1.42
Observations	134,766	134,612	134,612	134,612	134,612
Instructors	50,719	50,685	50,685	50,685	50,685
Institutions	50	50	50	50	50

Note: Table A13 presents coefficients from separate regressions on instructor characteristics. Column (1) is the evaluation integration measure itself. The remaining columns equal one hundred when the syllabus is evaluation-positive in the named category and zero otherwise. Academic rank uses the harmonized four-group ladder measure, with full professors as the omitted group. Indicators for missing gender and rank are included. ML Research measures how close an instructor's published work is to machine-learning research, relative to other work in the instructor's own field. Citation Count is the log of an instructor's citations. All specifications include syllabus text controls and absorb course and university-by-major-by-level-by-term fixed effects. Standard errors, clustered at the university level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A14: Instructor Characteristics and AI Outcomes: Additional Research Controls

	Teaching Integration (1)	Evaluation Integration (2)	AI Policy (3)	AI Permitted (4)
Female	0.04	0.12	2.63***	0.92*
	(0.26)	(0.15)	(0.60)	(0.52)
Associate Prof.	-0.54	-0.53***	0.07	-0.20
	(0.36)	(0.19)	(0.97)	(0.72)
Assistant Prof.	0.67	-0.29	1.88**	1.38**
	(0.42)	(0.21)	(0.91)	(0.68)
Non-TT Faculty	-0.67*	-0.58***	1.20**	0.55
	(0.37)	(0.19)	(0.58)	(0.44)
Any Publications	0.48**	0.17	0.43	0.62
	(0.21)	(0.16)	(0.40)	(0.44)
Citation Count	-0.26	-0.01	-0.26	-1.04
	(0.29)	(0.25)	(0.71)	(0.76)
ML Research	0.66***	0.40***	-0.20	0.42
	(0.20)	(0.10)	(0.25)	(0.30)
Publication Count	0.16	0.04	0.27	0.55
	(0.25)	(0.23)	(0.60)	(0.62)
Research Novelty	0.21	0.01	0.03	-0.05
	(0.23)	(0.12)	(0.32)	(0.38)
PhD Highly Selective	0.38	-0.43	1.95	-0.33
	(0.73)	(0.52)	(1.47)	(1.65)
Text Controls	Yes	Yes	Yes	Yes
Course FE	Yes	Yes	Yes	Yes
Univ.-Major-Level-Term FE	Yes	Yes	Yes	Yes
Outcome mean (%)	5.59	2.56	57.42	23.67
Observations	134,766	134,766	134,761	134,766
Instructors	50,719	50,719	50,717	50,719
Institutions	50	50	50	50

Note: Table A14 presents coefficients from separate regressions of each column outcome, measured in percentage points, on instructor characteristics, adding three research controls to the main specification. Publication Count is the log of an instructor's publications and Research Novelty measures how unusual the topic combinations in an instructor's published work are. PhD Highly Selective equals one when the instructor's doctorate comes from a highly selective institution, with an indicator for instructors whose doctoral institution is not observed. Academic rank uses the harmonized four-group ladder measure, with full professors as the omitted group. Indicators for missing gender and rank are included. ML Research measures how close an instructor's published work is to machine-learning research, relative to other work in the instructor's own field. Citation Count is the log of an instructor's citations. All specifications include syllabus text controls and absorb course and university-by-major-by-level-by-term fixed effects. Standard errors, clustered at the university level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A15: Alternative Measures of Machine-Learning Research

	Teaching Integration (1)	Evaluation Integration (2)	AI Policy (3)	AI Permitted (4)
ML Research (Baseline)	0.65***	0.40***	-0.18	0.43
	(0.20)	(0.10)	(0.24)	(0.31)
Global ML Similarity	0.92***	0.50***	-0.09	0.83***
	(0.21)	(0.16)	(0.29)	(0.25)
AI Vocabulary Density	1.03***	0.51**	0.30	0.57**
	(0.31)	(0.22)	(0.30)	(0.29)
Any AI or ML Publication	1.46***	0.48	-0.95	0.77
	(0.50)	(0.36)	(0.69)	(0.68)
Text Controls	Yes	Yes	Yes	Yes
Course FE	Yes	Yes	Yes	Yes
Univ.-Major-Level-Term FE	Yes	Yes	Yes	Yes
Outcome mean (%)	5.59	2.56	57.42	23.67
Observations	134,766	134,766	134,761	134,766
Instructors	50,719	50,719	50,717	50,719
Institutions	50	50	50	50

Note: Table A15 presents coefficients on alternative measures of machine-learning research from separate regressions of each column outcome, measured in percentage points, on instructor characteristics. Each row replaces the machine-learning research measure in the main specification with the named measure, keeping all other covariates. ML Research (Baseline) measures how close an instructor's published work is to machine-learning research, relative to other work in the instructor's own field. Global ML Similarity measures that closeness relative to all published work rather than to work in the instructor's own field. AI Vocabulary Density measures how frequently machine-learning terms appear in an instructor's published work. Any AI or ML Publication equals one when the instructor has published at least one AI or machine-learning paper. All specifications include syllabus text controls and absorb course and university-by-major-by-level-by-term fixed effects. Standard errors, clustered at the university level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A16: Instructor Characteristics and Teaching Integration by Broad Field

	Business (1)	Humanities (2)	STEM (3)	Social Sci. (4)	Other (5)
Female	1.58*	-0.21	0.00	-0.41	2.95**
	(0.82)	(0.40)	(0.31)	(0.37)	(1.31)
Associate Prof.	-1.02	-0.98	-0.74*	0.58	-2.14
	(1.42)	(0.73)	(0.39)	(0.51)	(3.21)
Assistant Prof.	-1.97	1.39*	0.32	1.80**	-4.38
	(1.51)	(0.85)	(0.50)	(0.85)	(2.91)
Non-TT Faculty	-2.01	-1.12	-0.48	0.33	-3.91
	(2.12)	(0.68)	(0.33)	(0.59)	(2.85)
Any Publications	0.33	0.29	0.77**	0.26	3.16
	(1.05)	(0.59)	(0.37)	(0.39)	(3.66)
Citation Count	-1.51**	0.27	-0.33*	0.58	-1.68*
	(0.70)	(0.37)	(0.18)	(0.37)	(1.01)
ML Research	-0.09	0.55	0.79***	0.47	4.61**
	(0.73)	(0.34)	(0.22)	(0.43)	(2.13)
Text Controls	Yes	Yes	Yes	Yes	Yes
Course FE	Yes	Yes	Yes	Yes	Yes
Univ.-Major-Level-Term FE	Yes	Yes	Yes	Yes	Yes
Outcome mean (%)	10.11	6.40	4.70	3.40	7.59
Observations	14,327	42,141	49,868	26,696	1,725
Instructors	5,544	16,323	18,838	11,061	971
Institutions	49	50	50	50	39

Note: Table A16 presents coefficients from separate regressions of teaching integration, measured in percentage points, on instructor characteristics, estimated within each broad field. Academic rank uses the harmonized four-group ladder measure, with full professors as the omitted group. Indicators for missing gender and rank are included. ML Research measures how close an instructor's published work is to machine-learning research, relative to other work in the instructor's own field. Citation Count is the log of an instructor's citations. All specifications include syllabus text controls and absorb course and university-by-major-by-level-by-term fixed effects. Standard errors, clustered at the university level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: Instructor Characteristics and Evaluation Integration by Broad Field

	Business (1)	Humanities (2)	STEM (3)	Social Sci. (4)	Other (5)
Female	0.58	0.43	0.02	-0.72***	3.31**
	(0.66)	(0.31)	(0.29)	(0.23)	(1.54)
Associate Prof.	-1.38	-0.86**	-0.22	-0.26	2.21
	(1.18)	(0.37)	(0.27)	(0.44)	(2.02)
Assistant Prof.	-2.54**	-0.31	0.50	-0.50	-4.67
	(1.26)	(0.52)	(0.32)	(0.64)	(3.91)
Non-TT Faculty	-1.62	-0.98**	-0.21	-0.33	-0.50
	(1.28)	(0.40)	(0.23)	(0.49)	(1.21)
Any Publications	-0.45	0.33	0.01	0.54**	5.58**
	(0.68)	(0.33)	(0.20)	(0.27)	(2.79)
Citation Count	-0.75	0.15	-0.06	0.42**	-0.87
	(0.61)	(0.30)	(0.17)	(0.20)	(1.70)
ML Research	0.15	0.80***	0.33**	0.14	2.75*
	(0.60)	(0.24)	(0.13)	(0.18)	(1.61)
Text Controls	Yes	Yes	Yes	Yes	Yes
Course FE	Yes	Yes	Yes	Yes	Yes
Univ.-Major-Level-Term FE	Yes	Yes	Yes	Yes	Yes
Outcome mean (%)	5.90	2.46	2.01	1.87	4.12
Observations	14,327	42,141	49,868	26,696	1,725
Instructors	5,544	16,323	18,838	11,061	971
Institutions	49	50	50	50	39

Note: Table A17 presents coefficients from separate regressions of evaluation integration, measured in percentage points, on instructor characteristics, estimated within each broad field. Academic rank uses the harmonized four-group ladder measure, with full professors as the omitted group. Indicators for missing gender and rank are included. ML Research measures how close an instructor's published work is to machine-learning research, relative to other work in the instructor's own field. Citation Count is the log of an instructor's citations. All specifications include syllabus text controls and absorb course and university-by-major-by-level-by-term fixed effects. Standard errors, clustered at the university level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A18: Instructor Characteristics and Teaching Integration by Selectivity Tier

	Selective (1)	Moderately Selective (2)	Broad Access (3)
Female	0.58	-0.43	0.22
	(0.39)	(0.39)	(0.41)
Associate Prof.	-0.56	-0.79	0.13
	(0.67)	(0.52)	(0.85)
Assistant Prof.	1.47***	0.72	-1.00
	(0.57)	(0.67)	(0.65)
Non-TT Faculty	-0.94	-0.37	-1.17**
	(0.73)	(0.52)	(0.57)
Any Publications	0.38	0.67***	0.08
	(0.40)	(0.26)	(0.60)
Citation Count	-0.14	-0.15	0.27
	(0.35)	(0.17)	(0.38)
ML Research	0.76	0.81***	-0.08
	(0.46)	(0.24)	(0.34)
Text Controls	Yes	Yes	Yes
Course FE	Yes	Yes	Yes
Univ.-Major-Level-Term FE	Yes	Yes	Yes
Outcome mean (%)	7.01	4.78	4.56
Observations	51,153	59,004	24,609
Instructors	18,652	23,140	8,927
Institutions	9	22	19

Note: Table A18 presents coefficients from separate regressions of teaching integration, measured in percentage points, on instructor characteristics, estimated within each selectivity tier. Tiers reflect the mean SAT score of the 2019 entering class and together cover every institution in the sample. Academic rank uses the harmonized four-group ladder measure, with full professors as the omitted group. Indicators for missing gender and rank are included. ML Research measures how close an instructor's published work is to machine-learning research, relative to other work in the instructor's own field. Citation Count is the log of an instructor's citations. All specifications include syllabus text controls and absorb course and university-by-major-by-level-by-term fixed effects. Standard errors, clustered at the university level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B: Data Construction

B.1 Main Sample

I collect course syllabi from 50 public universities across seven states. The sample covers 27 universities in Texas, nine in Florida, seven in Georgia, four in Indiana, and one each in Arizona, California, and Kentucky. I draw on each university's publicly accessible syllabus records to collect documents for the academic terms between Fall 2024 and Spring 2026, with the set of terms observed varying across universities (Table B1). For UT Austin and UT Dallas, I additionally recover syllabi dating back to Fall 2015, which provides a longer window for the longitudinal analyses.

Data sources. How universities make syllabi public varies across the 50 institutions in the sample. Some maintain dedicated PDF portals, while others use course management platforms with structured metadata or searchable systems organized by department and academic term. This variation determines how much course information I can recover directly from the source and how much I must extract from the syllabus text. Table B1 lists the universities in the sample and the number of syllabi I collect at each institution, together with characteristics that describe their size, selectivity, and research orientation.

University characteristics. To characterize the universities in the sample, I draw on IPEDS and the Opportunity Insights database of Chetty et al. (2020), which together cover enrollment, admissions selectivity, institutional resources, and students' socioeconomic composition (Table A2). I use these characteristics when examining how the integration of AI varies across universities.

Course and instructor information. For each syllabus in the sample, I recover six core metadata fields, covering the instructor's name, the course title, the department code and course number, the section identifier, and the academic term in which the course is offered. These fields determine how each syllabus is linked to a specific instructor, course, and academic unit.

Three of these fields, the department code, course number, and academic term, are typically recorded directly by each university's data system rather than extracted from the syllabus text itself. I recover these for effectively all syllabi in the sample, as they are encoded in the file structure or in structured metadata fields provided by each university's portal. For the small number of records where the academic term is not available from the source, I recover it by searching the syllabus text for explicit references to the semester and academic year.

Instructor names and course titles require a more involved extraction approach because their format and placement vary across institutions. For instructor names, I search the opening pages of each syllabus using text-matching rules ordered by precision, beginning with

Table B1: Universities in the Sample

University	Undergraduate Enrollment	Mean SAT	Research Exp. per FTE ($)	Graduate Share (%)	Syllabi
Texas (27 universities)					
East Texas A&M	9,034	1,055	487	36	1,717
Houston	42,050	1,240	3,925	19	6,676
Houston–Clear Lake	7,592	1,080	264	27	1,424
Houston–Downtown	15,483	1,035	126	10	1,951
Lamar	9,546	1,020	416	56	2,343
North Texas	37,529	1,105	2,064	28	2,780
North Texas at Dallas	3,760	905	372	18	335
Prairie View A&M	8,831	940	2,705	13	1,872
Sam Houston State	20,021	1,040	590	16	4,484
Stephen F. Austin	10,933	1,070	316	14	4,190
Tarleton State	14,650	1,055	1,512	16	1,766
Texas A&M	60,118	1,260	13,322	23	6,181
Texas A&M International	7,844	1,025	554	21	1,030
Texas A&M–Corpus Christi	8,998	1,075	2,979	34	1,008
Texas A&M–Kingsville	5,385	1,070	3,571	23	939
Texas A&M–San Antonio	7,619	960	526	10	1,009
Texas Southern	7,580	900	701	21	882
Texas State	36,613	1,085	3,256	12	5,391
Texas Tech	34,406	1,180	5,565	21	4,979
UT Arlington	37,854	1,110	2,391	31	2,834
UT Austin	43,718	1,360	17,781	20	13,393
UT Dallas	22,941	1,290	4,141	32	7,785
UT El Paso	23,509	980	4,815	17	1,498
UT Permian Basin	5,695	1,010	634	24	751
UT San Antonio	33,259	1,100	4,342	13	8,710
UT Tyler	8,231	1,115	2,926	27	775
West Texas A&M	7,772	1,060	834	27	2,412
Florida (9 universities)					
Central Florida	68,470	1,270	2,472	15	15,530
Florida	38,113	1,390	16,135	38	6,525
Florida A&M	8,704	1,090	3,138	16	2,672
Florida Atlantic	29,440	1,125	2,441	20	8,100
Florida Gulf Coast	15,924	1,125	513	13	2,188
Florida International	56,786	1,155	3,118	18	10,661
Florida State	36,242	1,315	5,398	27	4,147
North Florida	16,840	1,100	928	14	1,838
West Florida	11,464	1,125	910	34	1,304
Georgia (7 universities)					
Albany State	6,971	820	212	8	941
Fort Valley State	2,596	858	4,142	10	462
Georgia	33,422	1,275	12,019	25	5,785
Georgia College	5,863	1,150	132	20	1,700
Georgia Southern	25,275	1,070	1,384	15	1,240
Georgia State	33,129	1,160	5,516	21	1,881
Kennesaw State	44,419	1,120	360	11	6,484
Indiana (4 universities)					
IU Bloomington	48,857	1,285	2,376	19	5,430
IU Indianapolis	20,055	1,135	8,844	34	2,558
Purdue	41,678	1,325	7,364	26	4,179
Purdue Northwest	9,413	1,030	509	10	991
Other States (3 universities)					
Arizona State	70,646	1,210	5,827	19	4,430
San Jose State	29,480	1,125	73	26	5,288
Western Kentucky	17,217	915	412	13	1,747
Total (50)					**185,196**

Note: Table B1 presents characteristics and syllabi counts for each of the 50 universities in the sample, which covers 185,196 syllabi in total. I obtain undergraduate enrollment, mean SAT, research expenditures per full-time-equivalent student, and the graduate enrollment share from IPEDS. I compute mean SAT as the sum of the verbal and math section midpoints between the 25th and 75th percentile scores.

explicitly labeled fields and then moving to less structured patterns such as academic honorifics and university email addresses. I then standardize each extracted name so that the same instructor is identified consistently across documents. For course titles, I apply an analogous approach, supplemented by institution-specific rules where formatting differs substantially.

For the remaining syllabi where the text-based extraction does not recover the instructor name or course title, I send the opening pages of the document to a large language model that returns the missing field along with a reliability indicator. I use a standardized prompt, reported in the replication materials. Altogether, I recover instructor names for 99 percent of syllabi and course titles for 97 percent across the 50 universities.

Sample construction. I restrict the sample to the academic terms between Fall 2024 and Spring 2026. I then apply quality filters that exclude records without substantive syllabus content, covering course listing pages, preliminary course descriptions, empty course management system shells, and documents too short to describe the course and its policies. I additionally exclude course types that do not correspond to classroom instruction, such as practicum and independent-study registrations (Table B2).

Because instructors often teach several sections of the same course with identical syllabi, I deduplicate the syllabi at the section level, following Biasi and Ma (2022). I group syllabi by institution, department, course number, academic term, and normalized instructor name, and keep the longest document in each group. As such, the final estimation sample contains 185,196 unique course-section observations, covering 82,129 courses taught by 65,843 instructors across the 50 universities and 42 harmonized college majors.

Table B2: Sample Construction

	Records (1)	**Removed at Stage** (2)	**Retained (%)** (3)
Candidate syllabus records	214,492	—	100.0
After removing non-syllabus and duplicate records	201,795	12,697	94.1
After removing non-classroom course types	187,598	14,197	87.5
Final analysis sample	**185,196**	**2,402**	**86.3**

Note: Table B2 presents the construction of the analysis sample across the 50 universities. The first row counts all candidate records entering the sample construction in the covered academic terms. The second row removes records without substantive syllabus content along with duplicate documents. The third row removes course types that do not correspond to classroom instruction, including internships, independent study, thesis and dissertation registrations, applied music and physical education activity courses, and study abroad shells, as well as documents shorter than 150 words. The final row additionally removes practicum and clinical records. Column (2) presents the records removed at each stage, and column (3) the share retained relative to the first row.

B.2 Instructor Data

I build the instructor dataset from the syllabus metadata, which identify 65,843 unique instructors active between Fall 2024 and Spring 2026 across the 50 universities. I then link this roster to six external source families covering bibliometric records, education records, publicly available CVs and faculty profiles, department websites, state salary databases,

and residual text-based or name-based sources. Across these sources, I recover academic rank for 69 percent of instructors, gender for 82 percent, research indicators for 37 percent, and years since PhD for 21 percent. Table A2 shows coverage across the 50 universities.

OpenAlex. I use OpenAlex, an open bibliometric database covering publications, citations, and author records across all academic fields, to recover measures of instructors' research productivity and career trajectories. I match the instructor roster to the pool of authors affiliated with each university, using exact and progressively more flexible name comparisons and reviewing ambiguous cases by hand. Match rates remain below one hundred percent because many instructors in the sample do not maintain indexed research profiles.

From matched OpenAlex profiles, I recover the h-index, publication count, citations, research topics, and the years since first and last publication (Table B3). The first-publication year serves as a lower-bound career-stage proxy when direct PhD information is unavailable.

ORCID. Approximately 30 percent of instructors in the roster have ORCID identifiers linked through their OpenAlex profiles. I query these ORCID records, and supplement them when needed with name-based searches using institutional affiliation filters, to recover education records and related employment information. ORCID education records are the highest-priority source for PhD information because they are researcher verified, and they account for the largest share of the PhD completion years I recover across sources.

Instructors' CVs. I also recover instructors' educational backgrounds from publicly available CVs and faculty profile pages. I search the text of these documents for degree information and recover the year in which the instructor completed the PhD. These records provide PhD completion years for instructors who do not appear in ORCID, and they also help confirm academic rank where I do not observe titles elsewhere.

Faculty directories. I collect faculty directory information by downloading profile pages from departmental and institutional websites, from which I recover academic rank, PhD year, PhD institution, and research area. Because departments update these pages as faculty are hired and promoted, the recovered titles generally reflect instructors' current positions. Faculty directories are especially useful where salary data provide limited information on academic rank, particularly in Indiana.

State salary records. In this context, publicly available state salary records provide a further source of instructors' official titles. Because these records report the title under which each instructor is employed, they provide especially reliable information on academic rank. I match instructors to these records by name and normalize the reported job titles to a common four-level rank hierarchy, and Texas records additionally provide gender and hire date.

Table B3: Research Productivity among Matched Instructors by University

University	Instructors	Matched	Conditional on Match			
		(%)	Publications	Citations	h-index	Any ML Pub. (%)
Texas						
East Texas A&M	771	24.9	49	1,074	9	39.4
Houston	2,358	39.3	67	1,756	13	48.4
Houston–Clear Lake	457	44.9	36	878	8	34.7
Houston–Downtown	660	35.3	39	892	8	38.4
Lamar	576	43.9	61	1,224	9	42.2
North Texas	1,688	41.8	67	1,592	12	47.3
North Texas at Dallas	135	28.9	50	1,785	11	23.7
Prairie View A&M	526	46.2	64	1,250	9	41.1
Sam Houston State	1,735	40.0	32	604	7	35.8
Stephen F. Austin	775	11.9	27	349	5	19.8
Tarleton State	635	31.0	28	736	6	25.1
Texas A&M	3,561	53.4	85	2,554	16	51.5
Texas A&M International	410	36.6	32	756	7	33.3
Texas A&M–Corpus Christi	464	48.7	40	1,064	9	41.4
Texas A&M–Kingsville	391	48.6	52	1,308	10	42.1
Texas A&M–San Antonio	379	37.7	48	947	8	36.9
Texas Southern	399	33.1	37	774	7	25.2
Texas State	1,922	31.8	36	1,029	8	37.2
Texas Tech	1,602	41.6	49	1,396	10	41.4
UT Arlington	1,106	53.8	87	3,049	15	55.8
UT Austin	3,730	46.0	79	2,713	16	51.4
UT Dallas	2,077	34.4	77	2,130	15	60.5
UT El Paso	842	55.2	49	975	9	45.3
UT Permian Basin	206	36.4	66	2,883	11	50.7
UT San Antonio	2,472	31.8	56	1,321	12	42.9
UT Tyler	396	18.4	43	1,766	10	37.0
West Texas A&M	370	39.2	27	1,192	7	32.8
Florida						
Central Florida	3,545	38.0	53	1,212	11	47.4
Florida	2,768	35.7	77	2,455	15	50.8
Florida A&M	652	35.7	33	836	8	29.2
Florida Atlantic	1,686	39.7	42	1,006	10	38.6
Florida Gulf Coast	907	38.9	40	946	8	30.9
Florida International	3,048	30.6	54	1,319	11	38.3
Florida State	2,420	40.2	51	1,483	11	40.3
North Florida	1,018	40.4	37	668	8	40.5
West Florida	609	20.2	46	662	8	42.5
Georgia						
Albany State	337	20.8	17	327	4	32.1
Fort Valley State	133	37.6	59	1,664	11	42.9
Georgia	2,047	51.2	63	1,898	15	45.6
Georgia College	605	43.3	27	482	5	28.9
Georgia Southern	635	39.5	52	1,222	7	41.7
Georgia State	1,158	34.7	39	1,079	8	36.1
Kennesaw State	2,293	21.2	33	667	7	37.5
Indiana						
IU Bloomington	2,686	36.2	62	1,569	12	43.2
IU Indianapolis	1,240	23.3	47	1,323	10	39.6
Purdue	1,914	56.9	75	1,960	15	53.6
Purdue Northwest	400	39.5	39	1,162	7	38.3
Other States						
Arizona State	2,234	46.8	89	2,615	15	49.2
San Jose State	2,105	34.3	47	919	8	46.1
Western Kentucky	760	44.7	36	674	8	34.6
Total	**65,843**	**38.9**	**59**	**1,621**	**12**	**44.5**

Note: Table B3 presents research productivity for instructors linked to the analysis sample at each university. The roster contains 65,843 distinct instructors and 25,585 have an accepted research profile. Institution counts assign each instructor to every university where the instructor teaches. The total counts each distinct instructor once. I compute publication, citation, and h-index means over matched instructors with a defined value. Any ML Publication is the share of matched instructors with at least one publication related to machine learning, computed over matched instructors with retrieved publication records.

Other sources. For instructors whose academic rank I do not recover from salary databases or faculty directories, I extract rank information directly from syllabus text by searching instructor information blocks and signature lines for academic titles. For instructors without direct gender records, I impute gender using U.S. Social Security Administration first-name frequencies, assigning gender only when one category accounts for at least 75 percent of births associated with that first name.

Consolidating sources. Because several data sources may provide the same characteristic for a given instructor, I prioritize official administrative records over self-reported data, and self-reported data over web-based or syllabus-derived information. Applying this ordering, I classify instructors into four rank categories and recover gender, while OpenAlex provides a binary publication measure and the log h-index for matched researchers. I measure career experience using recovered PhD year, earliest publication year, and, for Texas instructors, hire year, and I include missing-data indicators for covariates with incomplete coverage.

ML Research. I measure the machine learning content of each instructor's research from the titles and abstracts of their publications. I represent each paper with SPECTER2, a language model trained on scientific text (Singh et al., 2023), and score its proximity to a benchmark corpus of machine learning research drawn from OpenAlex's artificial intelligence subfields and topics. I retain papers with an English-language abstract of at least 200 characters, rank each paper's score among all papers published in the same OpenAlex subfield and year, excluding the instructor's own papers from the reference group, and average these ranks at the instructor level. In the analysis, the measure is standardized among instructors with a measured research record and set to zero for other matched instructors. A global version ranks papers across all fields instead of within subfields.

B.3 Course Classification and Text Measures

To compare courses across the 50 universities, I map institution-specific departments and course numbers into common classifications, and I complement them with a set of measures constructed from the syllabus text itself. I describe each in turn.

B.3.1 Harmonization of College Majors

Across the 50 universities in the estimation sample, I observe more than 2,700 distinct department codes, reflecting different administrative naming conventions across states and institutions. A code such as "BIO" may appear in one place, while another university uses "BIOL" or related prefixes, so a systematic mapping is needed before cross-university comparisons make sense. I harmonize these codes into 42 named categories using the framework of Biasi and Ma (2022), with direct mappings for most codes and university-specific overrides where the same prefix refers to different fields at different institutions. For com-

posite or cross-listed courses, I assign the course to the first-listed department, and this procedure gives me a harmonized major for 98 percent of syllabi. The remaining records fall into a residual "Other" category.

For broader comparisons, I aggregate the 42 harmonized majors into five macro-fields, namely STEM, Humanities, Social Sciences, Business, and Other. The mapping follows mostly straightforward disciplinary boundaries, although a few boundary cases require CIP-based judgment.[1] Table B4 lists the full mapping from department codes to harmonized majors.

B.3.2 Major Characteristics

I complement the syllabus-level data with external major-level covariates that capture labor market conditions, occupational AI exposure, and student selection into each field. These covariates allow me to examine which characteristics of a major predict cross-field variation in AI classroom integration and also serve as controls in specifications that isolate the instructor-level component of AI adoption.

I draw on the Federal Reserve Bank of New York's College Labor Market data (Abel et al., 2014) to recover unemployment rates, underemployment rates, early- and mid-career median wages, and the share of graduates who go on to earn a graduate degree. I also measure each major's exposure to language modeling advances using the LMOE index from Felten et al. (2023), assigning occupation-level scores to majors through a CIP-to-SOC crosswalk weighted by BLS employment counts.

From the 2019–20 National Postsecondary Student Aid Study, I recover mean SAT composite scores, mean ACT scores, and mean college GPA for 23 broad major fields. At the institution-by-major level, IPEDS Completions and the College Scorecard provide degree counts, the share of female graduates, and post-graduation median earnings. In the main heterogeneity specifications, I focus on six covariates, namely LMOE, unemployment, mid-career wage, graduate-degree share, mean SAT, and share female.

B.3.3 Course Levels

I classify each course into one of three academic levels using the first digit of its course number, following Biasi and Ma (2022). Under this convention, courses numbered 100 through 299 are basic undergraduate, those numbered 300 through 499 are advanced undergraduate, and those numbered 500 and above are graduate. This heuristic follows a longstanding convention across public universities and covers the vast majority of the syllabi in the sample.

For the main specifications, I refine this classification into five tiers that distinguish freshman, sophomore, junior, senior, and graduate courses. I construct this finer classifi-

[1] I classify Education and Criminal Justice as Social Sciences following CIP conventions, Nursing as STEM following NSF and CIP usage, and Interdisciplinary Studies as Humanities based on the composition of its largest constituent departments.

Table B4: Harmonized Majors: Sample Distribution and Illustrative Sub-Fields

		Illustrative Sub-Fields		
Harmonized Major	N	(1)	(2)	(3)
Panel A: Business (N = 17,837)				
Business Administration	9,271	Management	Supply Chain	Entrepreneurship
Accounting	2,616	Auditing	Managerial Accounting	
Finance	2,436	Real Estate	Risk Management	Financial Planning
Marketing	2,129	Sales	Digital Marketing	
Hospitality	1,385	Tourism	Hotel Management	Culinary Arts
Panel B: Humanities (N = 56,455)				
English	9,390	Creative Writing	Literature	Comparative Literature
Communication	8,788	Journalism	Advertising	Public Relations
Art	8,131	Art History	Graphic Design	Digital Media
Interdisciplinary Studies	7,420	Honors	Gender Studies	Global Studies
Music	6,601	Music Education	Music Theory	Music Performance
Foreign Languages	5,852	Spanish	French	Linguistics
History	3,723	American History	European History	
Film & Theater	3,560	Theater	Film Studies	Drama
Philosophy	2,255	Religious Studies	Ethics	
Dance	735			
Panel C: STEM (N = 69,408)				
Mathematics	9,318	Statistics	Actuarial Science	Applied Mathematics
Biology	8,788	Microbiology	Neuroscience	Biochemistry
Engineering (Other)	7,939	Mechanical Eng.	Aerospace Eng.	Industrial Eng.
Health Sciences	7,765	Kinesiology	Speech Pathology	Occupational Therapy
Computer Science	5,214	Data Science	Cybersecurity	Software Engineering
Nursing	4,058	Graduate Nursing		
Chemistry	3,860	Biochemistry	Medicinal Chemistry	
Public Health	3,308	Pharmacy	Nutrition	Epidemiology
Information Systems	3,284	Information Technology	Management Info. Systems	
Physics	3,244	Astronomy	Space Science	
Agriculture	2,790	Agronomy	Animal Science	Food Science
Electrical Engineering	2,766	Computer Engineering	Electronics	
Architecture	2,305	Urban Planning	Interior Design	Landscape Architecture
Civil Engineering	1,419	Construction	Water Resources	
Geology	1,328	Earth Science	Meteorology	Atmospheric Science
Environmental Science	1,204	Sustainability	Oceanography	Marine Science
Biomedical Engineering	818	Bioengineering		
Panel D: Social Sciences (N = 37,643)				
Education	11,346	Curriculum & Instruction	Special Education	Reading Education
Psychology	6,346	Cognitive Science	Clinical Psychology	Behavioral Analysis
Political Science	3,747	International Relations	Public Policy	Government
Social Work	3,490	Counseling	Human Development	Family Studies
Criminal Justice	2,910	Criminology	Forensic Science	Law Enforcement
Economics	2,708	Agricultural Economics	Real Estate	
Sociology	2,051	Demography	Social Stratification	
Public Administration	1,945	Public Policy	Urban Planning	
Anthropology	1,581	Archaeology	Biological Anthropology	
Geography	1,519	Geographic Info. Systems	Urban Geography	
Panel E: Other (N = 3,853)				
Other	3,853	Legal Studies	Family & Consumer Sciences	

Note: Table B4 presents the distribution of the estimation sample across 42 harmonized majors, along with illustrative sub-fields that map to each category. The sample includes 185,196 section-deduplicated syllabi across the 50 universities in the sample from Fall 2024 through Spring 2026. Majors are sorted by macro-field, then by descending count within each field.

cation for all 50 universities, with institution-specific adjustments where numbering conventions differ.

For instance, UT Austin uses a three-digit numbering system in which the first digit encodes credit hours rather than academic level. For the Texas universities in the Texas Common Course Numbering System, the four-digit course number also encodes credit hours in its second digit. For the Florida universities, I use the Statewide Course Numbering System,

which provides comparable identifiers across institutions.

Across the full sample, 34 percent of syllabi correspond to basic undergraduate courses, 46 percent to advanced undergraduate, and 20 percent to graduate courses. Within the undergraduate tiers, freshman and sophomore courses account for 16 and 18 percent of the sample, while junior and senior courses account for 22 and 24 percent.

B.3.4 Course Characteristics

For four universities in the sample, I recover section-level grade distributions from publicly available registrar records. I use UT Austin and UT Dallas grade repositories, Texas A&M registrar files, and Purdue's public grade pages. From these records, I compute mean GPA on a 4.0 scale, the share of students receiving a D, F, or Withdrawal, and the share receiving an A. Purdue lists percentages rather than counts, so its grade data do not include enrollment.

I also recover section-level enrollment or capacity for a subset of universities from registrar and schedule databases, including IU Bloomington, IU Indianapolis, Florida, Houston, Florida Atlantic, Florida International, and Georgia. Because these sources differ somewhat in coverage and precision across universities, I combine them into a harmonized enrollment measure that uses observed enrollment when available and otherwise relies on seat counts, current capacity, or historical averages.

I also recover semester credit hours for a subset of the sample. At the Texas universities, credit hours are encoded in course numbering conventions, and for the Florida universities I use the Statewide Course Numbering System catalog. Credit hours are available for 28 of the 50 universities, covering roughly half the sections in the sample, and I include a missing indicator for the remaining universities in every specification that uses this variable.

B.3.5 Syllabus Text Characteristics

I construct nine text-based controls from the syllabus text, capturing the complexity of the document's prose, the prescriptiveness of the instructor's tone, and the presence of explicit course policies. The first three capture text complexity, namely the Flesch-Kincaid Grade Level, the Moving-Average Type-Token Ratio, and average sentence length. I additionally include the log of total word count, which absorbs the mechanical correlation between syllabus length and the likelihood of detecting any keyword-based outcome. Across the 50 universities, average syllabus length ranges from about 550 words at Albany State to more than 5,500 at Texas A&M International.

I also capture instructor tone through the prescriptive modal share, penalty language density, supportive language density, and binary indicators for explicit late-work and attendance policies. All continuous text controls are standardized to mean zero and unit standard deviation in the estimation sample. I recover all nine measures for every syllabus in the main sample.

I also identify whether a teaching assistant is mentioned in each syllabus, combining text-based detection with structured platform metadata where available. This approach

identifies a TA in approximately 17 percent of syllabi. Table B5 shows definitions and summary statistics for the nine text controls.

Table B5: Text Control Variables: Summary Statistics

Variable	Description	Mean	SD
Flesch–Kincaid Grade	Readability index (higher = more complex)	14.84	3.24
Lexical Diversity (MATTR)	Moving-average type–token ratio (50-word window)	0.72	0.03
Avg. Sentence Length	Mean words per sentence	21.99	5.12
Modal Prescriptive Share	Share of modal verbs that are prescriptive	0.27	0.16
Penalty Language Density	Penalty-related terms per 1,000 words	2.27	2.35
Supportive Language Density	Supportive/encouraging terms per 1,000 words	2.51	2.03
Has Late-Work Policy	Binary indicator for a late-work policy	0.56	0.50
Has Attendance Policy	Binary indicator for an attendance policy	0.58	0.49
Log Word Count	Natural log of full-text word count	7.63	0.76

Note: Table B5 presents summary statistics for the nine text-derived control variables used in the main specifications. I compute all metrics directly from the syllabus full text. Flesch–Kincaid Grade and Avg. Sentence Length are winsorized at the 99th percentile to limit the influence of syllabi lacking sentence punctuation. The sample includes the 185,196 section-deduplicated syllabi across the 50 universities.

B.3.6 AI-Related Content

To identify syllabi that contain AI-related content, I use ten keyword patterns in two tiers. The first covers policy-oriented post-ChatGPT terms, and the second covers broader references such as "artificial intelligence," "AI," "generative AI," and "large language model." I exclude terms with high false-positive rates and those that add little signal beyond these core patterns. The standalone acronym "AI" requires special treatment because it often refers to "Associate Instructor," building codes, and room numbers, so I apply a contextual filter blocking known false positives and requiring a nearby AI-related term. I use the resulting indicator of any AI mention to compute the AI-mention rates I present in Section 2 of the main paper. The four classification outcomes are constructed separately from the full syllabus text, as I describe in Appendix C.

B.4 Representativeness of Sample

I assess how well the collected syllabi cover the courses offered at the 50 universities in the sample. To do so, I match the Spring 2026 syllabi to each university's published course schedule for the same term, and I measure coverage as the share of offered courses with at least one matched syllabus. I recover same-term course schedules for 40 of the 50 universities, and coverage at the median university reaches 72 percent (Table B6).

Beyond the level of coverage, I examine whether the syllabi I collect resemble the courses each university offers. In this context, I compute the correlation between the field composition of the syllabi and that of the offered courses in the same term, and the alignment is strong at most universities. As a complementary measure of collection intensity, I also relate the number of Spring 2026 syllabi at each university to its undergraduate enrollment, recovering about ten syllabi per hundred enrolled students at the median institution. This

ratio is similar for the ten universities whose course schedules I do not observe (9.2 versus 10.3 at the median), which is reassuring about the intensity of collection at these institutions. Taken together, these results indicate that the syllabus sample tracks the courses offered at each university, which is the relevant margin for studying instructional practices.

Table B6: Spring 2026 Syllabus Coverage and Field Alignment

University	Syllabi (Spring 2026)	Undergraduate Enrollment	Syllabi per 100 Enrolled	Offering Coverage (%)	Field Correlation
Texas					
East Texas A&M	1,717	9,034	19.0	82.8	0.97
Houston	3,287	42,050	7.8	79.7	0.99
Houston Clear Lake	1,424	7,592	18.8	97.8	0.98
Houston Downtown	1,951	15,483	12.6		
Lamar	2,343	9,546	24.5	98.0	0.84
North Texas	2,780	37,529	7.4		
North Texas at Dallas	335	3,760	8.9		
Prairie View A&M	1,872	8,831	21.2		
Sam Houston State	4,035	20,021	20.2	44.2	0.99
Stephen F. Austin	1,844	10,933	16.9	89.2	0.93
Tarleton State	1,766	14,650	12.1	68.5	0.95
Texas A&M	5,012	60,118	8.3	81.1	1.00
Texas A&M Corpus Christi	1,008	8,998	11.2	70.7	0.97
Texas A&M International	1,030	7,844	13.1	89.5	1.00
Texas A&M Kingsville	939	5,385	17.4	80.6	0.98
Texas A&M San Antonio	1,009	7,619	13.2		
Texas Southern	882	7,580	11.6	59.7	0.90
Texas State	2,614	36,613	7.1	57.5	0.97
Texas Tech	1,081	34,406	3.1	32.0	0.83
UT Arlington	2,834	37,854	7.5	69.9	0.73
UT Austin	3,273	43,718	7.5	45.8	0.96
UT Dallas	1,871	22,941	8.2	98.1	0.99
UT El Paso	1,498	23,509	6.4	58.5	0.98
UT Permian Basin	751	5,695	13.2	89.5	0.97
UT San Antonio	2,146	33,259	6.5	58.0	0.99
UT Tyler	775	8,231	9.4		
West Texas A&M	601	7,772	7.7		
Florida					
Central Florida	3,957	68,470	5.8		
Florida	1,708	38,113	4.5	35.3	0.83
Florida A&M	1,352	8,704	15.5	63.0	0.87
Florida Atlantic	1,948	29,440	6.6	73.8	0.91
Florida Gulf Coast	2,188	15,924	13.7	89.2	0.97
Florida International	4,072	56,786	7.2		
Florida State	4,147	36,242	11.4	88.4	0.99
North Florida	1,753	16,840	10.4	86.0	0.92
West Florida	1,304	11,464	11.4	88.1	0.97
Georgia					
Albany State	941	6,971	13.5	84.4	0.97
Fort Valley State	462	2,596	17.8	56.1	0.92
Georgia	2,205	33,422	6.6	40.3	0.98
Georgia College	1,122	5,863	19.1		
Georgia Southern	1,049	25,275	4.2	25.4	0.53
Georgia State	1,721	33,129	5.2	19.4	0.30
Kennesaw State	4,178	44,419	9.4	91.6	1.00
Indiana					
IU Bloomington	2,705	48,857	5.5	44.8	1.00
IU Indianapolis	1,261	20,055	6.3	40.6	0.73
Purdue	1,698	41,678	4.1	49.0	0.98
Purdue Northwest	991	9,413	10.5	93.2	0.99
Other States					
Arizona State	4,430	70,646	6.3	45.0	0.97
San Jose State	5,288	29,480	17.9	80.3	0.98
Western Kentucky	1,747	17,217	10.1	74.2	0.94

Note: Table B6 presents the number of analysis-sample syllabi in Spring 2026, undergraduate enrollment from IPEDS, and syllabi per hundred undergraduates for all 50 universities. Offering coverage is the percentage of eligible Spring 2026 courses, defined by institution, subject, and course number, matched to at least one syllabus in the analysis sample. Before matching, I exclude independent study and directed reading, thesis and dissertation registrations, internships and practica, study-abroad shells, private lessons, and administrative placeholders. Laboratories and recitations count toward their parent course only when the same source identifies that parent course. Field correlation is the Pearson correlation across the five macro-field shares (STEM, Humanities, Social Sciences, Business, and Other), comparing collected syllabi with offered courses. Blank cells indicate that I do not observe a usable Spring 2026 offering source for that university, not that coverage is zero.

Appendix C: Classification Pipeline Details

This appendix provides the complete coding guidance for each of the four classification measures, including all worked examples of affirmative and negative cases drawn from actual syllabus text (Tables C1 and C2). The classification models and the human raters in the validation exercise all received this same guidance.

Table C1: Policy and Adoption Questions: Full Classification Guidance

Has AI Policy

Question. Does the syllabus contain an explicit policy, rule, or guideline about student use of AI or generative AI tools (such as ChatGPT, Copilot, or similar tools) for coursework?

Yes if the syllabus states rules, permissions, prohibitions, guidelines, or expectations about whether and how students may use AI tools for coursework. The policy may come from the instructor or the university and may appear anywhere in the syllabus. It may permit, prohibit, restrict, or conditionally allow AI use.

(i) "AI tools such as ChatGPT may not be used in this course"
(ii) "Students are welcome to use AI tools for assignments"
(iii) "Per university policy, AI-generated content must be cited"
(iv) "Use of generative AI is prohibited unless otherwise stated"
(v) "You may use AI for brainstorming but not for final submissions"
(vi) "All work must be your own; AI-generated submissions will be treated as plagiarism"

No if the syllabus teaches AI as a course topic but says nothing about whether students may use AI tools for their own coursework, or contains only a generic academic integrity statement that does not specifically mention AI or generative AI tools, or has no AI mention.

Permits AI

Question. Does the syllabus explicitly state that students are permitted, allowed, or welcome to use AI tools for coursework in this course?

Yes if the syllabus contains an explicit statement permitting, allowing, welcoming, or approving student use of AI tools for coursework. The permission may appear in a policy section, assignment description, course overview, or anywhere in the syllabus. It can be broad or targeted to specific assignments.

(i) "Students are welcome to use AI tools in this course"
(ii) "AI tools such as ChatGPT may be used for assignments"
(iii) "You are free to use generative AI"
(iv) "Per university policy, AI use is permitted with proper citation"
(v) "AI is allowed for brainstorming and drafting"

No if the syllabus only prohibits or bans AI without any permission, discusses AI only as a topic to study without permitting student use, requires AI disclosure without explicitly saying AI is permitted, or has no AI mention.

Note: Table C1 presents the complete coding guidance for the two policy questions. I classify these questions from the full syllabus text.

Prompt Structure and Model Configuration. I submit each of the four classification questions independently, so that a model's response to one question does not shape its evaluation of the remaining questions (Haaland et al., 2025; Saltiel, 2026). For each question, I provide the model with the full text of the syllabus and instruct it to base its response only on what is explicitly stated in the document.

For each syllabus-question pair, I construct a user message that inserts the question text, the corresponding coding guidance from Tables C1 and C2, and the full syllabus text. In

Table C2: Integration Questions: Full Classification Guidance

Teaching Integration

Question. Is AI part of what this course teaches: its class sessions, topics, or course materials?

Yes if the course dedicates class time or course content to AI: a schedule, module, or session entry on an AI topic, a class activity, lecture, discussion, or demonstration involving AI, or an assigned reading, video, tutorial, or other course material about AI. One concrete line is enough: a single schedule or module entry naming an AI topic or activity counts.

(i) "Week 7: AI in Public Administration" (a schedule entry)
(ii) "AI Ethics Discussion" (a one-line module element)
(iii) "In class, we will use ChatGPT to critique a sample essay" (a class activity)
(iv) "Watch: How Large Language Models Work" (an assigned reading or video about AI)

No if AI appears only outside the taught content: only in graded-task requirements with no class time or course material on AI (that is assessment, not teaching), only in rules about student use, only in a learning objective or general statement that the course will involve AI, or "AI" that does not mean artificial intelligence. Statements about AI use inside the course's AI policy or rules section are policy, not course content, even when they mention class exercises.

(i) "AI Analysis Project (20%): use ChatGPT to draft a memo" (graded work only, no AI class content)
(ii) "You may use AI tools if you cite them" (a use rule)
(iii) "Generative AI will be integrated throughout this course" (a general statement)
(iv) "Appreciative Inquiry (AI) reflection due Friday" (not artificial intelligence)

Evaluation Integration

Question. Does at least one graded or assessed task require students to engage with AI?

Yes if a specific assignment, project, exam, or task requires AI in the student's own work: the task requires students to use an AI tool, the deliverable must include or build on AI output, or a required step has students produce, compare, critique, or revise AI content. Look for the requirement in the task itself. A one-line graded item whose title names AI counts, unless the task is clearly about AI as a topic rather than using it. Course objectives are not tasks.

(i) "Homework will require the use of AI tools for problem solving" (a task requiring AI use)
(ii) "AI Analysis Project (20%): use ChatGPT to draft a memo" (a named graded assignment)
(iii) "Assignment 4: submit the transcript of your ChatGPT conversation" (AI output in a graded deliverable)
(iv) "Graded practice (15%): motivational interviewing with ChatGPT" (a required interaction)
(v) "Assignment 2 (10%): generate a response with ChatGPT and critique its accuracy" (a required critique step)
(vi) "Final project: build and train a machine learning model" (a build requirement)
(vii) "AI Assignment (10%)" (a one-line graded item)

No if no assessed task requires AI: optional, invited, or extra-credit use, course-wide permission, citation or disclosure rules, and task-like statements inside the AI policy section. Graded work that only tests AI as subject matter, or "AI" not meaning artificial intelligence, is also No.

(i) "You may use AI for the research paper" (optional use, even on a specific task)
(ii) "If you use AI, include your prompts and its responses" (a disclosure rule)
(iii) "The midterm covers machine learning concepts" (AI as subject matter)

Note: Table C2 presents the complete coding guidance for the Teaching Integration and Evaluation Integration questions. I classify both questions from the full syllabus text with AI policy statements removed.

addition to a binary classification, I require each model to provide a confidence rating, a brief justification, and a direct quotation from the syllabus when the answer is affirmative. This allows me to assess both the classification and the quality of the model's reasoning. Table C3 presents the user message template verbatim.

Table C3: User Message Template for Syllabus Classification

USER MESSAGE TEMPLATE

QUESTION: {question text}

CODING GUIDANCE: {guidance text}

SYLLABUS TEXT: {full_text}

INSTRUCTIONS
1. Read the FULL syllabus text carefully.
2. Answer the question with Yes or No.
3. State your confidence: high, medium, or low.
4. Provide a 2–3 sentence justification for your answer.
5. If your answer is Yes, quote the MOST RELEVANT passage(s) from the syllabus as evidence. If No, write "N/A".

Return ONLY a single JSON object with exactly these keys:

```
{
  "answer": "yes" or "no",
  "confidence": "high" or "medium" or "low",
  "justification": "your explanation here",
  "evidence": "quoted text or N/A"
}
```

Note: Table C3 presents the user message template populated for each syllabus-question pair. Placeholders in braces are replaced with the specific question text, coding guidance, and full syllabus text.

I classify each question with three open-weight models run locally, Gemma 4-31B, GPT-Oss 120B, and Qwen 3.6-35B, and take the label to be the modal answer of the three. For each of the three models, I set the temperature to zero and fix the random seed, aiming to maximize deterministic outputs across repeated runs.[1] Before classifying the two integration questions, I identify the AI policy statements in each syllabus and remove them from the text. The integration questions therefore read the course content itself, its schedule, activities, readings, and assignments, rather than the policy language about whether students may use AI. The two policy questions, where this language is itself the outcome of interest, are classified on the full syllabus text.

Human Validation. I recruited four research assistants to independently classify a sample of 60 syllabi across the full set of four measures.[2] I drew the syllabi using stratified sampling

[1] I also set the presence penalty and frequency penalty to zero and top_p to one for all three models.

[2] The rating design assigns 20 core syllabi to all four raters and rotates the remaining 40 so that each rater

by AI mention status, institution, and classification profile to ensure representation across the distribution of instructor responses. Each rater received the complete coding guidance (Tables C1 and C2), a brief orientation document describing the classification task, and the full text of each syllabus in PDF format. As shown in Table 3, the raters, working independently, agree closely both with one another and with the modal classification, which comfortably passes the Alternative Annotator Test.

Cross-Model Agreement. I classify each question with three open-weight models and take the label to be the modal answer of the three. Table C4 reports the pairwise agreement rates for the four questions. The three models agree on the large majority of syllabi, with agreement highest for the policy question, which has clear textual markers, and somewhat lower for the permission question, which requires more interpretive judgment.

Table C4: Cross-Model Agreement on the Four Questions

	Gemma vs. GPT-Oss		Gemma vs. Qwen		GPT-Oss vs. Qwen	
	Agreement (1)	Cohen's κ (2)	Agreement (3)	Cohen's κ (4)	Agreement (5)	Cohen's κ (6)
AI Policy	99.1	0.982	99.1	0.981	98.8	0.976
AI Permitted	94.7	0.854	96.1	0.896	94.2	0.838
Teaching Integration	97.9	0.780	98.2	0.835	97.6	0.758
Evaluation Integration	98.4	0.617	98.4	0.719	98.7	0.661
Syllabi	185,154					

Note: Table C4 reports, for each pair of classifiers, the share of syllabi on which the two give the same answer for each question, in percent, and Cohen's κ. Syllabi that never mention AI are coded as not having each measure for every classifier and are included in the comparisons. Each pair is computed on the syllabi covered by both classifiers. The Syllabi row reports the pooled Fall 2024 to Spring 2026 sample.

Qualitative Categories. The classification described above establishes whether a course integrates AI into its teaching or its evaluation, but it is silent on what that integration actually involves. A course that devotes a module to neural networks and a course that asks students to draft essays with ChatGPT both count as integrating AI, yet they reflect very different classroom choices. To distinguish between them, I build a qualitative layer on top of the classification. The natural input for this exercise is the passage of syllabus text that the classifier quotes as evidence when it answers a question affirmatively, since working with the quoted passage isolates the exact language describing how AI enters the course, rather than the syllabus as a whole.

Given these passages, the sorting itself is deliberately simple. I define one fixed list of words and phrases per category, and a passage joins a category when it contains at least one term from that category's list. The list for AI Methods, for instance, includes machine learning, neural networks, and deep learning, while the list for Using AI Tools includes use AI and AI tools. A passage can join several categories, and passages that match no list remain in a residual group. As such, no model judgment enters this step.

classifies 50 syllabi in total.

On the teaching side, the categories describe what the AI content of a course is about, distinguishing courses that teach the technology itself from courses in which students work with AI tools and courses that apply AI to the course's own subject. On the evaluation side, they instead describe what students are asked to do with AI in graded work, whether building AI models, engaging with AI output, or producing their own assignments with AI assistance (Table A3). In this way, the categories translate a binary signal of integration into a description of the choices instructors are actually making.

The lists must sort passages the way a careful reader would, and I verify this in a blind audit of teaching passages, checking that the category assigned by the lists matches the one a reader would choose after reading each passage in full. Moreover, the terms most distinctive of each category's passages line up closely with what each category is meant to capture (Figure A2). Altogether, these exercises establish that the passages are sorted sensibly and that the categories are distinct in their language.