Openness to AI in Higher Education*

Fernando Saltiel†

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

Please click here for the latest version.

Abstract

In this paper, I examine which university instructors are integrating generative AI into their classrooms. I construct a dataset of 148,200 course syllabi across 25 public universities in four U.S. states, linked to detailed instructor characteristics. I use a combination of open-weight large language models to classify each syllabus along dimensions of AI policy and concrete integration into teaching and evaluation, and I validate the measures against four independent human raters who agree on 95 percent of classifications. In Spring 2026, about 60 percent of syllabi contain an explicit AI policy, yet only about 6 percent show any classroom integration. Moreover, a variance decomposition reveals that instructors and the courses they teach, rather than institutions, drive classroom integration. I find that assistant professors and instructors with more recent publications are more likely to integrate AI into the classroom, with effects concentrated on the teaching rather than the evaluation margin. Altogether, these patterns are sharpest in STEM fields and at selective universities, where young, research-active faculty are leading the diffusion of AI into higher education classrooms.

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

†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; Humlum and Vestergaard, 2025; Dillon et al., 2025). These technologies are reshaping the task content of occupations and the returns to specific skills (Acemoglu and Restrepo, 2019; Felten et al., 2023; Eloundou et al., 2024). In this context, instructors at universities face real-time decisions about whether and how to incorporate AI into their courses, with potentially far-reaching implications for the human capital that students develop in college (Frankel and Kartik, 2022). That is, instructors serve as the primary agents of curricular chage, as they choose whether to address AI in their courses, whether to encourage or restrict its use, and whether to redesign their teaching and evaluation around these tools.

In this paper, I examine how university instructors have responded to generative AI in their course syllabi, and I analyze which professors are integrating these tools into the classroom. I construct a dataset of 148,200 course syllabi covering more than 60,000 unique courses taught by nearly 50,000 instructors across 25 public universities in Texas, Florida, Georgia, and Indiana for the 2024 through 2026 academic years.[1] Moreover, I link each syllabus to detailed instructor, course, major, and institutional characteristics from an auxiliary dataset that I assemble from various administrative data sources. This combination of syllabus text and linked instructor data allows me to move beyond documenting whether AI appears in syllabi and toward understanding who is shaping the integration of AI in higher education. I rely on course syllabi for this analysis, as they capture what professors actually communicate to students about AI, offering a revealed-preference measure that complements the stated-preference evidence from faculty surveys (Biasi and Ma, 2025).[2] To classify the information in these syllabi, I use open-weight large language models that recover richer dimensions of AI integration than keyword detection can provide.

I construct this dataset by collecting publicly available syllabi across four states, all of which require public universities to make course syllabi accessible online. The 25 universities range from selective research universities such as UT-Austin and the University of Georgia to broad-access institutions such as Florida A&M and West Texas A&M. For each syllabus, I recover the instructor's name, course title, department, and academic term, which together allow me to link each syllabus to a specific instructor, department, and course. I complement the syllabus data with instructor characteristics drawn from state salary databases, bibliometric records, and faculty directories, which allow me to observe each instructor's academic rank, gender, publication history, and career stage.[3] I similarly

[1] These syllabi are publicly available under state-level requirements that mandate online access to course syllabi at public universities.

[2] Recent work has shown that instructor policies materially shape student AI usage (Carvajal et al., 2024; Contractor and Reyes, 2025).

[3] Appendix B.3 describes the data sources, matching procedures, and priority rules used to construct the instructor-level characteristics. I recover gender and academic rank for 83 and 68 percent of the instructors in the sample, respectively.

recover institutional characteristics from IPEDS and the Mobility Report Cards of Chetty et al. (2020), along with major-level measures of mid-career earnings, AI exposure, and student composition across fields.

As a first approximation, I identify syllabi that mention artificial intelligence using a keyword-based approach. In the Spring 2026 cross-section, 62 percent of syllabi mention AI, up from essentially zero before the release of ChatGPT in November 2022.[4] The prevalence of AI mentions varies considerably across disciplines, with close to 80 percent of courses in Sociology mentioning AI compared to about half in Economics. I also find significant variation across universities, with syllabi at broad-access institutions more likely to mention AI than those at selective universities. Yet this binary indicator collapses fundamentally different instructor responses into a single measure, as a syllabus that prohibits AI use and one that builds a course around it both trigger the same keyword detection.

To recover what instructors actually do with AI in their classrooms, I develop a classification framework that captures multiple dimensions of the instructor response. I measure whether syllabi contain explicit AI policy language, whether they explicitly permit AI use, whether AI has been integrated into teaching through scheduled topics or class activities, and whether AI has been incorporated into graded evaluation. These distinctions capture qualitatively different levels of engagement. A policy statement about AI is nearly costless to produce, but an instructor who schedules a session on prompt engineering or requires students to use ChatGPT for a project is making a substantially different choice. I additionally separate institutional template language from instructor-authored text, which allows me to isolate the instructor's own disposition toward AI from the broader institutional policy environment. I validate these measures by having research assistants independently classify a stratified sample of syllabi across the full set of questions, following the approach recommended by Haaland et al. (2025). The four raters agree on 95 percent of classifications on average.

While the human raters establish a reliable benchmark, classifying close to 150,000 syllabi by hand would be prohibitively time-consuming. I therefore implement the classification using open-weight large language models with fixed, publicly available parameters (Korinek, 2023). I classify each syllabus using two independent models and employ a third model from a different family as a judge to resolve disagreements, following the LLM-as-judge framework of Zheng et al. (2024). The automated classification agrees with the modal human response in 95 percent of cases and passes the Alternative Annotator Test of Calderon et al. (2025), achieving a win-plus-tie rate of 96 percent.

I find that close to 60 percent of courses in Spring 2026 include an explicit AI policy in their syllabi, largely in line with the keyword-based measure discussed above. At the same time, only about one quarter of courses explicitly permit students to use AI, suggesting that most policies in place are not affirmatively endorsing student use of these tools. Direct AI integration into teaching is substantially rarer, as only about 5 percent of syllabi include concrete evidence that AI has been incorporated into course content. Changes to

[4] Using the longitudinal sample from UT Austin and UT Dallas, I find that the share of syllabi mentioning AI increased sharply in the semesters following the release of ChatGPT.

how students are evaluated remain the rarest form of classroom adoption, with fewer than 2 percent of syllabi incorporating AI into graded coursework. Altogether, about 6 percent of syllabi show any classroom integration, whether through teaching or evaluation.

I find significant differences in the extent to which AI has been integrated into the classroom across college majors. Business-oriented and applied fields lead by a wide margin, with about 20 percent of courses in Marketing and Information Systems having integrated AI, compared with roughly one percent in Music, Dance, and Geology. These differences broadly track occupational AI exposure, as the fields whose graduates are most likely to use generative AI at work are also where instructors are most likely to integrate it into the classroom (Felten et al., 2023; Bick et al., 2025). At the same time, majors in the social sciences and humanities are more likely to include AI policy language without comparable classroom integration. The policy response and the pedagogical response have not moved in lockstep across disciplines.

The differences across universities are similarly pronounced. Selective universities in the sample exhibit roughly 10 percent classroom integration, about double the rate at moderately selective and broad-access institutions. I find a related pattern across course levels, as graduate courses display less AI policy language but more classroom integration than undergraduate courses. These patterns are consistent with a growing body of evidence that AI adoption concentrates at more research-intensive organizations (McElheran et al., 2024; Yotzov et al., 2025), and they suggest that a similar gradient is emerging in higher education.

To decompose the variation in AI integration, I implement a Shapley-Owen variance decomposition that attributes variation to instructor, local course, university, major, and academic term components. I find that instructors account for roughly 25 to 30 percent of the variation in whether AI is actually integrated into the classroom. Courses play a similarly important role in shaping these outcomes, and I find that graduate courses and those with prior assessment structures involving take-home assignments are more likely to integrate AI into teaching (Biasi and Ma, 2025). By contrast, the university component is negligible for classroom integration, explaining only 0.3 percentage points of the variance, even though it accounts for about 7 percentage points of the variance in AI policy. Altogether, the results indicate that it is instructors and the courses they teach, rather than institution-wide directives, that drive whether AI is actually incorporated into the classroom.

I examine which observable instructor characteristics predict classroom AI integration, conditional on local course and university-by-major-by-level-by-term fixed effects. I find that assistant professors are more likely than full professors to have integrated AI into their courses, and that instructors who have published more recently are similarly more likely to do so.[5] These associations remain stable once I include the full set of instructor covariates simultaneously. They are also robust to controlling for text-based syllabus characteristics and to replacing local course fixed effects with course-by-year fixed effects. The effects are concentrated on the teaching margin, as assistant professors are more likely to schedule

[5]On the policy margin, the patterns look quite different. Women are substantially more likely than men to include AI policy language in their syllabi, yet this gender gap largely disappears once I turn to classroom integration. The forces shaping who addresses AI formally and who teaches with it are not the same.

AI topics and assign AI-related readings but are no more likely to redesign graded assessments around these tools. Measures of cumulative research output, including h-index and total publication count, add essentially no predictive power once I control for the core covariates. Instead, what appears to matter is not the volume of an instructor's research but rather proximity to active engagement with new methods. These instructor-level effects are sharpest in STEM fields and at selective universities, where assistant professors are about 2 percentage points more likely than full professors to have integrated AI into the classroom. This finding fits in with Biasi and Ma (2022), who show that research-active instructors teach courses significantly closer to the knowledge frontier, with assistant professors exhibiting the lowest education-innovation gap across academic ranks. In that sense, it is young, research-active faculty at the most research-intensive universities who are leading the diffusion of AI into higher education classrooms.

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). Recent evidence shows that generative AI is diffusing quickly on the extensive margin, yet that realized productivity gains depend on complementary organizational changes and task-level redesign (David, 1990; Brynjolfsson and Hitt, 2000). Moreover, adoption is strongly heterogeneous across organizations and individuals, fitting in with canonical evidence on technology diffusion (Griliches, 1957; Hall and Khan, 2003; Comin and Hobijn, 2010). I contribute to this literature by documenting how AI is being adopted in higher education, a large sector in which instructors face real-time decisions about whether and how to incorporate these tools into their professional practice.

I further contribute to a growing literature examining how universities, instructors, and students are responding to artificial intelligence in educational settings. Prior work has documented faculty attitudes using surveys (Stohr et al., 2024; Ravselj et al., 2025), analyzed institutional AI policy statements (McDonald et al., 2025), and shown that instructor policies materially shape student AI usage, with heterogeneous effects across subgroups (Carvajal et al., 2024; Contractor and Reyes, 2025; Cruces et al., 2026). Fitting in with this literature, Chirikov (2026) examines how instructors regulate AI use in their syllabi and documents sizable within-institution variation. I extend this body of work by linking syllabi to detailed instructor characteristics, which allows me to formally decompose the sources of variation in AI integration and to show that gender predicts AI policy adoption, whereas assistant-professor status and publication recency predict classroom integration.

This paper further fits in with an extensive literature on instructor heterogeneity and higher education production. An influential body of work has shown that instructor effects are sizable and consequential for student outcomes across both short- and longer-run margins (Hoffmann and Oreopoulos, 2009; Carrell and West, 2010; Braga et al., 2016; De Vlieger et al., 2016; Feld et al., 2020). Moreover, instructor employment type and career stage shape classroom practices and student pathways (Bettinger and Long, 2004, 2010; Carrell et al.,

2010; Fairlie et al., 2014; Figlio et al., 2015; Chen et al., 2021). Most closely related to my paper is the pioneering work of Biasi and Ma (2025), who construct a dataset of 2.2 million syllabi and show that instructor and course assignment processes are the primary drivers of curricular content variation. They further document large within-institution dispersion and sizable effects on student outcomes.[6] I extend their approach by applying the syllabi-as-data framework and Shapley-Owen variance decomposition to AI policy and classroom-integration outcomes across 25 universities. I find that instructor and course components account for the vast majority of explained variation in classroom AI integration, while the university component is large for policy but negligible for classroom measures. These results indicate that institutions set the stage for whether AI is formally addressed, yet instructors choose whether and how to respond.

Lastly, I broadly contribute to a growing literature using text-as-data methods in economics (Baker et al., 2016; Gentzkow et al., 2019; Ash and Hansen, 2023; Dell, 2025). A recent wave of work has shown that large language models can classify text with accuracy comparable to human raters across a range of economic applications (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). At the same time, concerns have emerged regarding the sensitivity of LLM-based classifications to model choice and prompt phrasing (Ludwig et al., 2025). I contribute to this methodological literature by taking advantage of open-weight large language models to recover nuanced information about AI policy stances from unstructured syllabus text. The classification captures dimensions that go beyond what keyword detection or dictionary approaches can achieve. These include distinguishing institutional policy language from instructor-authored text and separating explicit AI permission from concrete integration into teaching and evaluation. Moreover, I introduce a boilerplate decomposition that separates institutional template language from instructor-authored text.[7]

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. Section 5 examines heterogeneity across universities and majors. Section 6 studies the drivers of AI integration, including the decomposition and course-level patterns. Section 7 examines who integrates AI into the classroom. Section 8 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

[6] See also Biasi and Ma (2022) for earlier work on the education-innovation gap and the role of instructor research activity in shaping syllabus content.

[7] The open-weight models ensure reproducibility, as all model parameters are fixed and publicly available, and the classification can be conducted without sharing data with external providers (Korinek, 2023).

include detailed information on course content, topics covered, specific assignments and course policies, which allow me to observe how instructors incorporate AI into their classes (Biasi and Ma, 2022). Following the rapid rollout of generative AI tools in higher education, I assemble a new dataset of publicly available syllabi from 25 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 across four states (Texas, Florida, Georgia, and Indiana), all of which have implemented policies that require public universities to make course syllabi publicly accessible.[8] I recover course syllabi for 25 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. The state-level disclosure requirements allow me to recover the syllabi for a large share of courses offered at these institutions. As such, the sample covers a broad range of academic disciplines and course levels, ranging from introductory undergraduate courses to graduate seminars. For UT-Austin and UT-Dallas, I additionally recover syllabi dating back to 2015–2016, which allow me to analyze outcomes before and after the rollout of ChatGPT.

For each syllabus, I recover a set of core course characteristics through a multi-step extraction approach that combines structured information recorded in each university's data system with information drawn from the syllabus text itself.[9] The core characteristics recovered from the syllabi include the instructor's name, the course title, department and course number, and the academic term in which the course is offered. Together, these fields allow me to link each syllabus to a specific instructor, academic department, and course. I keep one syllabus for a course-instructor combination in each academic term, so that each observation in the dataset corresponds to a unique course taught by a specific instructor in a given semester.

The main estimation sample includes syllabi for the courses offered between Fall 2024 and Spring 2026. I focus on these four academic terms given the recent rollout of state-level disclosure laws in Indiana and Georgia. Altogether, I recover 148,218 unique syllabi covering 60,202 unique courses taught by 49,834 different instructors in the 25 universities in the main sample, and 34 percent of these syllabi are from classes offered in Spring 2026 (Table A1).[10]

Beyond the main course characteristics recovered from the syllabi, I extract a set of text-

[8] The disclosure requirements vary across states in terms of how many years of syllabi universities must maintain and how they are made available. Appendix B.1.1 describes the policy context and collection approach for each university.

[9] For a small share of syllabi where the text-based extraction does not recover the instructor name or course title, I use a local large language model to extract these characteristics. Appendix B.1.2 describes the full extraction procedure.

[10] I additionally collect 55,800 and 29,400 syllabi for UT Austin and UT Dallas between 2015 and 2024, respectively.

based features from each syllabus that capture relevant information about the course and the instructor. For instance, these measures include the document's length and writing complexity, as well as the prescriptiveness of the instructor's language, which vary considerably across institutions and disciplines (Table A2). I additionally recover whether the syllabus includes specific policies about late-work or attendance requirements, and whether the class has a teaching assistant.[11]

As a first approximation to understanding how AI is being integrated into the classroom, I identify syllabi that include AI-related content using a keyword-based text analysis approach. In this context, the keyword set includes terms that cover broader engagement with AI, such as artificial intelligence, generative AI, and large language model, along with specific terms like ChatGPT.[12] For each syllabus with at least one keyword match, I additionally recover the surrounding text, which provides context on how AI is referenced in each course. This allows me to provide an initial description of the general prevalence of AI-related content across courses, disciplines, and institutions in the sample.

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 that I recover from multiple data sources. Specifically, I draw on publicly available salary records, bibliometric databases, instructors' CVs, faculty directories, and institutional profiles to construct a detailed instructor-level dataset capturing information on instructors' backgrounds and career trajectories. These data allow me to understand how instructors' characteristics, covering their academic rank, gender, research productivity, and career experience, among others, drive AI adoption outcomes in higher education.[13]

To recover information on instructors' research productivity, I first match the instructor roster to OpenAlex, an open-access bibliometric database (Priem et al., 2022). I match 35 percent of instructors, ranging from 12 percent at Stephen F. Austin to 57 percent at Purdue (Table A3), reflecting institutional differences in the composition of their faculty members. From the matched profiles, I recover their publication counts and dates, the number of citations, and their h-index, which I use to construct measures of research productivity and impact. I also capture information about instructors' educational backgrounds and career trajectories by drawing on ORCID education records, publicly available faculty CVs, and departmental profile pages.[14] The combination of these data sources allows me to measure whether professors have any publications, along with information on their doctoral training, including their PhD institution and completion year.[15]

I also collect information on instructors' academic rank from state salary databases,

[11] In Appendix B.2.1, I describe how I construct the full set of textual characteristics from the syllabi in the sample.

[12] The full keyword list and contextual filtering rules are described in Appendix B.2.2.

[13] Appendix B.3 describes the full set of data sources, matching procedures, and the priority rules that I use to construct this complementary dataset of instructor-level characteristics.

[14] Appendix B.3 describes each data source used in this setting in detail.

[15] I recover PhD completion years and institutions for 18 percent and 24 percent of instructors, respectively.

which record each employee's official institutional title across all four states in the sample. I classify the instructors into tenure-track positions, distinguishing among full, associate, and assistant professors, and non-tenure-track faculty, which includes lecturers, adjuncts, and instructors of practice. I also consider instructors' gender, by drawing on records from the Texas salary database, along with a name-based probabilistic approach that uses Social Security Administration baby-name frequency data for the other instructors. When multiple data sources provide overlapping information for a given instructor, I prioritize official administrative records over self-reported and text-extracted information (Appendix B.3.2). Altogether, I recover information on instructors' gender and on their academic rank for 83 and 68 percent of the sample, respectively. Table A3 presents the full set of instructor characteristics and their availability across the 25 universities in the sample.

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 combine information on institutional characteristics from multiple data sources (Appendix B.4). First, I use data from the Integrated Postsecondary Education Data System (IPEDS) to capture measures of institutional size, research intensity, faculty compensation, and average SAT scores for each of the 25 universities in the sample. I also draw on the Mobility Report Cards from Chetty et al. (2020) to recover the median parental income for each institution. Moreover, I classify the 25 universities in the sample into three tiers of institutional selectivity based on Barron's Admissions Competitiveness Index, from selective research universities to broad-access institutions. Table A4 presents these characteristics for each university, showing sizable differences in research resources and faculty compensation across the three selectivity tiers.

I similarly recover a set of major-level characteristics that capture the labor market conditions, academic orientation, and student composition across fields. In particular, I capture the mid-career median earnings and unemployment rate for each major from the Federal Reserve Bank of New York's College Labor Market data (Abel et al., 2014). I also recover the gender composition and the share of graduates who go on to earn a graduate degree. I also construct a measure of average SAT scores in each major, using information from the National Postsecondary Student Aid Study (Appendix B.5.1). 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 compared to those in Nursing.

Since the 25 universities in the main sample use more than 1,500 distinct department codes, I harmonize them into 42 majors following Biasi and Ma (2025).[16] For instance, this step classifies courses from departments in areas such as Neuroscience, Microbiology, and Biochemistry into a harmonized Biology major (Table B11). To better understand heterogeneous AI integration across disciplines, I further aggregate the 42 majors into five macro-

[16] Appendix B.5.2 describes the mapping from department codes to the 42 harmonized categories in detail.

fields, covering STEM, Humanities, Social Sciences, Business, and Other. Table A5 presents these characteristics for the harmonized majors in the sample, showing significant heterogeneity in AI exposure, mid-career earnings and average SAT scores across these fields.

I also recover the academic level of each course from the course numbering conventions used by each university. Since nearly all universities in the sample follow a standard system in which the leading digit of the course number indicates the year of study, I can identify whether courses are designated as freshman through senior level, or as graduate courses (Appendix B.6.1). I use this five course-level classification as the main variable, and also classify courses as basic undergraduate, advanced undergraduate, or graduate following Biasi and Ma (2022). I also take advantage of the Florida Statewide Course Numbering System, which provides harmonized course identifiers across universities in Florida. This allows me to analyze the differences in syllabi for the same course taught at different institutions in Florida.

Representativeness. I validate the representativeness of the syllabus corpus by comparing the distribution of syllabi across academic fields against external benchmarks drawn from registrar data, state enrollment records, and federal degree-completion statistics. For the universities where I have access to direct section-level data, the department-level distribution of syllabi closely tracks the distribution of actual course sections, confirming that the corpus reflects the underlying composition of course offerings at each institution (Appendix B.8).

2.4 Descriptive Analysis

Altogether, the estimation sample includes 148,000 unique syllabi covering over 60,000 unique courses across the 25 universities. To assess whether these syllabi are representative of the courses offered at these institutions, I compare the distribution of syllabi across departments against the distribution of course sections at each university. I find that the two distributions closely track each other across the universities where external benchmarks are available (Table B19), indicating that the syllabi in the estimation sample are broadly representative of course offerings at these institutions.[17]

Table 1 presents descriptive statistics for the main sample, focusing on the Spring 2026 cross-section across the 25 universities. The text-based analysis finds that 62 percent of the syllabi in this semester mention artificial intelligence. Half of the syllabi are taught by tenure-track faculty, while only 37 percent are taught by instructors who have published at least one academic paper, consistent with the composition of the instructor workforce at these institutions.

The next two columns in Table 1 split the sample by whether the syllabus mentions artificial intelligence.[18] Syllabi that mention AI tend to be considerably longer, averaging close

[17] I describe the representativeness analysis in detail in Appendix B.8. I compare the field-level distribution of syllabi against external benchmarks drawn from registrar and schedule data for the eight universities where such data are available.

[18] For this comparison, I focus on syllabi in the Spring 2026 semester. Table A6 presents the corresponding evidence for the full 2024–2026 period.

to 1,000 more words than those that do not. Moreover, syllabi that mention AI are 10 percentage points more likely to be taught by a woman, while the share of tenure-track faculty and those with publications is effectively identical across the two groups. Undergraduate courses are more likely to mention AI than graduate courses, and syllabi at broad access universities are overrepresented among those that discuss AI.

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 of syllabi in Economics. These patterns suggest that the prevalence of AI in the syllabi is shaped more by discipline-specific norms than by the broader field of study. I similarly find significant variation across the universities in the sample that holds even within selectivity levels (Figure A1).

To further understand the extent to which AI has been integrated into these syllabi, I draw on the longitudinal sample of syllabi for UT Austin and UT Dallas and present the evolution of AI mentions since the 2019–2020 academic year. Figure 1 (Panel b) shows that a small share of syllabi already mentioned 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 universities. By the Spring of 2026, 64 percent of syllabi at UT Austin and 44 percent at UT Dallas mention artificial intelligence.[19]

While this analysis establishes that AI-related content has become pervasive in college syllabi, this approach does not distinguish between syllabi that encourage AI use and those that restrict it. Two instructors who both mention AI in their syllabi may hold fundamentally different policy stances toward its use in the classroom. I address this limitation in the next section by introducing a classification framework that uses open-weight large language models to extract richer dimensions of AI policy from the syllabus text.

3 Measuring AI Integration in the Classroom

3.1 AI Integration into Higher Education

The descriptive analysis in Section 2 reveals that over half of syllabi now mention artificial intelligence, yet this binary indicator collapses fundamentally different instructor responses into a single measure. 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 trigger the same keyword detection, yet they represent diametrically opposed pedagogical stances. On the student side, recent evidence indicates that explicit prohibition of AI tools substantially reduces intended AI use, suggesting that these policy distinctions have real consequences for student behavior (Contractor and Reyes, 2025). The same limitation applies to the institutional policy margin. Many AI policies are boil-

[19] This difference across the two universities is consistent with the broader cross-sectional patterns in Figure A1.

erplate text inserted at institutional direction, and the near-universality of AI mention at some institutions suggests institutional diffusion of template language rather than genuine pedagogical engagement (McDonald et al., 2025; Chirikov, 2026). As such, moving beyond binary detection to recover what instructors actually *do* with AI in their classrooms is a first-order empirical concern.

Instructors play a central role in shaping the human capital that students acquire in higher education, and whether and how they choose to integrate AI into their classrooms is a first-order question about what the educational system produces. Importantly, integrating AI into a course is not the same as keeping course content up to date with advances in the instructor's own field, a process that flows naturally from research activity. It is instead a different kind of decision, one that requires deliberate pedagogical redesign around a general-purpose technology that cuts across fields (Bresnahan and Trajtenberg, 1995). A history professor who redesigns essay assignments around AI critique is making a pedagogical choice unrelated to the history research frontier, and a biology instructor who incorporates AI-based exercises is adapting a technology that is not specific to biology.

The first dimension I consider is whether AI has entered what instructors actually *teach*. A policy statement permitting or discouraging AI is nearly costless to produce, but an instructor who schedules a session on prompt engineering, assigns readings about AI applications in their field, or designs a hands-on AI exercise is making a qualitatively different choice. A typical undergraduate semester offers only 14 to 16 weeks of instructional time, and allocating any of these scarce sessions to an AI-related topic comes at the direct cost of displacing traditional content. These are costly, observable decisions that reveal the instructor's judgment about what knowledge students need (Comin and Hobijn, 2010; Bloom et al., 2025). The strength of this motivation varies across the curriculum, as not all courses face the same technological frontier (Eloundou et al., 2024; Dell'Acqua et al., 2025).[20] As such, teaching integration is likely to concentrate among certain disciplines and instructor types, with prevalence rates well below those of simple AI mention.

The second dimension captures whether AI has entered how instructors *assess* students. A named graded assignment requiring ChatGPT, or a requirement to submit AI conversation transcripts, represents a reconception of what the assessment measures. This is not primarily about whether instructors have changed how they monitor or detect student AI use. Rather, these assignments shift the assessed quantity from "can you produce good output unaided" to "can you evaluate and improve AI-generated output," making the student's analytical contribution visible and separately assessable (Frankel and Kartik, 2022).[21] Evaluation redesign is the costliest form of integration because it requires rethinking what the assignment measures, how it is graded, and what constitutes successful student performance. An assignment where students must critique, revise, or compare their own work against AI-generated output requires the instructor to reconceive not only the task but also

[20] Dell'Acqua et al. (2025) show that AI capabilities are uneven across tasks in a field experiment with management consultants, with AI users performing substantially *worse* on tasks outside the AI capability frontier.

[21] Assignments requiring students to evaluate AI-generated content assess a meta-cognitive skill that is increasingly valued in the labor market (Noy and Zhang, 2023; Brynjolfsson et al., 2025).

the rubric by which it is evaluated. This amounts to creating new tasks within the educational process that did not exist before AI (Acemoglu and Restrepo, 2019). Evaluation integration should therefore be the rarest form of classroom adaptation to AI.

3.2 Measuring AI Integration into the Classroom

Following the approach recommended by Haaland et al. (2025) for incorporating qualitative analysis into automated text classification, I work with research assistants who closely read a sample of AI-mentioning syllabi to catalog the full range of instructor responses to AI. This exploratory phase confirms that instructor responses organize along the two dimensions motivating the classification, with the classroom actions dimension further decomposing into teaching content and evaluation design. I construct a set of binary measures organized along these domains, with attitude measures applied to instructor-authored text remaining after removing institutional boilerplate and classroom action measures applied to the full syllabus text.[22]

Before examining how instructors integrate AI into teaching and evaluation, I first capture whether the instructor addresses AI at all and what stance they take toward its use. I consider whether the syllabus contains any explicit policy regarding student use of AI tools, which serves as the extensive-margin gate. I additionally consider whether the instructor goes further and explicitly permits AI use, going beyond the presence of any policy to recover the instructor's affirmative stance toward adoption. The gap between these two measures is itself informative, as an instructor who addresses AI but does not permit its use is making a qualitatively different choice than one who explicitly welcomes it.

The teaching domain captures whether AI has entered course content through three measures applied to the full syllabus text. I first consider whether the course schedule dedicates at least one session, week, or module to an AI-related topic. Scheduling an AI topic represents a direct allocation of the instructor's most constrained resource, fitting in with the revealed-preference logic developed in Section 3.1. I additionally examine whether the syllabus describes an activity, demonstration, or exercise involving AI during scheduled class time. This is a higher-cost commitment than scheduling a topic, as it requires the instructor to manage real-time classroom interaction with AI tools. I also capture whether the instructor assigns readings, videos, tutorials, or other instructional materials specifically about AI, signaling that the instructor views AI literacy as knowledge worth developing (Javadian Sabet et al., 2024). These three measures range from the most modular form of teaching integration (assigning a reading) to the most resource-intensive (designing a hands-on exercise), consistent with evidence that higher education responds to shifts in the skill content of labor demand (Conzelmann et al., 2023; Javadian Sabet et al., 2024).

The evaluation domain captures whether AI has been incorporated into graded coursework through three measures. I first consider whether the syllabus describes a specific, named assignment that explicitly requires students to use an AI tool as part of the graded

[22]The boilerplate decomposition uses document-level clustering to identify sentences common across multiple syllabi within an institution. After removing these shared sentences, the residual text captures instructor-specific language about AI. Details and validation appear in Appendix B.2.2.

deliverable. I additionally capture whether any assignment requires students to submit AI-generated output, such as transcripts, screenshots, or generated text, as part of the graded submission. I also examine whether students must engage with AI-generated content as a structured step in the assignment workflow, by evaluating, comparing, critiquing, or revising AI output. This last form of engagement captures the most intensive evaluation redesign, as it requires students not only to use AI but to exercise judgment over its output. Together with the teaching measures, these evaluation measures allow me to characterize not just whether AI has entered the classroom but along which margin.

To capture these dimensions from syllabus text, I construct the measures as binary questions, following guidance in the psychometrics literature showing that binary response formats yield higher inter-rater reliability compared to multi-category scales (Dolnicar et al., 2011). I additionally include brief coding guidance for each question that clarifies the classification criteria, following standard practice in content analysis (Lombard et al., 2002; O'Connor and Joffe, 2020). The guidance was iteratively refined through multiple rounds of human review, incorporating concrete examples of both affirmative and negative cases drawn from actual syllabus text. Table 2 presents the eight classification questions with their accompanying guidance.

I incorporate human raters to establish the validity of these constructs before deploying any automated classification. In particular, four research assistants independently classify a sample of 60 syllabi across the full set of eight measures. I draw the syllabi using stratified sampling by AI mention status, institution, and classification profile to ensure representation across the distribution of instructor responses.[23] This overlapping design allows me to assess inter-rater reliability across the full set of questions (Krippendorff, 2004).

Table C10 presents the inter-rater reliability results. I find that the four raters agree on 95 percent of classifications on average across the eight questions. Agreement is higher for questions with clear textual markers and somewhat lower for questions requiring more interpretive judgment, with the full breakdown reported in Appendix C.4. These results provide a reliable benchmark for the subsequent LLM-based classification (Dell, 2025; Haaland et al., 2025).

3.3 LLM Classification Implementation

While the human raters described in Section 3.2 agree on 95 percent of classifications, coding close to 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). This is particularly relevant in this setting, as the same keyword appears in syllabi that prohibit AI use, permit it, or redesign entire courses around it. A growing literature has shown that large language models can classify text with accuracy comparable to human raters across a range of economic applications (Bartik et al.,

[23]The rating design assigns 20 core syllabi to all four raters and rotates the remaining 40 so that each rater classifies 50 syllabi total.

2024; Braghieri et al., 2024; Lagakos et al., 2025; Ash et al., 2025).

On the other hand, recent work has raised concerns about the sensitivity of LLM-based classifications to model choice and prompt phrasing (Ludwig et al., 2025). Proprietary models pose additional challenges, as providers may update or deprecate their systems without notice.[24] I implement the classification using open-weight models with fixed, published parameters, which can be run locally without sharing data with external providers (Korinek, 2023). I classify each syllabus using two independent primary models and employ a third model as a judge to resolve disagreements, following the LLM-as-judge framework of Zheng et al. (2024). 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 API calls for each of the eight classification questions, so that a model's response to one question does not shape its evaluation of the remaining questions (Haaland et al., 2025; Saltiel, 2026). In addition to requiring a binary Yes/No 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. The full prompt structure is documented in Appendix C.1.

I classify each question using two open-weight primary models, GPT-Oss 120B and Qwen 3.5-35B, which rank among the top-performing open-weight models on standard evaluation leaderboards (Chiang et al., 2024). 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, aiming to maximize deterministic outputs.[25] Each primary model independently classifies all syllabi, providing two complete sets of classifications. For the six teaching and evaluation questions, I use a revised system prompt that guards against confusing AI policy language with evidence of classroom integration.[26] Full prompt structures, model configurations, and random seeds are documented in Appendix C.

I additionally implement a third open-weight model, GLM-4.7-Flash, as a judge that serves two roles in the classification pipeline (Zheng et al., 2024). The judge belongs to a different model family than the two primary classifiers, which guards against the self-preference bias whereby LLM judges systematically favor outputs from models within their own family (Zheng et al., 2024; Wataoka et al., 2025). First, when the two primary classifiers disagree, the judge receives both models' justifications and evidence and makes an independent determination.[27] Second, when both primary classifiers agree that a teaching or evaluation measure is present, the judge selects the single most compelling evidence passage from the syllabus text, drawing on both models' suggested evidence. This evidence

[24] Barrie et al. (2024) document that submitting the same prompt to GPT-4o yields results that vary month-to-month, even absent publicly announced model changes.

[25] The full set of inference parameters for all three models is documented in Appendix C.1.

[26] The two policy questions use the original system prompt, since for these questions AI policy language is the relevant evidence. Appendix C presents both prompt variants verbatim.

[27] I anonymize the primary model identities and randomize their ordering across cases to prevent positional bias (Zheng et al., 2024). The adjudication prompt differs slightly across the two question domains. Appendix C presents both variants verbatim.

selection step is important because the curated passages serve as the input for the qualitative analysis of how instructors integrate AI into the classroom, as described in Section 4.[28] When both primary classifiers agree that a measure is absent, the classification is recorded without invoking the judge.

I find that the two primary classifiers largely agree. Among syllabi sent to the classifiers, the two models reach the same classification in 98 percent of cases, with the judge adjudicating the remaining 2 percent (Table C9). The judge rules the measure present in approximately half of adjudicated cases and absent in the other half for the six teaching and evaluation questions (Table C8).[29] This balanced split indicates that the cases the judge resolves reflect genuine ambiguity in the syllabus text rather than systematic differences between the two primary models.

3.4 Validation

I validate the judge-based classification 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 LLM-Judge or the held-out rater is more likely to agree with this benchmark. The LLM passes the test if its win-plus-tie rate, the share of comparisons where it matches the consensus at least as well as the held-out rater, exceeds 50 percent. Since human raters can only classify a small share of the nearly 150,000 syllabi in the sample, passing this test establishes that the automated approach can credibly extend the human classifications to the full set of syllabi.

Table 3 reports the validation results. I find that the average leave-one-out agreement rate among the four human raters reaches 97 percent, and the LLM-Judge agrees with the modal human response in 95 percent of syllabus-question pairs, comparable to the 97 percent average agreement among human raters documented in Section 3.2.[30] As a result, the LLM-Judge passes the Alternative Annotator Test, achieving a win-plus-tie rate of 96 percent, well above the 50-percent threshold indicating human-level performance.

I also assess whether the judge-based classification is robust to the choice of underlying model. I compare the accuracy of each individual primary model (GPT-Oss and Qwen) against the human modal response (Appendix C.4). All three approaches achieve accuracy above 90 percent on every question, with the models performing comparably overall. These results fit in with the ensemble literature showing that aggregation reduces the influence of

[28] The evidence selection mode applies only to the six teaching and evaluation questions, where the curated passages support the downstream qualitative taxonomy. For the two policy questions, the judge operates only in adjudication mode. Appendix C presents the evidence selection prompt verbatim.

[29] Across the six teaching and evaluation questions, the judge sides with each primary model at broadly similar rates (Table C8), confirming that the adjudication does not systematically favor either classifier.

[30] Agreement is highest for the evaluation questions (98–100%) and lowest for scheduled AI topic (91%), paralleling the pattern observed in human inter-rater reliability (Appendix C.4).

idiosyncratic model behavior (Dietterich, 2000), though even the individual models meet the validation threshold.

Altogether, these validation results establish that the judge-based classification reliably recovers the AI integration measures at scale, achieving agreement with human raters that is comparable to the agreement among the raters themselves. I therefore use the judge-based canonical classification to characterize AI integration across the full sample of syllabi.

4 The Prevalence and Evolution of AI Integration in Higher Education

4.1 Prevalence of AI Integration

Figure 2 presents the results on the extent to which artificial intelligence has been integrated into higher education in the Spring 2026 semester. First, the LLM-based approach indicates that close to 60 percent of courses in the sample include an explicit AI policy in their syllabi.[31] This share is largely in line with the text-based measure that captured syllabi mentioning artificial intelligence discussed in Section 2.4. At the same time, the share of courses that explicitly allow students to use AI is significantly smaller. Altogether, just 26 percent of courses allow students to use artificial intelligence. As such, the majority of courses with policies in place are not explicitly permitting students to use AI in their classes.[32]

Figure 2 also shows that direct AI integration into teaching is substantially rarer. I find that only about 4 percent of syllabi include a scheduled module or topic devoted to AI, and about 2 percent assign AI-related readings or materials. In-class AI exercises appear in fewer than 1 percent of courses. Altogether, 5 percent of syllabi display concrete evidence that AI has been integrated into teaching, whether through course content, class activities, or assigned readings. The gap between the close to 60 percent of courses with an AI policy and the 5 percent with direct teaching integration suggests that the classroom response to generative AI has primarily taken the form of formal acknowledgment. Direct changes to what students are actually asked to learn remain far less common, fitting in with McDonald et al. (2025).

Changes to how students are evaluated represent the deepest form of classroom adoption, and this margin remains the rarest. Only about 1 percent of syllabi include a named graded assignment that requires students to use AI, and structured engagement with AI-generated content or submission of AI output are similarly uncommon. These low rates fit in with the fact that redesigning how students are graded around a new technology requires greater changes to the course than scheduling a new topic or assigning a reading (Frankel and Kartik, 2022).[33] Altogether, 2 percent of syllabi incorporate AI into evaluation, and 6

[31] Figure A2 presents corresponding evidence across the full 2024-2026 sample.

[32] Figure A4 shows that the prevalence of these measures is broadly similar across the two main models used in the analysis (GPT-Oss and Qwen).

[33] Tables A7 and A8 provide examples of what these measures look like in practice, from dedicated AI modules and in-class exercises to AI-assisted projects and assignments built around evaluating AI output.

percent incorporate AI into either teaching or evaluation.

Figure A3 presents the extent to which the teaching and evaluation measures are jointly present across syllabi in the main sample. Among the courses with any classroom integration, 69 percent appear only on the teaching margin, while 15 percent appear only on evaluation. Moreover, the three teaching measures capture different ways that instructors bring AI into the classroom. About half of the courses with a scheduled AI topic do not also assign AI-related readings, and roughly three quarters do not include an in-class AI exercise. In other words, a course that devotes a week to large language models and a course that assigns students to read about AI are integrating it into instruction through different channels. A similar pattern holds within evaluation, where most courses with a graded AI assignment do not also require students to submit AI-generated output.[34]

4.2 AI Integration over Time

The first panel of 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 permitted 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 the share of courses with any teaching or evaluation use rose only to about 10 percent by Spring 2026.[35] 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.

The second panel of Figure 3 shows a similar pattern over the past four semesters in the broader sample.[36] Here too, the share of syllabi that explicitly permit AI rises quickly, moving from roughly one in eight courses in Fall 2024 to more than one in four by Spring 2026. By contrast, classroom integration increases more gradually over the same period, rising from about 4 percent to over 7 percent. Most of that increase again comes through teaching rather than evaluation, although the broader point is simply that formal permission moved faster than direct classroom use. Table A9 shows the corresponding semester-by-semester changes for the underlying policy, teaching, and evaluation measures, and it shows the same broad increase across the four semesters.[37] Altogether, even in this shorter window,

[34] Among teaching-positive syllabi, about 60 percent show exactly one of the three teaching signals, while about 7 percent show all three. Among evaluation-positive syllabi, the corresponding shares are about 40 and 30 percent. Cross-domain overlap is also low, with fewer than one in six classroom-positive syllabi combining both teaching and evaluation integration.

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

[36] This panel uses the balanced sample of the 12 universities observed in all four paper semesters. Figure A5 presents the corresponding evidence for the full sample.

[37] Table A9 uses different samples from Figure 3. Panel A shows the full 25-university sample, while Panel B shows a separate course-balanced sample.

colleges appear to have moved faster in formalizing permission for AI than in incorporating it into classroom practice.

5 Heterogeneity Across Universities and Majors

5.1 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. In fact, Contractor and Reyes (2025) show that students in science majors at Middlebury are more likely to use AI than their peers in the humanities. Similarly, Handa et al. (2025) find that students in STEM-related majors were more likely to be early users of Claude.

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 20 percent of courses in Marketing have integrated AI through teaching or evaluation, compared to just one percent of courses in Music, Dance, and Geology. Altogether, 12 percent of courses in Business majors have integrated AI into the classroom, which more than doubles the adoption rate for majors in the humanities, STEM and social science.[38] While this integration is more likely to occur through the teaching margin, rather than through AI-related assignments (Figure A7), the two margins are strongly correlated across majors. Majors that 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 A10). This mirrors the differential prevalence of AI mentions across fields presented earlier. At the same time, this table shows that 35 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 expliciy AI policy, they are much more likely to permit students to use it when they do. This indicates that majors in Social Sciences in the Humanities have moved rapidly on the policy margin without making corresponding changes to integrate AI into the classroom.[39]

I draw on the longer UT Austin and UT Dallas panel to understand the evolution of these differences across majors. Figure A11 shows that while Business majors were more likely to engage with AI content in the classroom prior to ChatGPT, the cross-field differences grow substantially after Fall 2022. Figure A9 presents corresponding evidence for the broader sample, showing that Business majors have most rapidly integrated AI into

[38] There are significant differences in adoption within the four broad majors in the sample, as 12 percent of English courses show direct evidence of classroom integration, compared to just 4 percent of History classes.

[39] 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 A6).

their courses in the past four semesters. This result thus fits in with the fact that business graduates are the most likely to use generative AI at work (Bick et al., 2025).

5.2 Heterogeneity Across Universities

The second panel of Figure 4 presents the corresponding evidence for classroom AI integration across universities in the Spring 2026 semester. I find that selective universities are much more likely to have incorporated generative AI into their classes than other universities, with about 10 percent of syllabi providing direct evidence of integration. This integration rate doubles the corresponding adoption rate at moderately selective and broad-access universities (Table A11). For instance, while more than 12 percent of courses at UT-Austin and UT-Dallas have already integrated AI into the classroom, this share remains below 6 percent at both University of Houston and West Texas A&M. Moreover, this integration-selectivity gradient is more pronounced in AI-based teaching activity, rather than through the evaluation margin (Figure A8).

The gradient in AI integration across the selectivity spectrum stands in marked contrast to the prevalence of explicit AI policies across the universities in the sample (Figure A6). That is, while all courses at Florida Atlantic include an AI policy in the Spring 2026 semester, this type of language appears in less than half of the syllabi at University of Florida.[40] In fact, Figure A12 shows that universities with higher SAT scores are less likely to address AI in their syllabi but more likely to integrate it into actual course activities. While actual classroom integration has increased across all selectivity levels in these past semesters (Figure A10), it remains most common for the selective universities in the sample. 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), suggesting that a similar gradient may apply in higher education.

In Table A13, I further examine the extent to which AI has been differentially integrated across different course levels. Explicit policy language is more common in undergraduate courses than in graduate courses.[41] By contrast, graduate classes are most likely to have integrated AI into teaching, and they also exhibit the highest overall rate of classroom AI integration. Altogether, 7 percent of graduate courses had incorporated AI into teaching or evaluation 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.

5.3 Drivers of Heterogeneous Adoption

Empirical Framework. 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

[40] Table A12 presents the prevalence of the main AI measures for the 25 universities in the sample.

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

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 associated with 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 5 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 at the same time. The pairwise results show that universities with lower average SAT scores are more likely to include AI policy language, confirming the negative selectivity gradient shown above. 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 become largely indistinguishable from zero in the multivariate specification.

I further analyze how these characteristics shape the extent of classroom integration in the second panel of Figure 5. First, universities with larger graduate sectors are more likely to integrate AI into the classroom. Similarly, universities with higher research expenditures per student are more likely to have integrated AI into their courses. This pattern also emerges for institutions with higher average faculty salaries, as a one SD increase in salaries is conditionally associated with a 1.5pp increase in classroom 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 the classroom.[42] Moreover, this result is stronger on the teaching-integration margin, where research expenditures and faculty salary remain individually significant even in the multivariate specification (Figure A13).[43]

This evidence thus indicates that the 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). My results suggest that a similar gradient may be emerging in higher education.

Major-Level Results. Figure 6 shows how major-level characteristics shape AI policy adoption conditional on university and course-level fixed effects. Across both the univariate and multivariate specifications, 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. This suggests that majors with greater AI exposure are also more likely to address it in their syllabi, once other differences across fields are accounted for.[44]

The second panel of Figure 6 presents the corresponding results for classroom AI integration. Here again, majors with lower graduate-degree shares are more likely to integrate AI into the classroom, as are those with lower average SAT scores. On the other hand, instructors in higher-paying majors are more likely to have integrated AI into their courses. The main result, however, is the AI-exposure gradient. Majors with greater AI exposure are significantly more likely to have integrated these tools into the classroom, with a one-standard-deviation increase associated with about a 2 percentage point increase in classroom integration. By contrast, the share of female students does little on this margin.

Altogether, these results indicate that instructors in majors with greater AI exposure are more likely to have integrated these tools into their courses. This pattern is especially clear on the teaching margin (Figure A14), where AI exposure is strongly associated with AI-based teaching but much less so with evaluation. This fits in with the evidence presented in Section 5.1, where the fields with the highest rates of classroom integration are also those where students and workers are most likely to use generative AI.

[42] Institutions with higher mean SAT scores and higher median parental income are also more likely to integrate AI into the classroom, yet these coefficients are insignificant in the multivariate specification.

[43] There is no corresponding relationship between universities' research environment and AI-focused assessments.

[44] Figure A14 presents the corresponding results for AI permission, teaching, and evaluation separately.

6 Drivers of AI Integration

6.1 Decomposition Approach

The results presented so far show that AI has been increasingly integrated into the classroom, albeit with substantial heterogeneity across universities, college majors, and courses. The extent of this integration may also vary across instructors, who are ultimately responsible for teaching the course and designing the syllabus. As such, to understand the extent to which these different layers contribute to shaping AI integration in syllabus n, 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, following the notation introduced above, $Y_n^{(k)}$ denotes syllabus-level outcome k for syllabus n. In equation (2), $\alpha_{i(n)}$ 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.[45]

I estimate equation (2) separately for the main outcomes of interest, and use the resulting specifications to carry out a variance decomposition of AI integration across these different margins.[46] 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 surrounding institutional and disciplinary context. In that sense, the $U \times M \times L \times T$ component captures the common background conditions that shape AI adoption within a given university-major-course-level-term cell.

Shapley–Owen Decomposition. Because the instructor, course, and contextual components in equation (2) overlap substantially in the data, a simple sequential decomposition would depend on the order in which these components are introduced. To address this, I implement a Shapley–Owen decomposition following Biasi and Ma (2022), which computes the average marginal contribution of each block of fixed effects to adjusted R^2 across all possible orderings of inclusion.[47] As such, it yields partial contributions that are additive,

[45] Throughout this section, I use $U \times M \times L \times T$ as shorthand for the university-by-major-by-course-level-by-term fixed effects captured by $\phi_{u(n),m(n),l(n),t(n)}$.

[46] 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. I further estimate the decomposition separately by broad field, selectivity tier, and course level to assess whether the relative importance of these margins varies across settings.

[47] Formally, for a given block j, the Shapley contribution equals $\sum_{T \subseteq V \setminus \{j\}} \frac{|T|!(K-|T|-1)!}{K!} \left[R^2(T \cup \{j\}) - R^2(T)\right]$,

order-invariant, and that sum to the total explained variation. Intuitively, the procedure asks how much knowing the instructor, the course, or the broader context helps explain AI integration in a syllabus, once I account for the fact that these margins overlap. I also implement a five-way version that disaggregates the contextual block into separate university, major, and term fixed effects. This allows me to unpack whether the broader institutional or disciplinary environment plays a distinct role beyond what instructors and courses already capture.

6.2 Decomposition Results

Figure 7 presents the results from the variance decomposition of equation (2) for the main measures of AI integration. The first result is that AI policy is shaped largely by instructors. Instructors explain about 36 percent of the variance in whether a syllabus includes AI policy language, while local course fixed effects explain about 23 percent and the university-by-major-by-level-by-term component explains about 17 percent. 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. In both cases, a sizable share of the variation remains unexplained.

The same figure also shows a different pattern once I turn to realized classroom use. Instructors explain about one quarter of the variance in any teaching, any evaluation, and any classroom integration. Local course fixed effects explain a similarly large share, slightly above instructors for teaching and classroom integration and only slightly below for evaluation. By contrast, the university-by-major-by-level-by-term component explains very little on these margins. Thus, most of the systematic variation in whether AI is actually used in the classroom comes from instructors and local courses rather than the broader academic environment.[48]

These results are similar in the additional decompositions. Figure A16 shows that once I replace local course fixed effects with course-year fixed effects, instructors explain the largest share of the variance for all five main outcomes, including any classroom integration. The first panel of Figure A17 shows that the relative magnitudes of these components are similar across the five broad major groups, while the second panel shows the same across selectivity tiers. Figure A18 shows a similar pattern across undergraduate and graduate courses, although the teaching and classroom margins are somewhat more course-driven in graduate courses.[49]

Table 4 then separates the broader component into university, major, and term fixed

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 $K = |V|$. I use adjusted rather than raw R^2 because the fixed-effects blocks differ sharply in the number of categories they contain.

[48] Figure A15 presents the corresponding results for the specific AI measures. The same broad pattern appears there as well, although some measures are somewhat more instructor-driven and others somewhat more course-driven.

[49] Figure A16 reports the specification with course-year fixed effects, while Figure A17 and Figure A18 report the corresponding decompositions by broad field, selectivity tier, and course level.

effects. The first column shows that AI policy remains primarily instructor-driven in the five-way decomposition, with instructors explaining about 37 percent of the variance and local courses about 24 percent. Universities explain about 7 percent of the variance in AI policy, whereas majors and terms each contribute relatively little. This suggests that universities may play some role in shaping whether AI use is formally regulated in the syllabus, although that role remains clearly smaller than that of instructors and local courses.

The last two columns of Table 4 show a different result for realized classroom integration. Local courses explain about 31 percent of the variance in any teaching and any classroom integration, while instructors explain about 26 to 27 percent, and the university component is close to zero. This stands in contrast to the policy results, where universities still explain a modest share of the variance. In that sense, the forces that shape whether AI appears in syllabus policy are not the same as the forces that shape whether AI is actually incorporated into teaching. This broader pattern is consistent with Biasi and Ma (2022), who similarly show that instructors and courses explain much more of the relevant curricular variation than schools do.[50]

6.3 Course Characteristics

I next examine which types of courses are more likely to adopt AI policy language and to integrate AI into the classroom. To do so, I estimate a modified version of equation (2) that replaces the local-course fixed effects with observable course characteristics, including course level, log enrollment, credit hours, and whether the course has a teaching assistant.[51]

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.[52]

On the other hand, when it comes to whether AI has actually been incorporated into teaching or evaluation, the picture looks quite different. Figure 8 shows that freshman and sophomore courses are no more likely than graduate courses to integrate AI into classroom practice, and junior and senior courses are, if anything, somewhat less likely to do so. 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 schedule AI topics or assign AI-related readings.[53] In this context, graduate courses may

[50] Table A14 reports the corresponding five-way decomposition for the full set of specific AI outcomes and shows the same contrast between the policy-related measures and the realized measures of classroom integration.

[51] Table A15 reports the corresponding univariate estimates.

[52] Tables A17 and A18 report the multivariate coefficients separately by broad field and selectivity tier, showing that these patterns are broadly similar across different parts of higher education.

[53] These results are driven primarily through AI being incorporated into teaching rather than through evaluation (Table A16).

have more natural scope for incorporating AI into teaching, precisely because their content engages with frontier knowledge (Biasi and Ma, 2025).[54]

The course fixed effects in equation (2) are defined locally and therefore capture university-specific courses that are not directly comparable across institutions. As such, these course effects may reflect both how a course is implemented at a particular university and broader patterns that hold across the same harmonized course at different institutions. To separate these two margins, I focus on the Florida subsample and estimate a specification that includes both local course fixed effects and harmonized-course-by-term fixed effects.[55] Table A20 shows that local course fixed effects explain about 19 percent of the variation in classroom AI integration, and harmonized-course-by-term effects account for an additional 8 percent. These results indicate that some harmonized courses are more likely to integrate AI into the classroom across multiple universities in the same term, and not only because of how those courses are implemented at a particular institution.

7 Who Integrates AI into the Classroom?

7.1 Instructor Characteristics

The results presented in Section 6 show that instructors play a key role in shaping how AI is incorporated into the classroom. To understand which types of instructors are more likely to integrate AI into their courses, Table 5 presents descriptive evidence on how the main outcomes vary across instructor characteristics. On the policy margin, I find that female instructors are more likely to include explicit AI policy language in their syllabi, fitting in with the earlier evidence in Section 2. In particular, 66 percent of female instructors have at least one syllabus with an AI policy, compared with 57 percent of male instructors.[56] By contrast, the corresponding differences across academic rank and publication activity are much smaller. Assistant professors are somewhat more likely than full professors to adopt AI policy language, yet the spread across rank groups is limited, and the difference between instructors with and without publications is smaller still.

Table 5 shows a different pattern once I turn to whether AI is actually incorporated into the classroom. Here, the gender difference largely disappears. Instead, classroom integration is more common among assistant professors and among instructors with publication records. About 14 percent of assistant professors integrate AI into the classroom, compared with roughly 11 percent of professors, associate professors, and non-tenure-track faculty. I find a similar gap across research activity, as 13 percent of instructors with publications

[54] In the UT Austin and UT Dallas subsample, where I additionally observe pre-2024 average GPA and the historical share of course points allocated to take-home assignments, higher-GPA courses are more likely both to adopt AI policy and to integrate AI into the classroom, and courses whose pre-AI assessment structure relied more heavily on take-home assignments are more likely to integrate AI into teaching and evaluation (Table A19).

[55] This is feasible because universities in Florida share a common state course code structure, which allows courses to be consistently harmonized across institutions.

[56] Table A21 shows that female instructors are also somewhat more likely to explicitly allow their students to use AI in their classes.

integrate AI into the classroom, compared with 10 percent of those without publications. Table A21 shows that these differences arise mainly through the teaching margin rather than through AI-focused assessments.[57] In that sense, the instructors who are quickest to formalize AI in the syllabus are not exactly the same as those who are most likely to incorporate it into classroom practice.

7.2 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, and years since last publication, while $\phi_{c(n)}$ and $\lambda_{u(n),m(n),l(n),t(n)}$ absorb local-course and $U \times M \times L \times T$ fixed effects as in equation (2). I estimate a version of Equation (3) that includes one covariate at a time and then a version that includes all of the main variables at the same time. To account for the differences in syllabi structure across instructors, I also estimate a specification that directly controls for the text-based characteristics described in Section 2.[58]

Empirical Results. The first panel of Figure 9 presents the results for AI policy adoption. I find that women are far more likely than men to have an explicit AI policy in their classes, even conditional on the courses they teach and on broader university, major, and term fixed effects. This is consistent with the broader engagement of female instructors with AI documented in Section 7.1, and suggests that women are responding to the arrival of generative AI more actively through formal course policies. Assistant professors and non-tenure-track faculty are also more likely to include an explicit policy in their syllabi relative to full professors, and instructors who have published more recently similarly show higher rates of AI policy adoption.

The second panel of Figure 9 shows a very different pattern once I turn to actual classroom integration. I find that assistant professors are about one percentage point more likely than full professors to have integrated AI into their courses. Research activity also matters on this margin, as instructors who have published more recently are more likely to integrate AI into the classroom, and those with any publications show a similarly positive relationship. In both cases, the pairwise and multivariate estimates are closely aligned, suggesting that these associations are not driven by composition across the other covariates. Non-tenure-track faculty, by contrast, are no more likely than full professors to integrate AI into

[57] The differences in AI-focused assessments are much narrower across these instructor groups. More broadly, Table A22 shows that the same ordering appears across fields and across the selectivity spectrum.

[58] 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.

their courses. Altogether, these results indicate that it is young tenure-track faculty with active research agendas who are leading the integration of AI in the classroom, fitting in with the descriptive patterns documented in the previous subsection.

This finding is consistent with Biasi and Ma (2022), who show that instructors with more recent publications teach courses significantly closer to the knowledge frontier, with assistant professors exhibiting the lowest education-innovation gap across academic ranks. It also fits with a growing body of work showing that younger and more recently trained workers tend to lead the adoption of new tools in the workplace (McElheran et al., 2024; Bick et al., 2025).

Figure A19 further shows that the patterns documented above for classroom integration are concentrated on the teaching side rather than in AI-focused evaluation. Assistant professors are more likely to schedule AI topics and assign AI-related readings, yet they are actually less likely to incorporate AI into graded assignments and evaluation.[59] The publication measures follow a similar pattern, with research activity predicting higher rates of teaching integration but showing little relationship with evaluation. This asymmetry suggests that career incentives may shape the form that adoption takes, as assistant professors appear willing to experiment with AI in their course materials while avoiding more consequential changes to grading and assessment (Goehring et al., 2024).

Table A24 confirms that it is not the volume of an instructor's research that predicts AI integration. I find that measures such as H-index, total publication count, and PhD-institution selectivity add essentially no predictive power once I control jointly for the core covariates. Instead, the instructors who are integrating AI most actively are recent graduates with active research agendas, not those with the longest publication records.

These findings remain broadly stable across several robustness exercises. Figure A20 shows that adding the text-based syllabus controls leaves the classroom integration results largely unchanged, and the policy coefficients similarly remain significant. This suggests that the instructor associations documented above are not driven by differences in syllabi structure across instructors. Figure A21 yields similarly stable estimates when I replace local-course fixed effects with course-by-year fixed effects, confirming that the results are not driven by time-varying course composition.

Finally, Table A25 examines whether these patterns replicate in the UT Austin and UT Dallas subsample, where I observe the same instructors across multiple semesters. The policy differences are substantially larger in this subsample, with female instructors about seven percentage points more likely and assistant professors about eight percentage points more likely to adopt AI policy language, while the classroom-integration results remain consistent with the full-sample estimates.

[59] Table A23 presents these results separately for seven specific AI outcomes, confirming that assistant professors are more likely to bring AI into course content and assigned readings but less likely to incorporate it into graded assignments and student deliverables.

7.3 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. In the first panel of Figure 10, I examine the extent to which instructor characteristics differentially shape AI integration for STEM and non-STEM majors. I find that in STEM fields, assistant professors are about 1.7 percentage points more likely than full professors to have integrated AI into their courses. Instructors with publications are also significantly more likely to have done so. In non-STEM fields, the career-stage and publication differences are more muted and less precisely estimated. In that sense, the instructors who are most actively experimenting with AI in the classroom are early-career researchers in the disciplines where these tools are most directly relevant (fitting in with the education-innovation gap documented by Biasi and Ma 2022).[60]

I further examine these patterns across five broad fields. In Table A26, I find that the career-stage premium in classroom integration is significant only in STEM, while the corresponding estimates in Humanities, Social Sciences, and Business are all close to zero. I find a more diffuse pattern on the policy margin, where the main instructor covariates predict AI policy language across most disciplines rather than in STEM alone (Figure A23, panel (a)).[61]

Since AI has been differentially integrated into the classroom at selective institutions compared with the rest of the sample, I further examine whether the instructor-level patterns also vary across institutional settings. The second panel of Figure 10 shows the effects across levels of institutional selectivity. I find that at selective universities, assistant professors are about 2.4 percentage points more likely than full professors to have integrated AI into the classroom. Instructors who have published more recently are similarly more likely to have done so. At less selective universities, the career-stage differences are much smaller, although having publications remains positively associated with classroom integration. Selective universities are thus not only more likely to integrate AI into the classroom altogether, but it is specifically early-career, research-active instructors at these institutions who are experimenting most actively with these tools (Table A22).[62]

On the policy margin, I find that the career-stage differences are similarly concentrated at selective institutions, where assistant professors are about seven percentage points more likely than full professors to include AI policy language (Figure A23, panel (b)). At less selective universities, these differences are largely absent. When I decompose the non-selective group further, I find that the career-stage premium in classroom integration is strongly positive only at the selective tier, while at moderately selective and broad-access institutions it is much smaller (Table A28).[63] In that sense, while the publication premium in classroom AI integration is present across most of the sample, it is assistant professors at

[60] Figure A22, panels (a)–(b), decomposes these patterns by teaching and evaluation margins.

[61] Table A27 presents the corresponding results for AI policy across the five broad fields.

[62] As with the field-level results, I find that these selectivity differences operate mainly through the teaching margin (Figure A22, panels (c)–(d)).

[63] Having any publications, on the other hand, is positively associated with classroom integration at both moderately selective and broad-access institutions (Table A28).

the most selective research universities who are leading the way.

Altogether, these results show that the instructor-level differences documented in the previous subsection are sharpest in STEM fields and at selective institutions. This fits in with a growing body of evidence that the adoption of new technologies tends to be most pronounced where organizational resources and human capital are already well developed (McElheran et al., 2024; Biasi and Ma, 2025). By contrast, I find that the female policy premium appears across nearly every field and selectivity tier I examine.

8 Conclusion

In this paper, I have examined how university instructors are responding to generative AI in their course syllabi, drawing on a dataset of close to 150,000 syllabi across 25 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 60 percent of Spring 2026 syllabi, yet that this diffusion has not translated into comparable changes to what instructors teach or how they assess students, with only about 6 percent of syllabi showing any classroom integration. Where classroom integration does appear, it is far more concentrated on the teaching margin than on evaluation, suggesting that instructors are more willing to experiment with what they present in the classroom than with how they measure student performance. I have further shown that this gap between policy and practice varies substantially across fields, course levels, and institutions. Business and applied fields lead in classroom integration, and the majors whose graduates face the most AI exposure in the labor market are also where instructors are most likely to integrate these tools into teaching. Selective and research-intensive universities show substantially higher rates of classroom integration, and graduate courses are more likely to have integrated AI into teaching than introductory courses, even though introductory courses carry more AI policy language.

To understand what drives these patterns, I have decomposed the variation in AI outcomes using a Shapley-Owen framework 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 in classroom integration, while the university component is negligible. Institutions appear to shape the conditions for formal policy diffusion, but they do not determine whether instructors actually redesign their courses around AI. Among observable instructor characteristics, assistant professors and instructors with recent publications are the most consistent predictors of classroom integration, and these effects are concentrated on the teaching margin rather than on evaluation. Measures of cumulative research output, including h-index and total publication count, add essentially no predictive power, suggesting that what matters is not the volume of an instructor's research but rather proximity to active engagement with new methods. These instructor-level effects are sharpest in STEM fields and at selective universities, fitting in with evidence that research-active instructors teach courses closer to the knowledge

frontier (Biasi and Ma, 2022). In that sense, it is young, research-active faculty at the most research-intensive universities who are leading the diffusion of AI into higher education classrooms.

Linking the syllabus-based measures developed here to student-level administrative records would allow researchers to study whether exposure to AI-integrating instructors changes the skills students acquire and the returns they earn in the labor market. More broadly, the finding that classroom integration is concentrated among early-career researchers in frontier-oriented environments raises the question of whether this pattern will persist as the technology matures, or whether adoption will eventually diffuse more broadly across ranks, fields, and institutions. How those individual decisions accumulate across higher education is likely to shape the next chapter of AI in the classroom.

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.

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.

Bettinger, E. P. and Long, B. T. (2004). Do college instructors matter? the effects of adjuncts and graduate assistants on students' interest and success. Working Paper 10370, NBER.

Bettinger, E. P. and Long, B. T. (2010). Does cheaper mean better? the impact of using adjunct instructors on student outcomes. *Review of Economics and Statistics*, 92(3):598–613.

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

Biasi, B. and Ma, S. (2025). Frontier knowledge in college and student success. Working paper, Yale SOM / NBER affiliation version. October 6, 2025 version.

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

Bloom, N., Hassan, T. A., Kalyani, A., Carvalho, M., Lerner, J., and Tahoun, A. (2025). The diffusion of new technologies. *Quarterly Journal of Economics*, 140(2):1299–1365.

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. and Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives*, 14(4):23–48.

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., Page, M. E., and West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *Quarterly Journal of Economics*, 125(3):1101–1144.

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.

Carvajal, D., Franco, C., and Isaksson, S. (2024). Will artificial intelligence get in the way of achieving gender equality? Discussion Paper 03, NHH Norwegian School of Economics.

Chen, K., Hansen, Z., and Lowe, S. (2021). Why do we inflate grades? the effect of adjunct faculty employment on instructor grading standards. *Journal of Human Resources*, 56(3):774–811.

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.

Conzelmann, J. G., Hemelt, S. W., Hershbein, B., Martin, S. M., Simon, A., and Stange, K. M. (2023). Skills, majors, and jobs: Does higher education respond? Working Paper 31572, NBER.

Covington, M. A. and McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2):94–100.

Cruces, G., Fernández Meijide, D., Galiani, S., Gálvez, R. H., and Lombardi, M. (2026). Does generative AI narrow education-based productivity gaps? evidence from a randomized experiment. Working Paper 34851, NBER.

David, P. A. (1990). The dynamo and the computer: An historical perspective on the modern productivity paradox. *American Economic Review*, 80(2):355–361.

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.

Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2025). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Organization Science*. Forthcoming. Harvard Business School Working Paper 24-013.

Dietterich, T. G. (2000). Ensemble methods in machine learning. *Lecture Notes in Computer Science*, 1857:1–15.

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.

Fairlie, R. W., Hoffmann, F., and Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review*, 104(8):2567–2591.

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.

Figlio, D. N., Schapiro, M. O., and Soter, K. B. (2015). Are tenure track professors better teachers? *Review of Economics and Statistics*, 97(4):715–724.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3):221–233.

Frankel, A. and Kartik, N. (2022). Improving information from manipulable data. *Journal of the European Economic Association*, 20(1):79–115.

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

Goehring, G., Mezzanotti, F., and Ravid, S. A. A. (2024). Technology adoption and career concerns: Evidence from the adoption of digital technology in motion pictures. Working Paper 32844, NBER.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, 25(4):501–522.

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.

Hall, B. H. and Khan, B. (2003). Adoption of new technology. Working Paper 9730, NBER.

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.

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.

Jones, M. P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association*, 91(433):222–230.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., and Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. Technical Report Research Branch Report 8-75, Naval Technical Training Command, Naval Air Station Memphis.

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.

National Center for Education Statistics (2020). Classification of instructional programs (CIP). https://nces.ed.gov/ipeds/cipcode. 2020 Edition.

Noy, S. and Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192.

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*.

Ravselj, D., Kerzic, D., Tomazevic, N., Umek, L., and many others (2025). Higher education students' use of generative AI: A multi-country study. *International Journal of Educational Technology in Higher Education*.

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.

Stohr, C., Ou, A. W., and Malmstrom, H. (2024). Perceptions and usage of AI chatbots among university students: A cross-disciplinary study. *Computers and Education: Artificial Intelligence*, 7:100231.

Wataoka, K., Hida, T., Baba, R., Oba, J., and Mori, S. (2025). Self-preference bias in LLM-as-a-judge. In *Proceedings of the 13th International Conference on Learning Representations (ICLR)*.

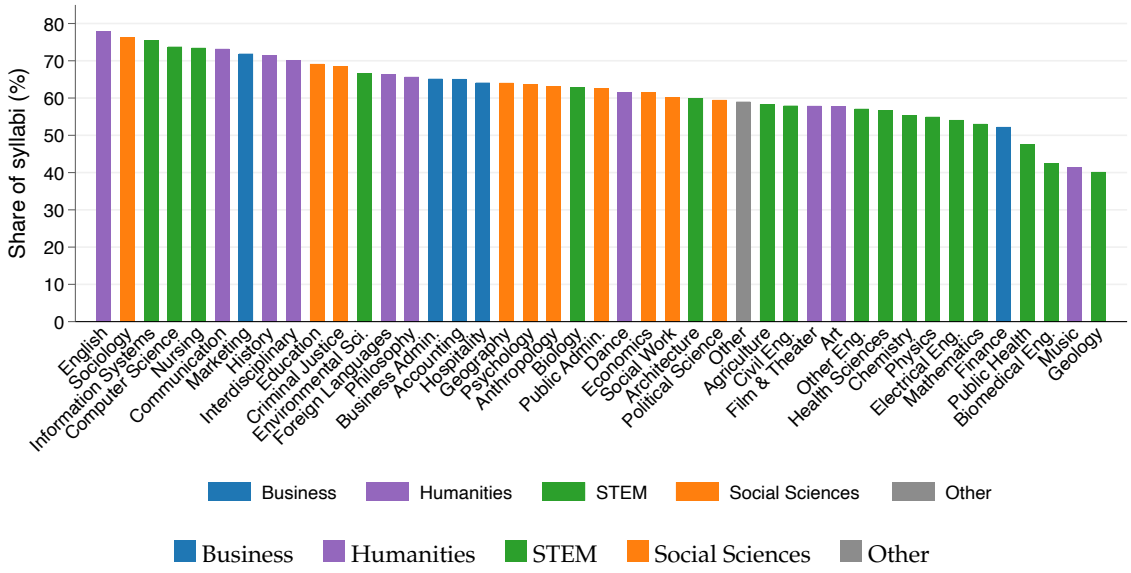
Yotzov, I., Barrero, J. M., Bloom, N., and Davis, S. J. (2025). Firm data on AI. Working Paper 34836, NBER.

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., and Stoica, I. (2024). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. In *Advances in Neural Information Processing Systems*, volume 36. NeurIPS 2023.

Main Exhibits

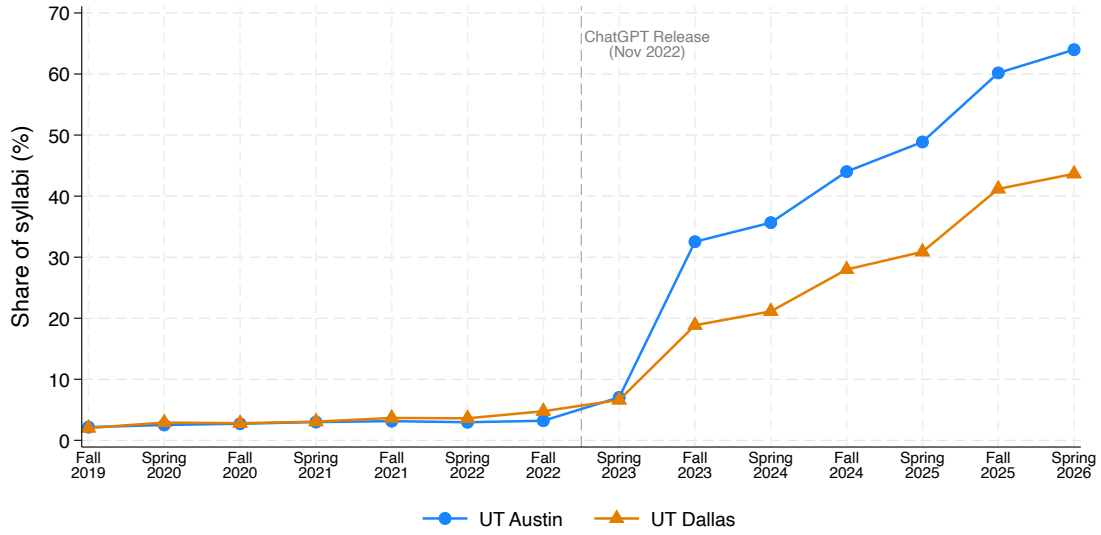
Figure 1: AI prevalence across disciplines and over time

(a) By Major



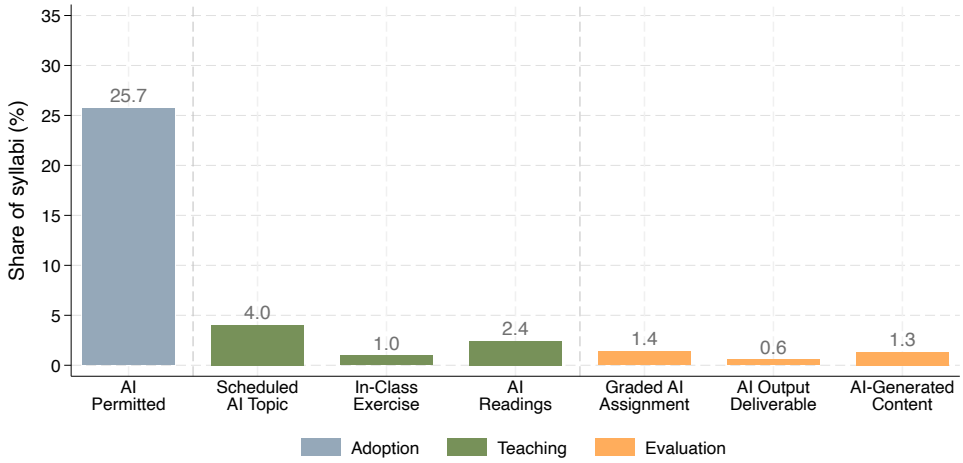
Business Humanities STEM Social Sciences Other

(b) Time Series



Note: Figure 1 shows the prevalence of AI mentions in the syllabi. Panel (a) reports the Spring 2026 share of syllabi mentioning AI by harmonized major, ordered from highest to lowest prevalence and colored by macro-field. Panel (b) reports 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.

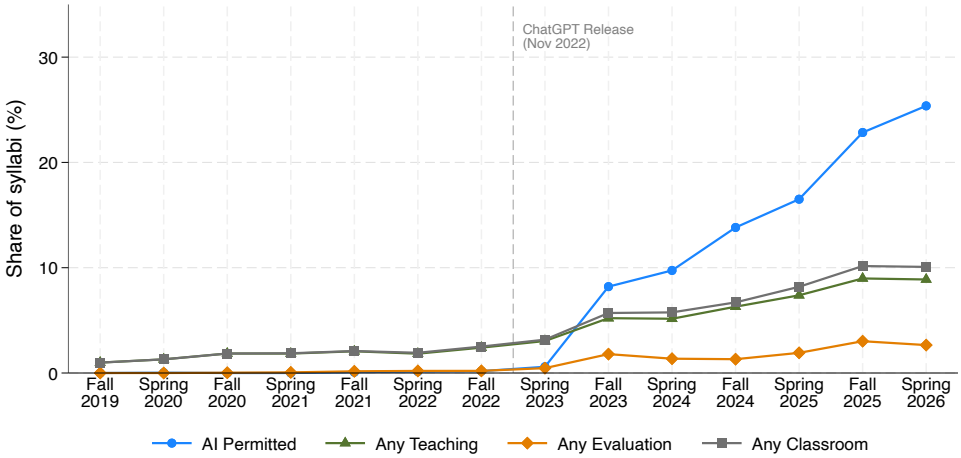
Figure 2: Spring 2026 Prevalence of AI Policy, Teaching, and Evaluation Measures



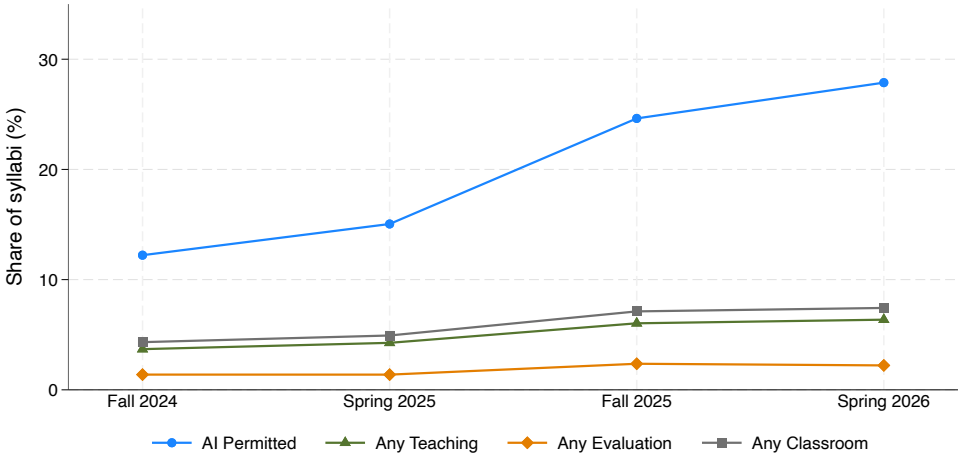
Note: Figure 2 shows the prevalence of AI policy, teaching integration, and evaluation integration measures in Spring 2026. The bars report the share of syllabi with each measure. In Spring 2026, 58.9% of syllabi include an AI policy, 5.1% display any teaching integration, and 1.8% display any evaluation integration.

Figure 3: AI Policy and Classroom Integration Over Time

(a) UT Austin and UT Dallas (balanced)



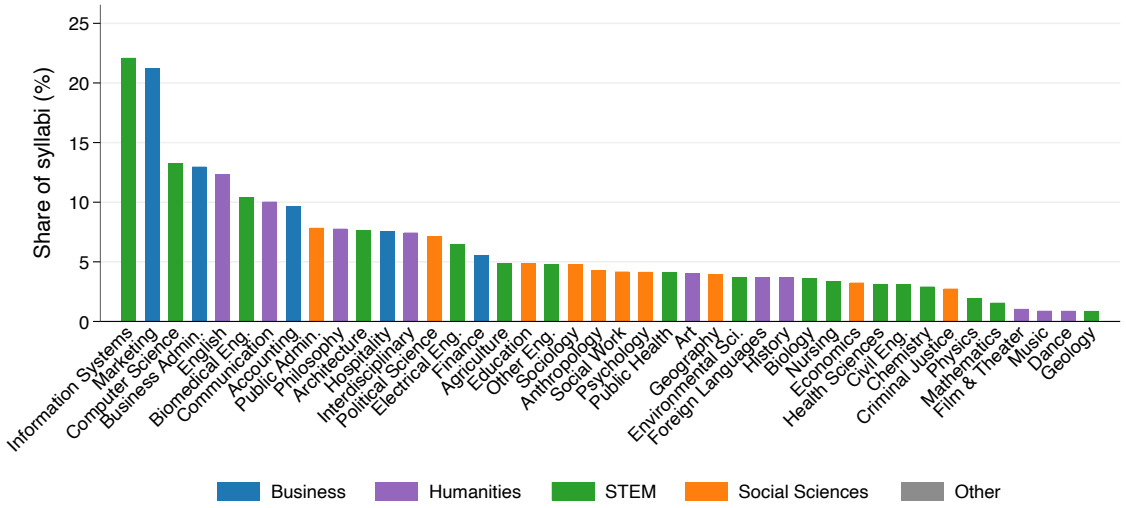
(b) Cross-sectional sample (balanced)



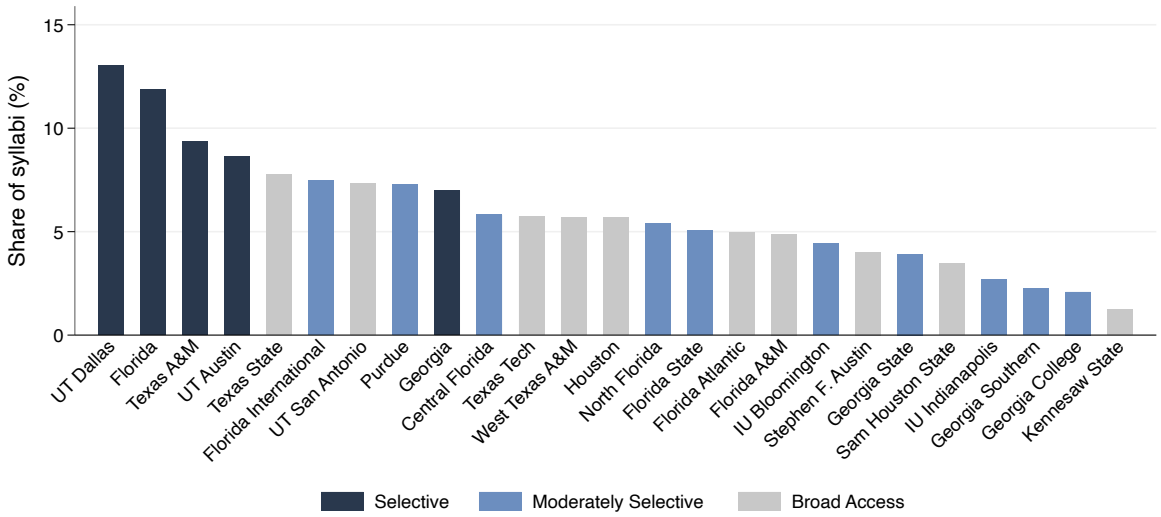
Note: Figure 3 presents changes over time in AI policy and classroom integration. Panel (a) reports the balanced-course panel for UT Austin and UT Dallas from Fall 2019 through Spring 2026, where balance requires a course to appear in at least one term in both 2024–2025 and 2025–2026. Panel (b) reports the broader balanced-university sample, which includes the 12 universities observed in all four semesters. The lines track AI policy, any teaching integration, any evaluation integration, and any classroom integration. In panel (b), AI policy rises from 38.6% in Fall 2024 to 59.7% in Spring 2026, while any classroom integration rises from 4.3% to 7.4%.

Figure 4: Classroom AI Integration Across Majors and Universities

(a) By Major



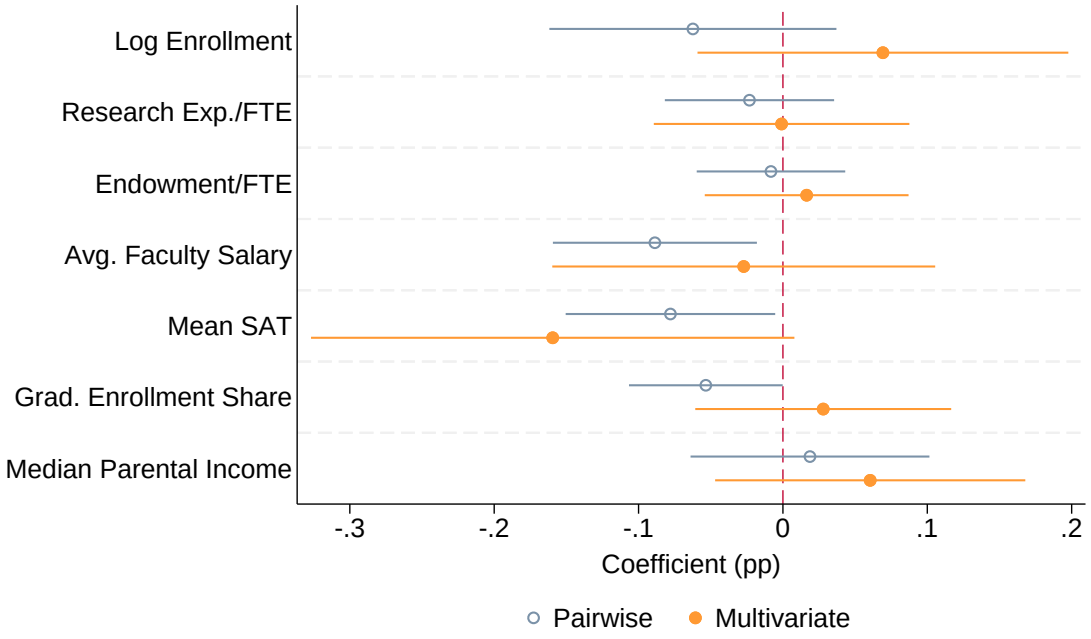
(b) By University



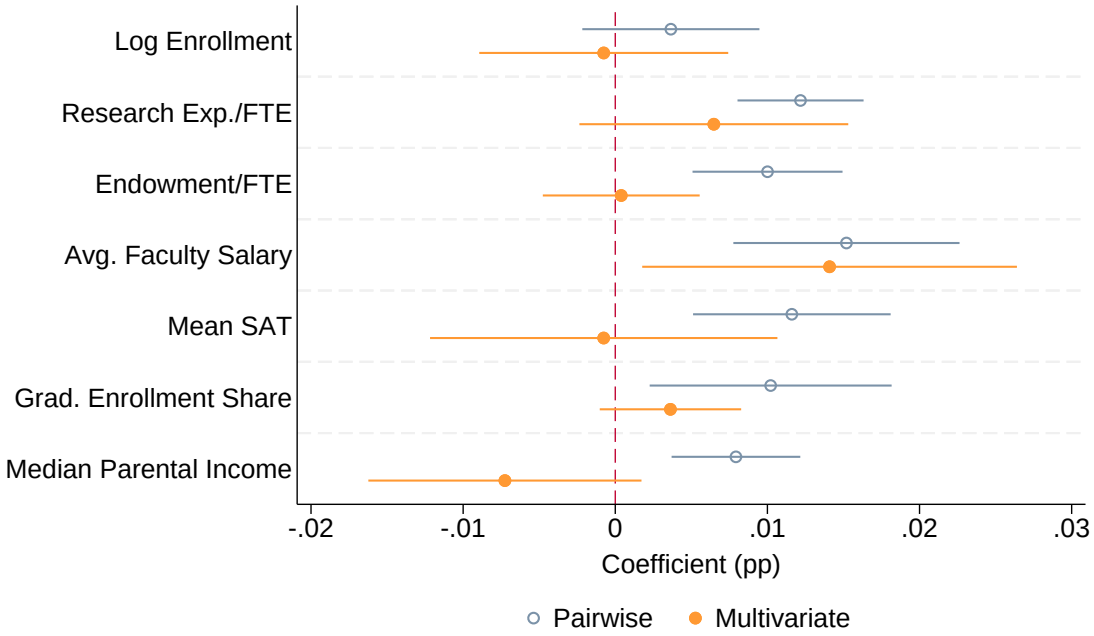
Note: Figure 4 presents the share of syllabi with any classroom AI integration (teaching or evaluation) in Spring 2026. Panel (a) shows prevalence across 42 harmonized majors, with bars colored by broad field (Business, Humanities, STEM, Social Sciences, Other). Panel (b) shows prevalence across the 25 universities, ordered from highest to lowest, with bars colored by selectivity tier (Selective, Moderately Selective, Broad Access).

Figure 5: University Covariates and AI Integration

(a) AI Policy



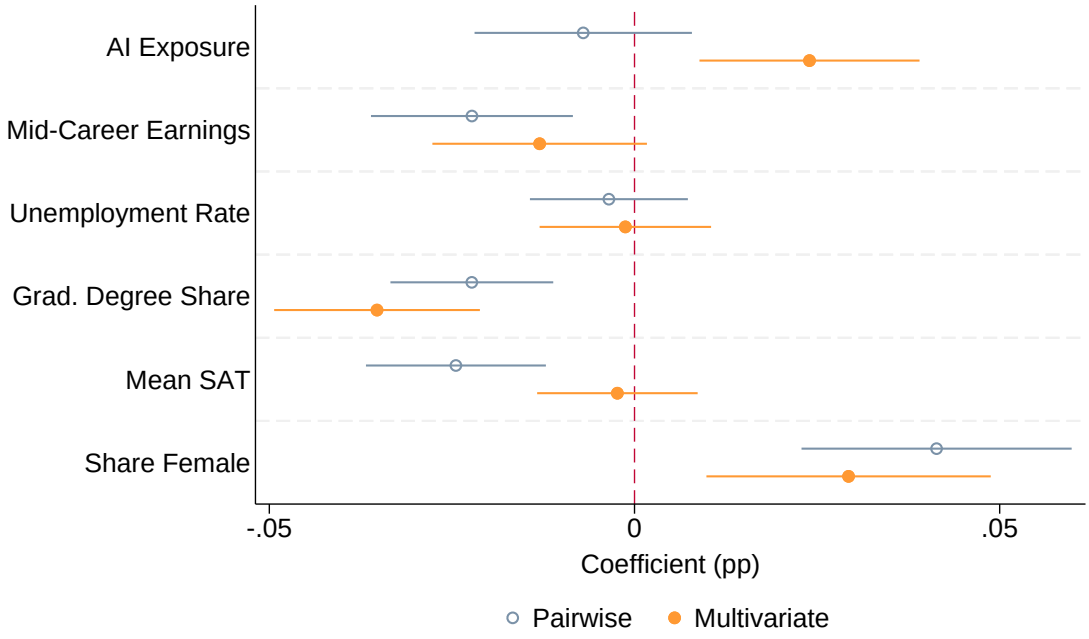
(b) Any Classroom



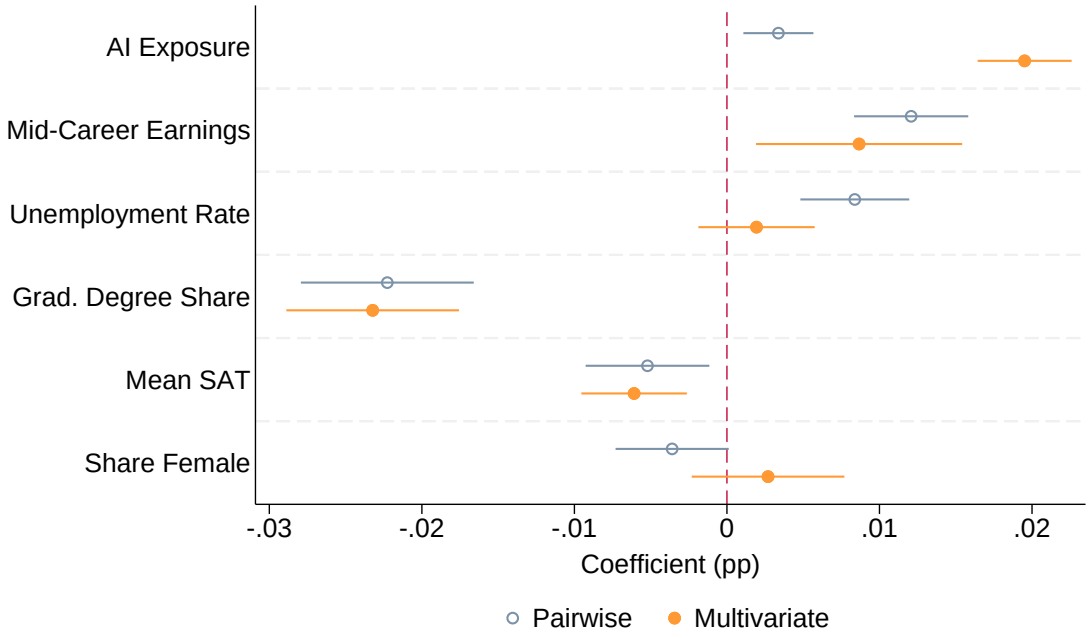
Note: Figure 5 presents coefficient estimates from regressions of AI integration on seven university-level covariates. Panel (a) reports results for having an AI policy; panel (b) reports results for any classroom AI integration. Circles report pairwise estimates; diamonds report multivariate estimates including all covariates simultaneously. Bars represent 90% confidence intervals. Fixed effects: major × course level × term. Standard errors clustered at the university level.

Figure 6: Major Covariates and AI Integration

(a) AI Policy

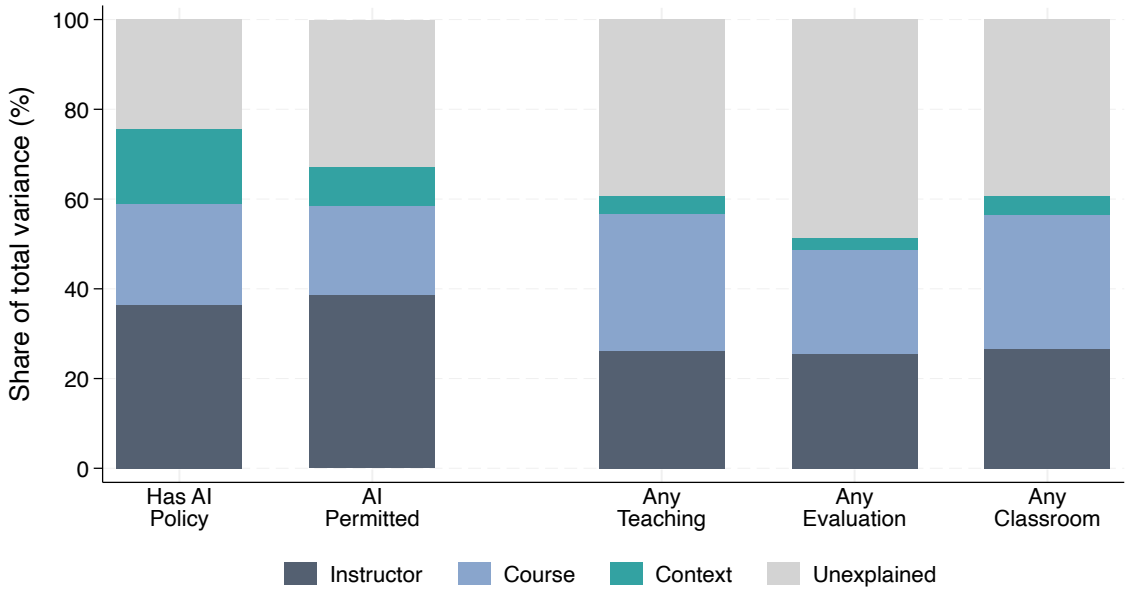


(b) Any Classroom



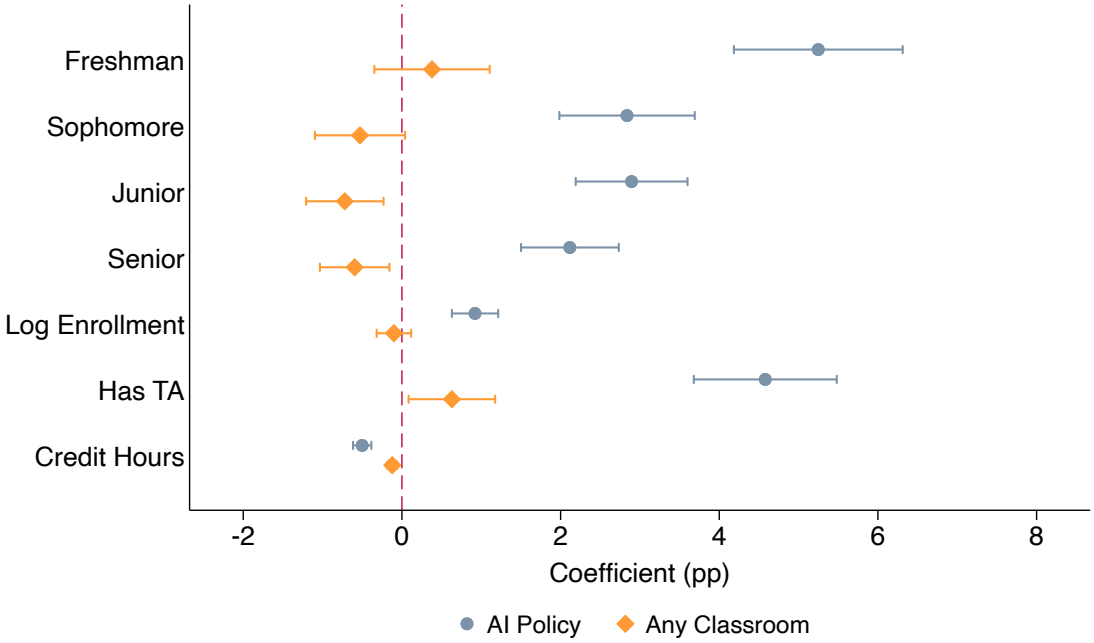
Note: Figure 6 presents coefficient estimates from regressions of AI integration on six major-level covariates. Panel (a) reports results for having an AI policy; panel (b) reports results for any classroom AI integration. Circles report pairwise estimates; diamonds report multivariate estimates including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

Figure 7: Variance Decomposition of the Main AI Outcomes



Note: Figure 7 shows the baseline three-way Shapley–Owen variance decomposition for the five main AI outcomes. Each bar decomposes the total variance of the outcome into the shares attributed to instructor fixed effects, local course fixed effects, university-by-major-by-level-by-term fixed effects, and the unexplained residual. The bars sum to 100 percent of total variance.

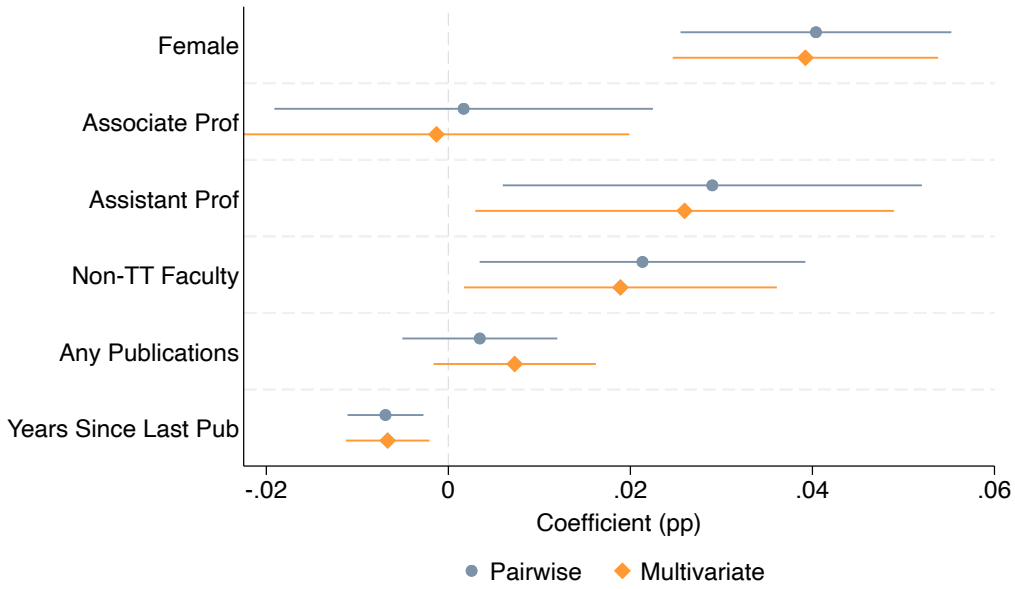
Figure 8: Course Characteristics and AI Integration



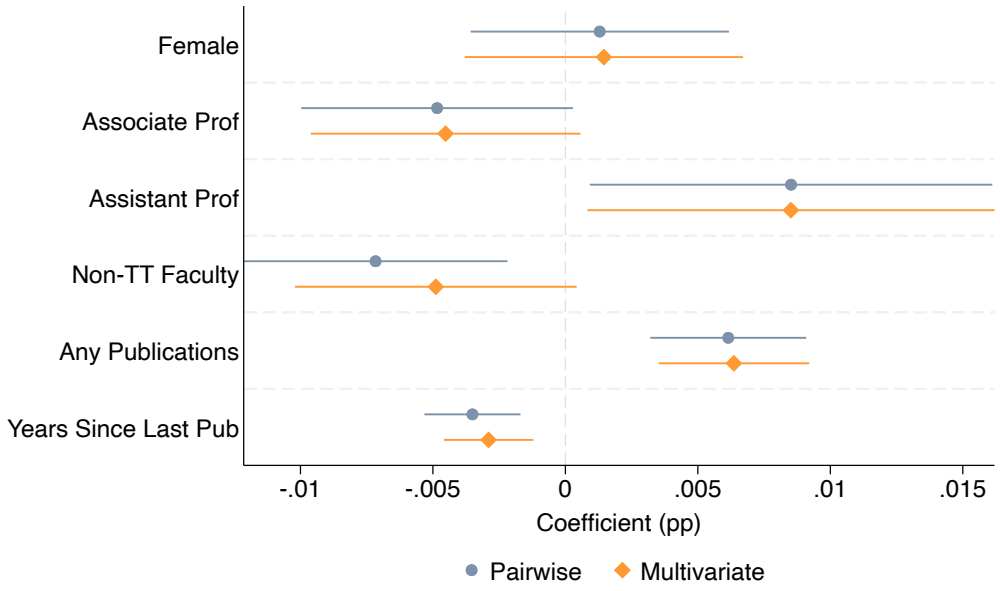
Note: Figure 8 shows the multivariate course-characteristic coefficients for Has AI Policy and Any Classroom. The specification replaces local course fixed effects with observable course characteristics, includes instructor fixed effects and university-by-major-by-term fixed effects, and omits Graduate 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.

Figure 9: Instructors and AI Integration in Higher Education

(a) AI Policy

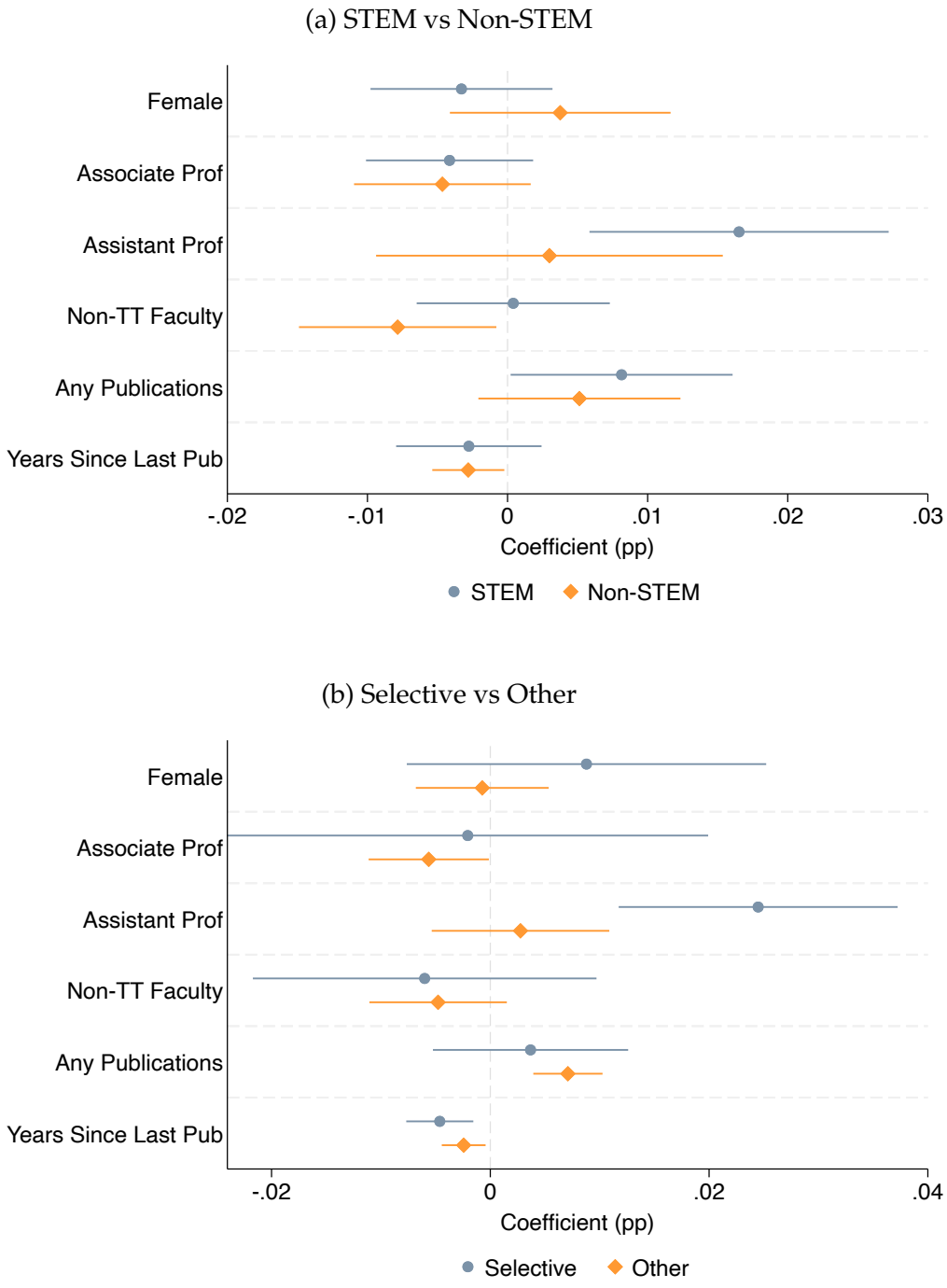


(b) Any Classroom



Note: Figure 9 shows coefficient estimates from regressions of AI outcomes on instructor-level covariates. Circles (blue) report pairwise estimates from separate regressions; diamonds (orange) report multivariate estimates including all covariates simultaneously. All continuous covariates are standardized to one standard deviation. Bars represent 90% confidence intervals. The specification absorbs local course fixed effects and university×major×level×term fixed effects. Standard errors are clustered at the university level. The sample covers the 25 primary universities from Fall 2024 through Spring 2026.

Figure 10: Heterogeneity in Classroom AI Integration



Note: Figure 10 shows heterogeneity in the multivariate instructor-characteristic coefficients for any classroom AI integration. Panel (a) compares STEM and non-STEM fields; panel (b) compares selective (Barron's tier 1) and other institutions. Bars represent 90% confidence intervals. The specification absorbs local course and university×major×level×term fixed effects. Standard errors are clustered at the university level. The sample covers the 25 primary universities from Fall 2024 through Spring 2026.

Table 1: Descriptive Statistics

	Overall (1)	**AI Mentioned** (2)	**No AI Mentioned** (3)
Syllabi Characteristics			
AI Mentioned	0.62	1.00	0.00
Number of Words	2,702	3,076	2,081
Has TA	0.17	0.18	0.15
Instructor Characteristics			
Female	0.47	0.51	0.41
Tenure-Track	0.51	0.51	0.50
Any Publications	0.37	0.37	0.38
University Characteristics			
Selective	0.20	0.20	0.20
Moderately Selective	0.43	0.37	0.53
Broad Access	0.37	0.43	0.26
Major Characteristics			
Business	0.09	0.09	0.08
STEM	0.37	0.35	0.40
Social Sciences	0.22	0.22	0.20
Humanities	0.31	0.32	0.30
Course Characteristics			
Undergraduate	0.76	0.78	0.71
Graduate	0.24	0.22	0.29
Syllabi	59,250	36,966	22,284

Note: Table 1 shows descriptive statistics for the Spring 2026 cross-section across the 25 universities in the main sample. The first column reports means for the full Spring 2026 sample; the second and third columns report means separately for syllabi that do and do not contain AI-related keywords. Female and Tenure-Track are reported among syllabi with non-missing instructor information for those characteristics. University selectivity follows the three-tier Barron's classification. Business, STEM, Social Sciences, and Humanities denote the harmonized macro-fields, and Undergraduate combines basic and advanced undergraduate courses. Table A6 reports the corresponding evidence for the full 2024–2026 sample period.

Table 2: AI Integration Classification Questions and Guidance

Has AI Policy — Does the syllabus contain an explicit policy, rule, or guideline about student use of AI or generative AI tools for coursework? *Guidance:* Syllabi that state rules, permissions, prohibitions, or expectations about student use of AI tools for coursework, whether from the instructor or the university. A generic academic integrity statement that does not specifically mention AI does not qualify.
Permits AI — Does the syllabus explicitly state that students are permitted, allowed, or welcome to use AI tools for coursework? *Guidance:* Syllabi that explicitly permit, allow, or welcome student use of AI tools for coursework, whether broad or targeted to specific assignments. The permission may appear in a policy section, assignment description, or course overview. A syllabus that only prohibits AI, or that discusses AI only as a course topic without permitting student use, does not qualify.
Scheduled AI Topic — Does the course schedule, calendar, or module list dedicate at least one class session, week, or module to an AI-related topic? *Guidance:* A schedule entry names an AI-related topic with a date, week number, module label, or session number. A syllabus that mentions AI only in a policy or guidelines section without any AI-related entry in the course schedule does not qualify.
In-Class AI Exercise — Does the syllabus describe an in-class activity, demonstration, or exercise involving AI that occurs during scheduled class time? *Guidance:* An activity, exercise, demonstration, or workshop involving AI tools takes place during scheduled class time. A policy statement or an AI assignment done outside class does not qualify.
AI Readings and Materials — Does the syllabus assign readings, videos, tutorials, or other instructional materials specifically about AI? *Guidance:* Specific readings, articles, videos, tutorials, podcasts, or certifications about AI or generative AI appear in a reading list, schedule, or assignment description. A topic title in a schedule without a named reading, or a policy section without assigned materials, does not qualify.
Named Graded AI Assignment — Does the syllabus describe at least one specific, named assignment that explicitly requires students to use an AI tool, where the AI use is part of the graded deliverable? *Guidance:* At least one named, graded assignment cannot be completed without AI, and the AI use contributes to the student's grade. General permission to use AI on existing assignments without requiring it for any specific deliverable does not qualify.
AI Output Deliverable — Does any assignment require students to submit AI-generated output (transcripts, screenshots, generated text or images) as part of the graded submission? *Guidance:* AI-generated artifacts, such as conversation transcripts, screenshots of interactions, or AI-generated text, must be submitted as part of a designed assignment. A conditional disclosure policy requiring students to report AI use if it occurs does not qualify.
Engagement with AI Content — Does any assignment require students to engage with AI-generated content as a structured step in the assignment workflow? *Guidance:* Students must actively work with AI output as a designed workflow step, by producing, evaluating, comparing, critiquing, revising, or responding to AI-generated content. A general suggestion to use AI for brainstorming without a structured engagement step does not qualify.

Note: Table 2 presents the eight binary classification questions used to measure AI integration in course syllabi. Policy and adoption measures are applied to the full syllabus text. Teaching and evaluation measures are also applied to the full syllabus text. Each measure is classified independently through a separate API call. Appendix C presents the full coding guidance with worked examples for each question.

Table 3: Agreement with Modal Human Response (%)

	Individual Raters						
Question	**R1**	**R2**	**R3**	**R4**	**Mean**	**LLM-Judge**	**AAT**
Has AI Policy	88	100	100	96	96	93	95
Permits AI	100	96	98	94	97	93	95
Scheduled AI Topic	87	92	90	93	90	91	93
In-Class AI Exercise	98	98	98	100	98	93	95
AI Readings	94	94	96	94	94	92	92
Graded AI Assignment	98	100	100	100	100	98	100
AI Output Deliverable	98	100	100	100	100	98	98
Engagement with AI	100	100	100	100	100	98	98
Average	**95**	**97**	**98**	**97**	**97**	**95**	**96**

Note: Table 3 compares the automated LLM-Judge classification to human raters across the eight headline measures. Each rater column reports leave-one-out agreement with the modal response of the remaining raters. Mean is the average across the four human raters. LLM-Judge reports the agreement rate between the judge-adjudicated canonical classification and the human modal response. AAT reports 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.

Table 4: Shapley–Owen Decomposition of AI Integration

	Has AI Policy (1)	**Any Teaching** (2)	**Any Classroom** (3)
Instructor	37.3	26.4	26.9
Course	23.5	31.3	30.9
University	6.6	0.3	0.3
Major	1.2	1.1	1.1
Term	2.9	0.2	0.2
Unexplained	28.5	40.7	40.6
Adjusted R^2	71.5	59.3	59.4
Observations	98,146	98,146	98,146

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

Table 5: Instructor Characteristics and AI Integration

	AI Policy (1)	**Any Classroom** (2)
Gender		
Female	66.1	11.0
Male	56.8	11.3
Academic Rank		
Professor	60.3	11.0
Associate Professor	62.8	11.5
Assistant Professor	64.8	14.4
Non-Tenure-Track	59.3	11.5
Research Activity		
Any Publications	61.6	12.6
No Publications	60.4	10.2
Publication Count Tercile		
Low	64.2	11.8
Medium	65.5	13.3
High	56.0	12.8
H-Index Tercile		
Low	64.8	11.8
Medium	64.7	13.6
High	56.0	12.7
PhD at Highly Selective Institution		
Yes	61.5	14.7
No	64.9	11.9
Instructors	47,020	

Note: Table 5 shows the prevalence of the main AI outcomes by instructor characteristics. Each cell reports the share of instructors for whom at least one syllabus exhibits the column outcome, in percent. Unit of observation is the instructor (one row per institution–instructor pair). Publication count and h-index terciles are computed among instructors with positive values. PhD institution selectivity follows the Barron's classification. The sample covers the 25 primary universities from Fall 2024 through Spring 2026.

Supplementary Appendices

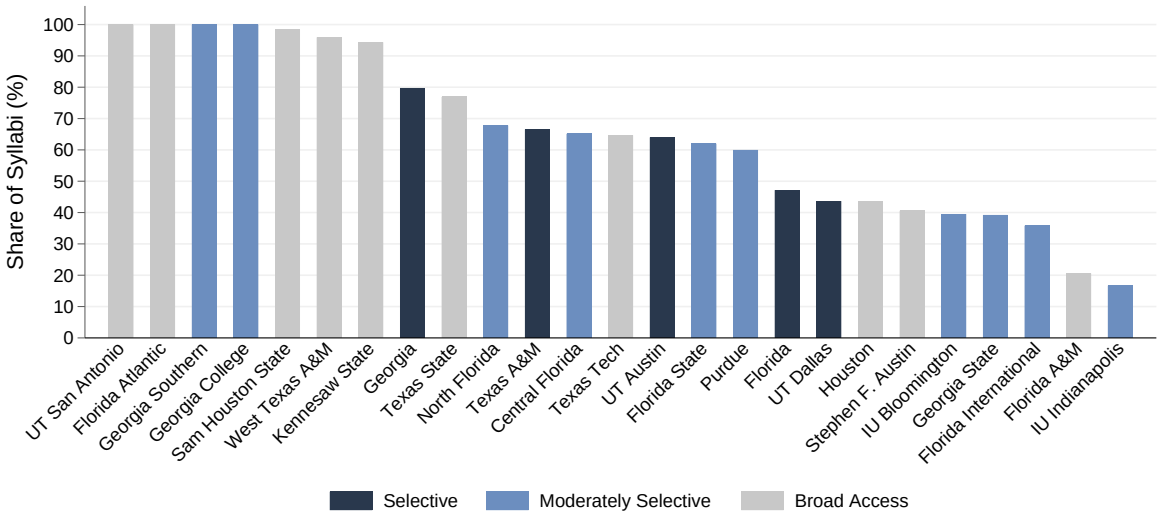
– **Appendix A**: Additional Figures and Tables . . . . . p. A2

– **Appendix B**: Data Construction . . . . . p. A54

– **Appendix C**: Classification Pipeline Details . . . . . p. A88

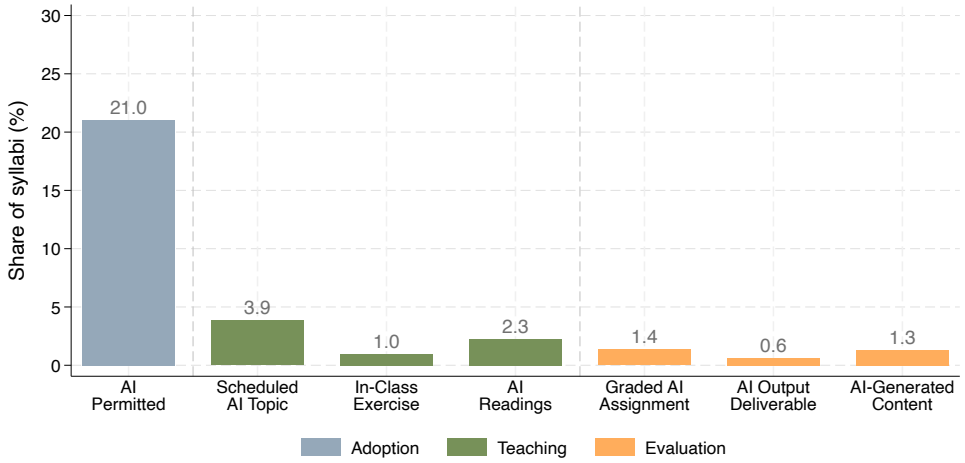
Appendix A: Additional Figures and Tables

Figure A1: AI Mentions by University



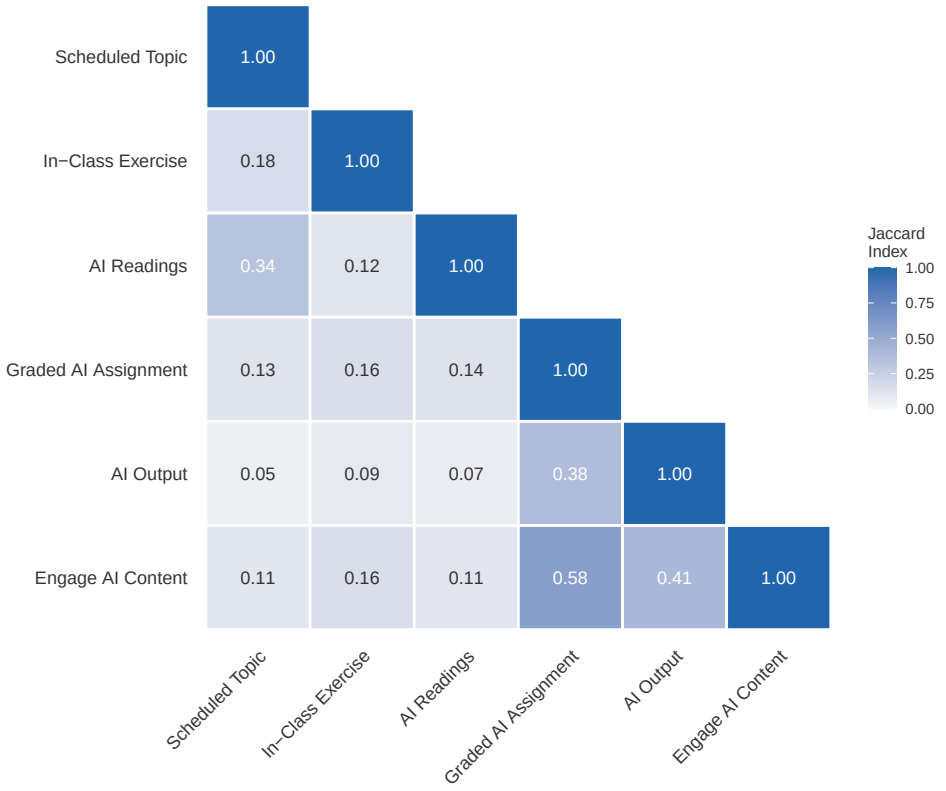
Note: Figure A1 shows the Spring 2026 prevalence of AI mentions by university for the 25 universities in the main sample, ordered from highest to lowest. The bars report the share of syllabi containing at least one AI-related keyword in the full syllabus text.

Figure A2: Pooled Prevalence of AI Policy and Classroom Integration Measures



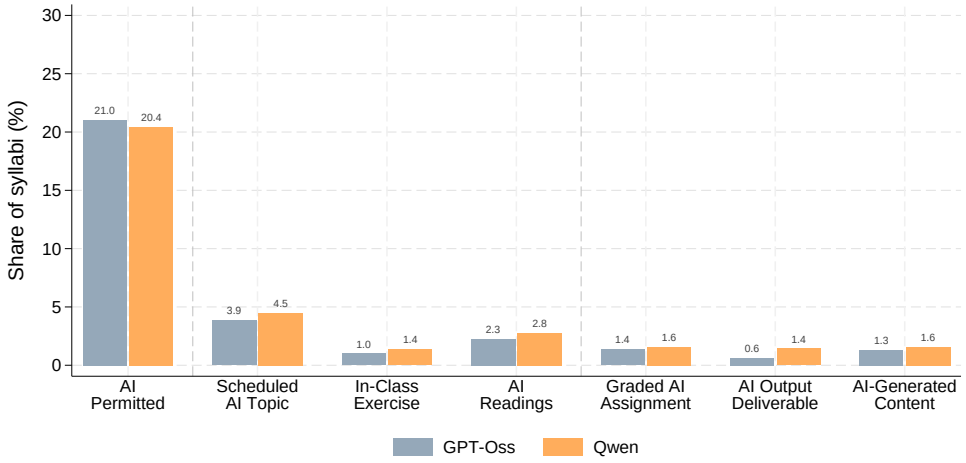
Note: Figure A2 shows the pooled prevalence of AI policy, teaching integration, and evaluation integration measures across the 25 universities for Fall 2024 through Spring 2026. The bars report the share of syllabi with each measure. In the pooled sample, 51.5% of syllabi include an AI policy, 4.8% display any teaching integration, and 1.7% display any evaluation integration.

Figure A3: Jaccard Overlap Across Classroom Integration Measures



Note: Figure A3 shows the Jaccard similarity matrix across the six classroom integration measures for the 25 universities, Fall 2024 through Spring 2026. Each off-diagonal entry reports $|A \cap B|/|A \cup B|$, the share of syllabi positive on either measure that are positive on both. Low cross-domain values indicate that the teaching and evaluation measures capture largely distinct dimensions of classroom integration.

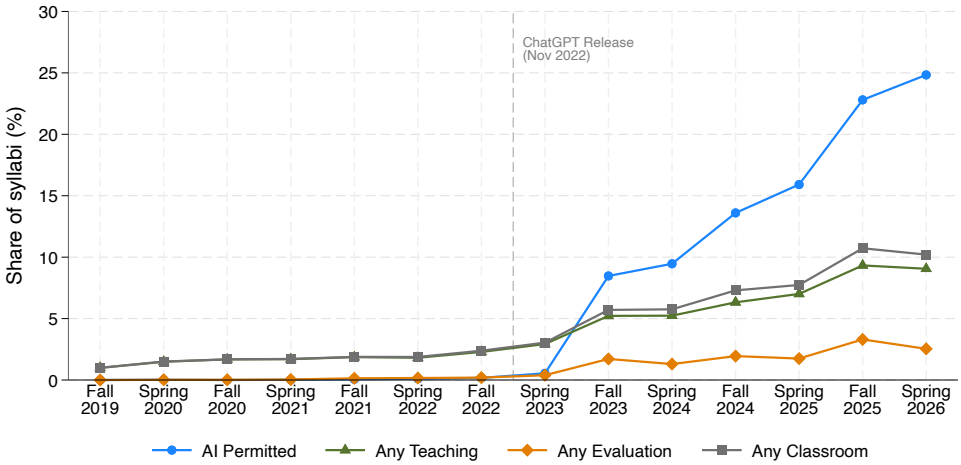
Figure A4: Prevalence Across GPT-Oss and Qwen



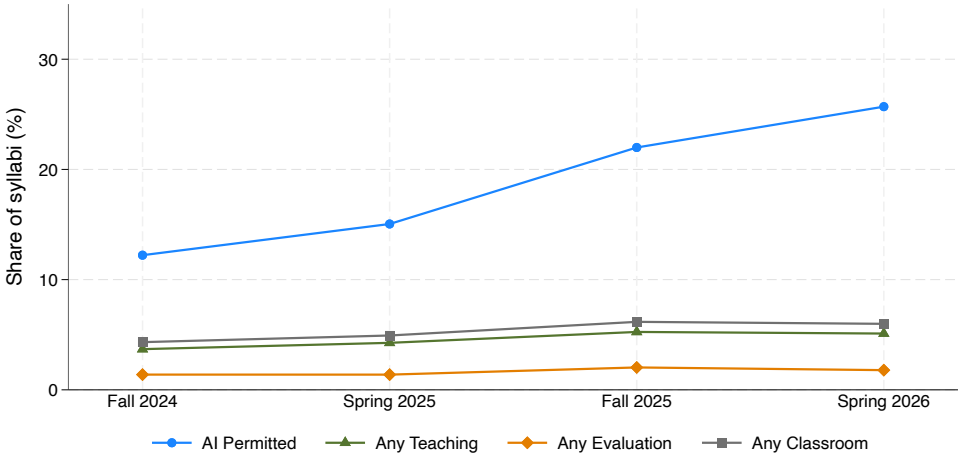
Note: Figure A4 shows the pooled prevalence of each AI integration measure under the two base models, GPT-Oss and Qwen, for the 25 universities across Fall 2024 through Spring 2026.

Figure A5: AI Policy and Classroom Integration Over Time, Full Samples

(a) UT Austin and UT Dallas (full sample)



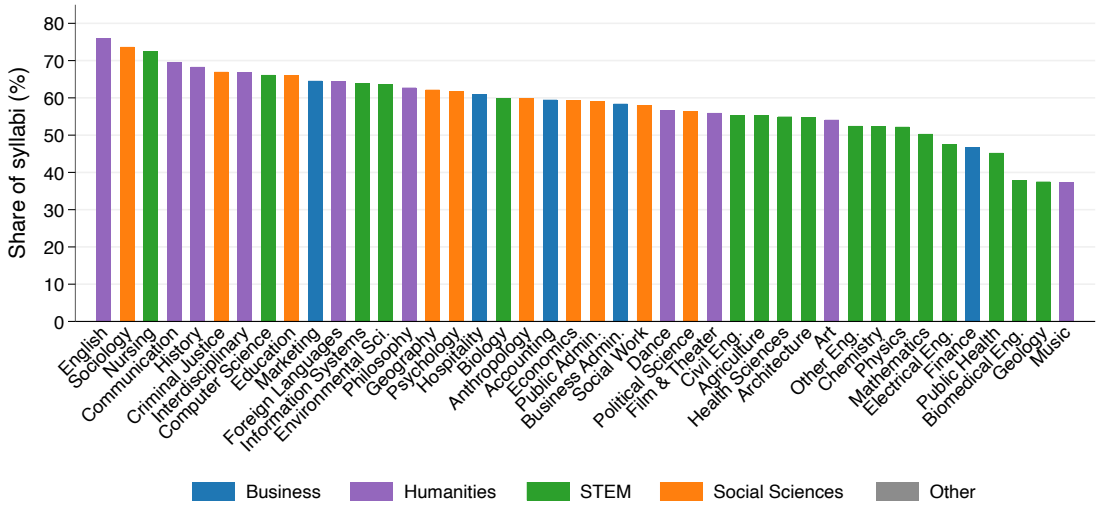
(b) Cross-sectional sample (full sample)



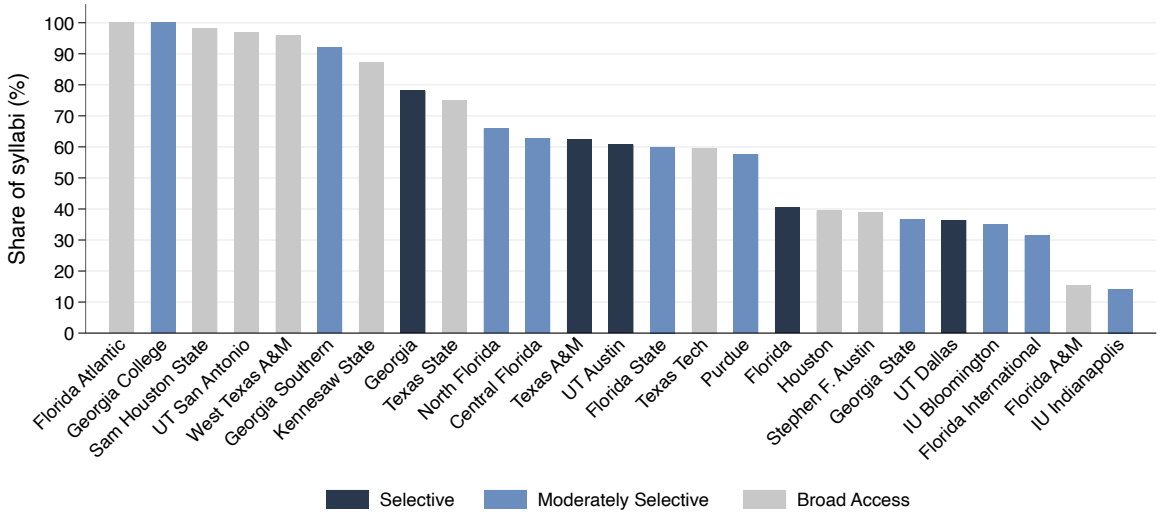
Note: Figure A5 presents the corresponding over-time evidence in the unbalanced samples. Panel (a) includes all syllabi from UT Austin and UT Dallas from Fall 2019 through Spring 2026, while panel (b) includes all 25 universities across the four semesters. The same four outcomes as in Figure 2 of the main text are plotted in each panel.

Figure A6: AI Policy Prevalence Across Majors and Universities

(a) By Major



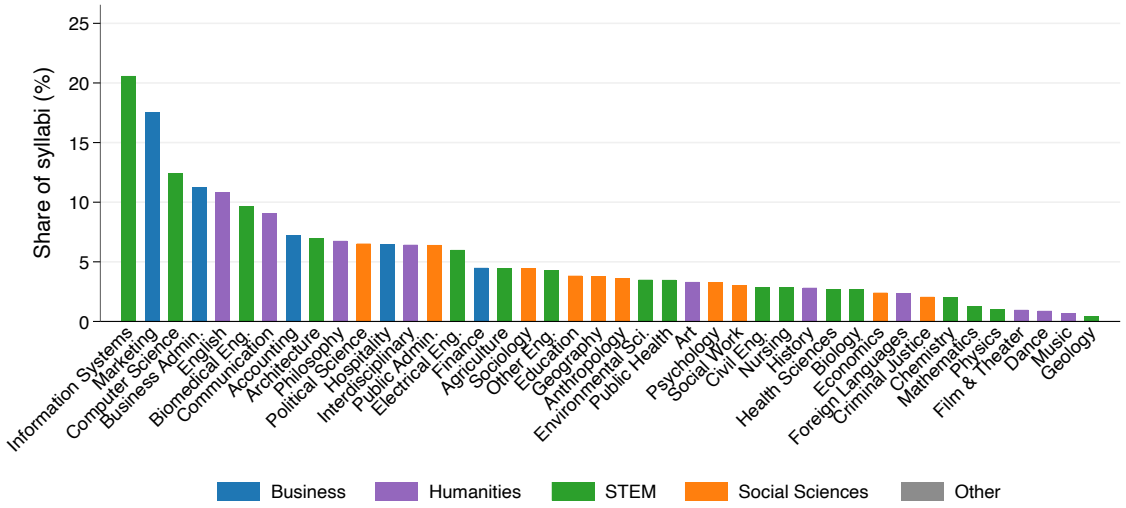
(b) By University



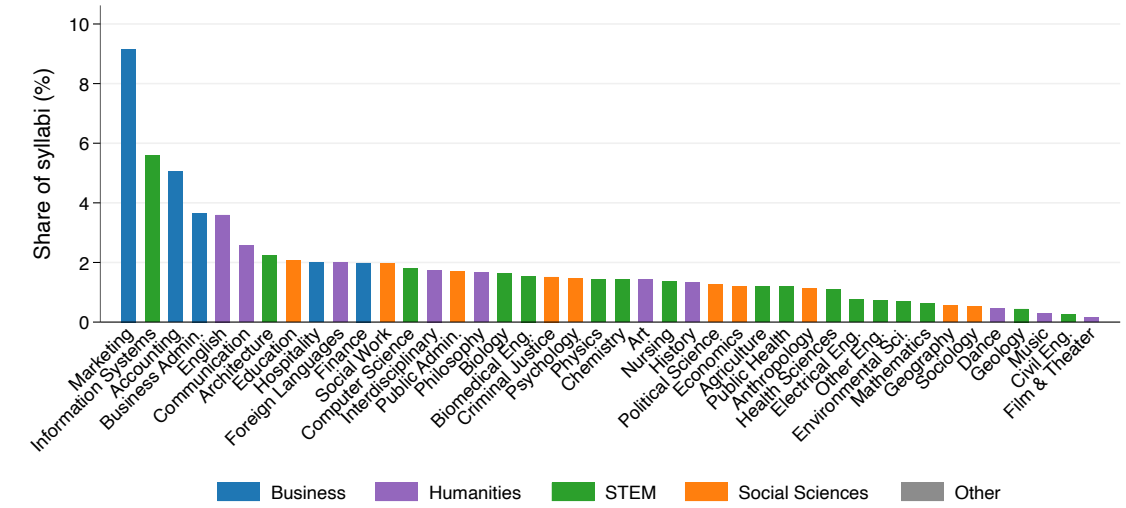
Note: Figure A6 presents the share of syllabi containing an AI policy statement in Spring 2026. Panel (a) shows prevalence across 42 harmonized majors, with bars colored by broad field (Business, Humanities, STEM, Social Sciences, Other). Panel (b) shows prevalence across the 25 universities, ordered from highest to lowest, with bars colored by selectivity tier (Selective, Moderately Selective, Broad Access).

Figure A7: Teaching and Evaluation AI Integration by Major

(a) Any Teaching



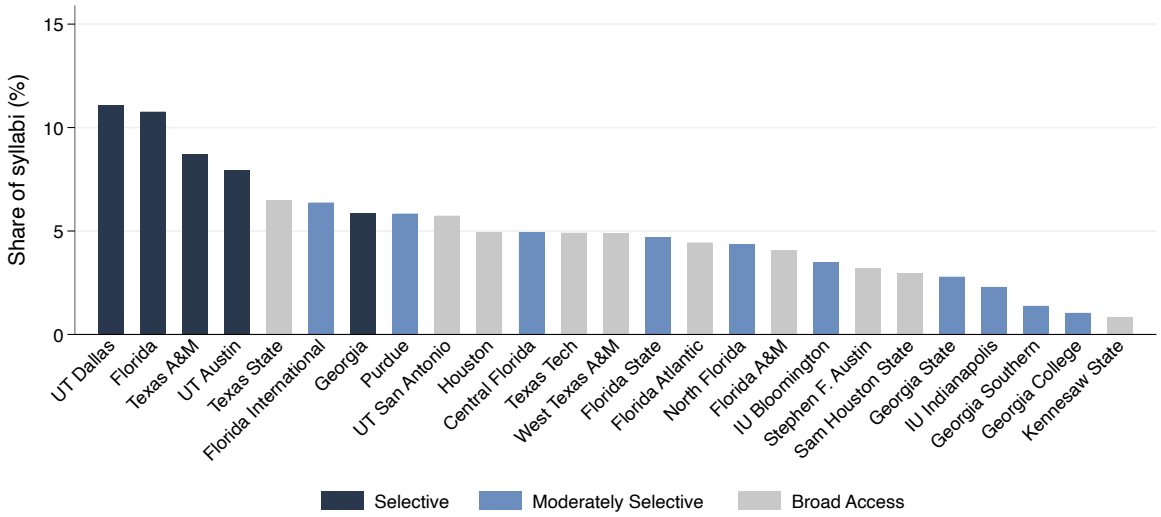
(b) Any Evaluation



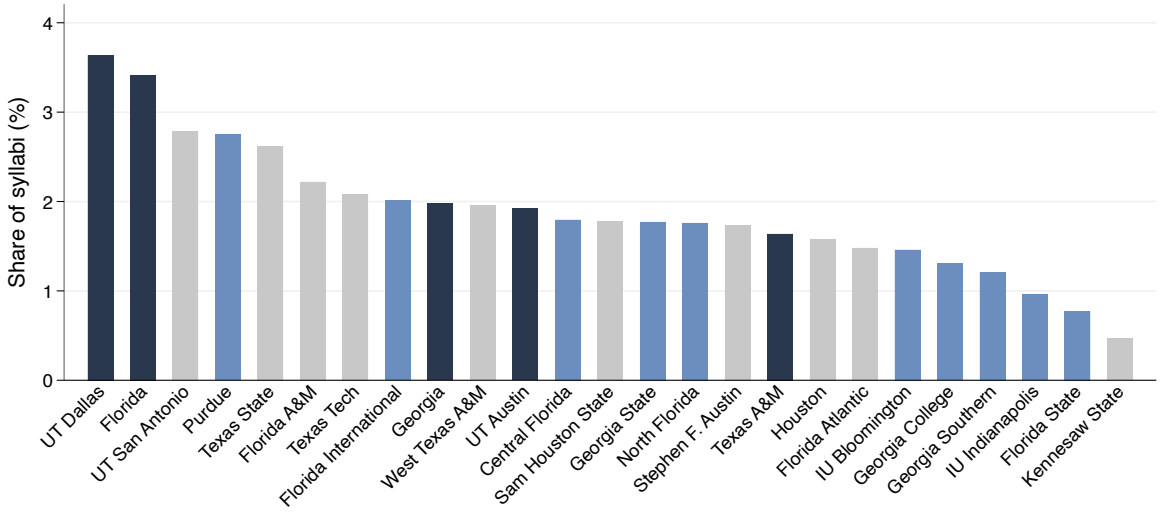
Note: Figure A7 presents the share of syllabi with AI teaching integration (panel a) and AI evaluation integration (panel b) across 42 harmonized majors in Spring 2026. Bars are colored by broad field (Business, Humanities, STEM, Social Sciences, Other).

Figure A8: Teaching and Evaluation AI Integration by University

(a) Any Teaching

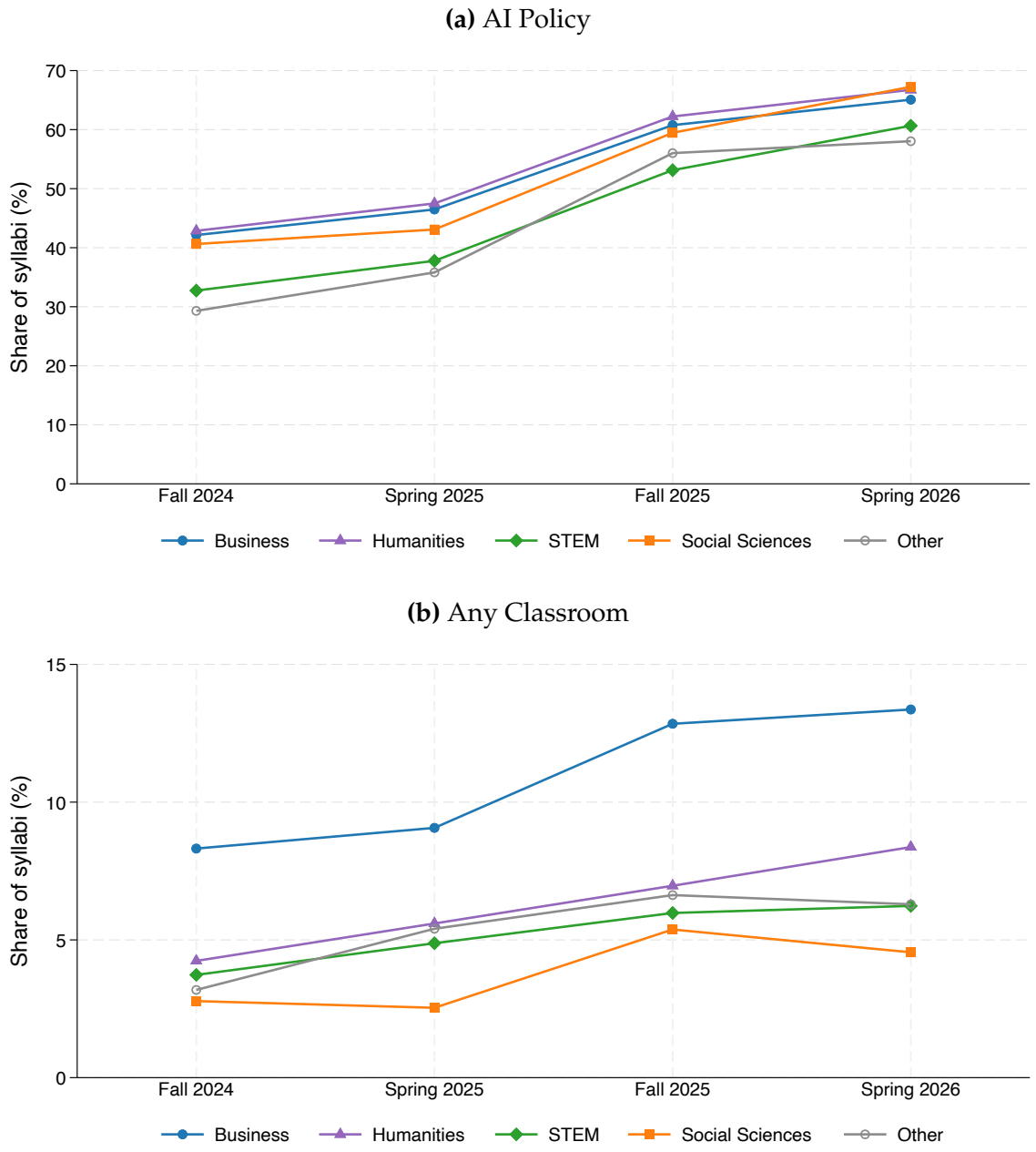


(b) Any Evaluation



Note: Figure A8 presents the share of syllabi with AI teaching integration (panel a) and AI evaluation integration (panel b) across the 25 universities in Spring 2026, ordered from highest to lowest prevalence. Bars are colored by selectivity tier (Selective, Moderately Selective, Broad Access).

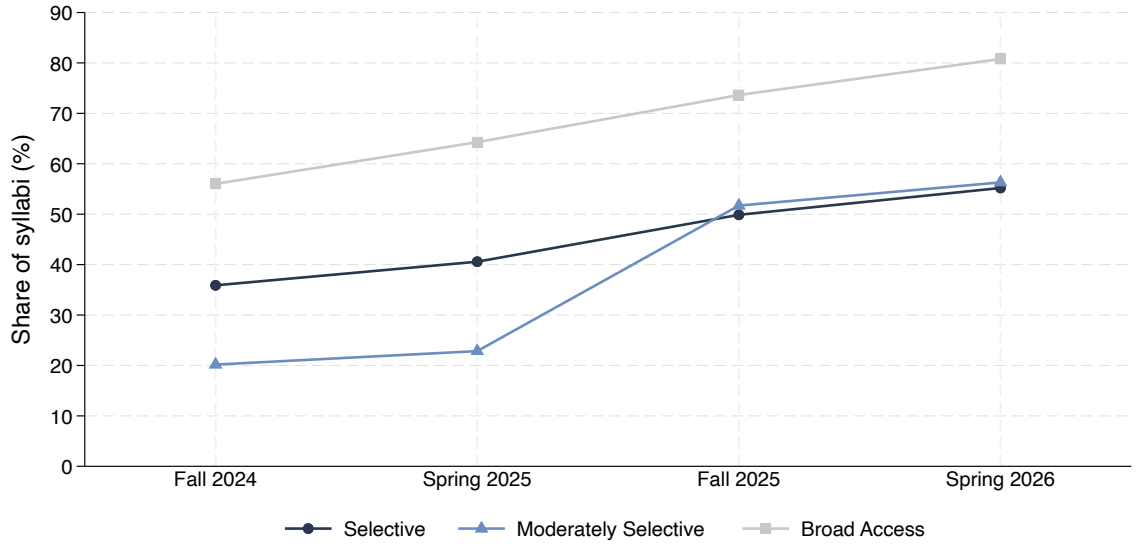
Figure A9: AI Integration by Broad Field, Balanced Panel



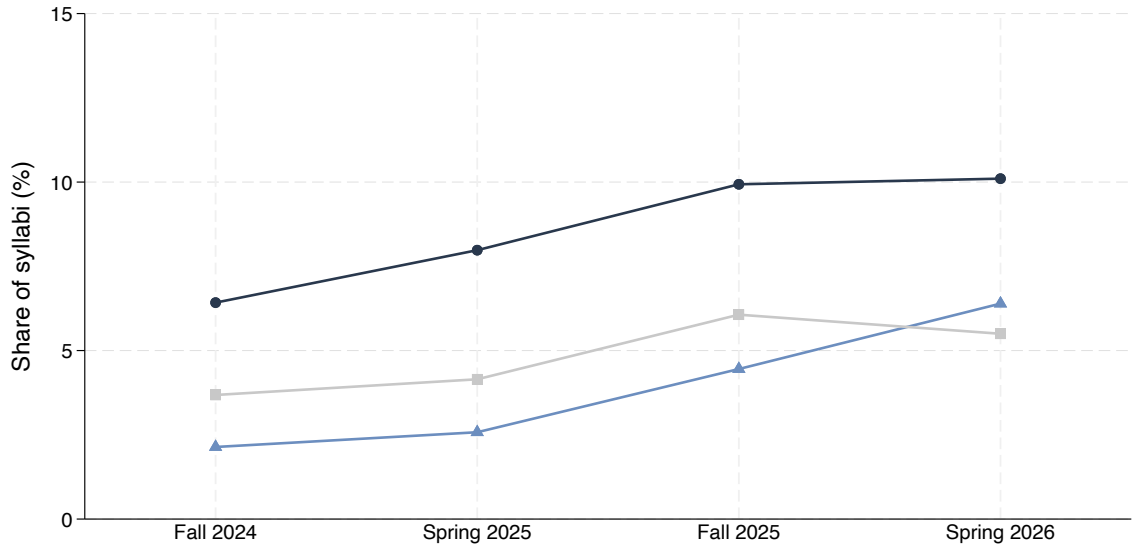
Note: Figure A9 presents the evolution of AI policy prevalence (panel a) and classroom AI integration (panel b) across the five broad field groups over four semesters. The sample is limited to courses observed in at least one term in both 2024–2025 and 2025–2026 across the 25 universities.

Figure A10: AI Integration by Selectivity Tier, Balanced Panel

(a) AI Policy

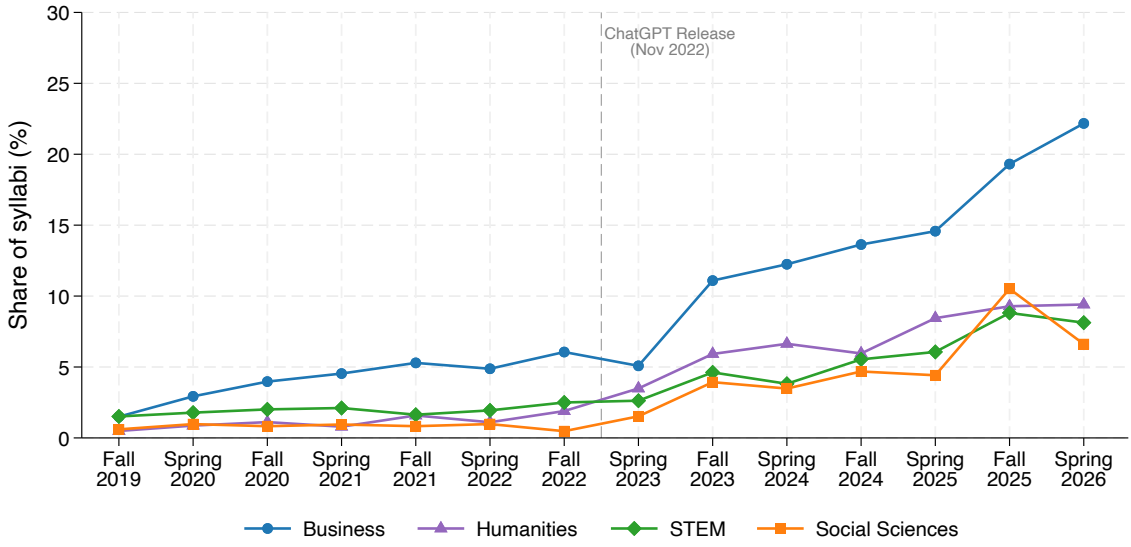


(b) Any Classroom



Note: Figure A10 presents the evolution of AI policy prevalence (panel a) and classroom AI integration (panel b) across the three selectivity tiers over four semesters. The sample is limited to courses observed in at least one term in both 2024–2025 and 2025–2026 across the 25 universities.

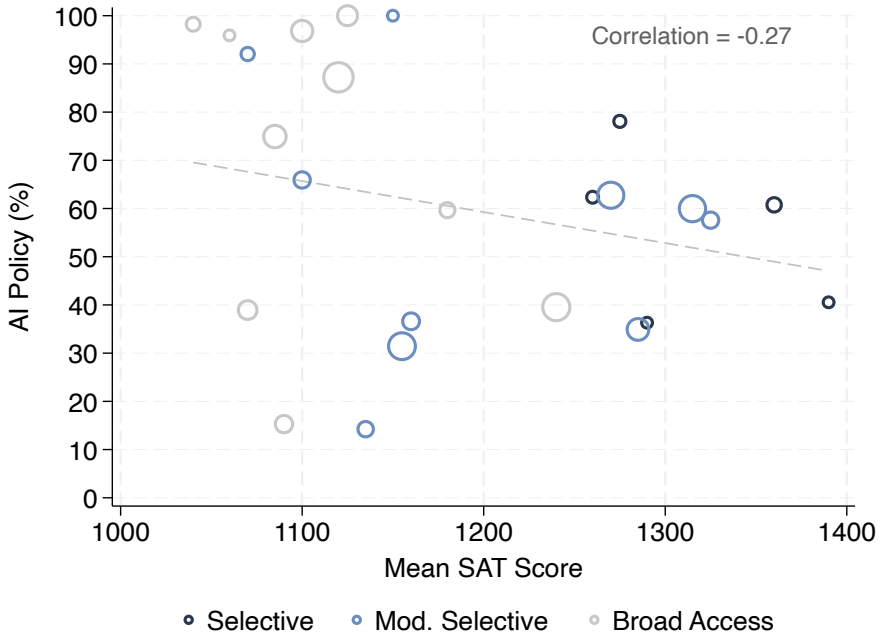
Figure A11: Classroom AI Integration by Broad Field, Longitudinal



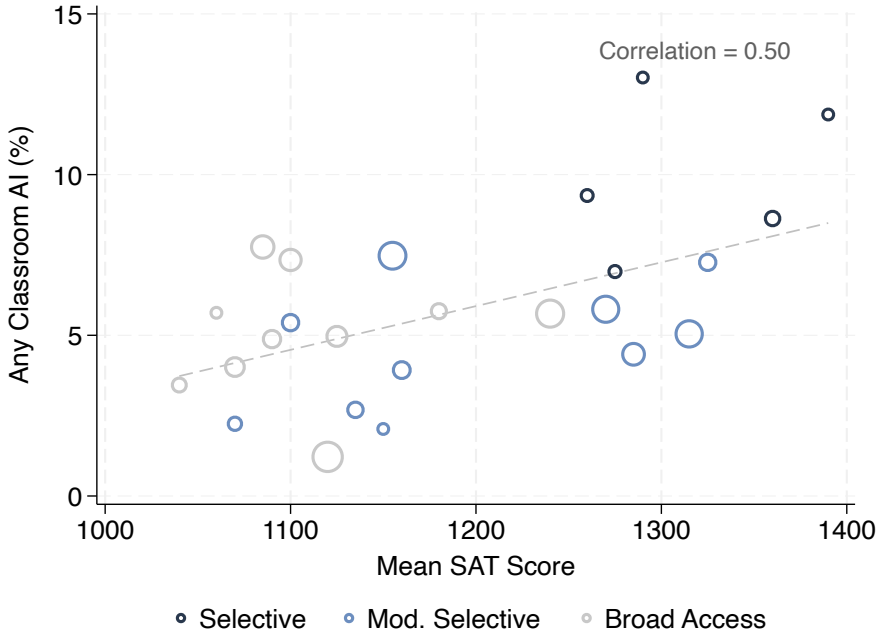
Note: Figure A11 presents the evolution of classroom AI integration across four broad field groups (Business, Humanities, STEM, and Social Sciences), excluding Other. Each line connects term-level means for courses in the balanced panel. The vertical dashed line marks the launch of ChatGPT in November 2022. The longitudinal sample covers UT Austin and UT Dallas from Fall 2019 through Spring 2026.

Figure A12: AI Integration and Mean SAT Score

(a) AI Policy

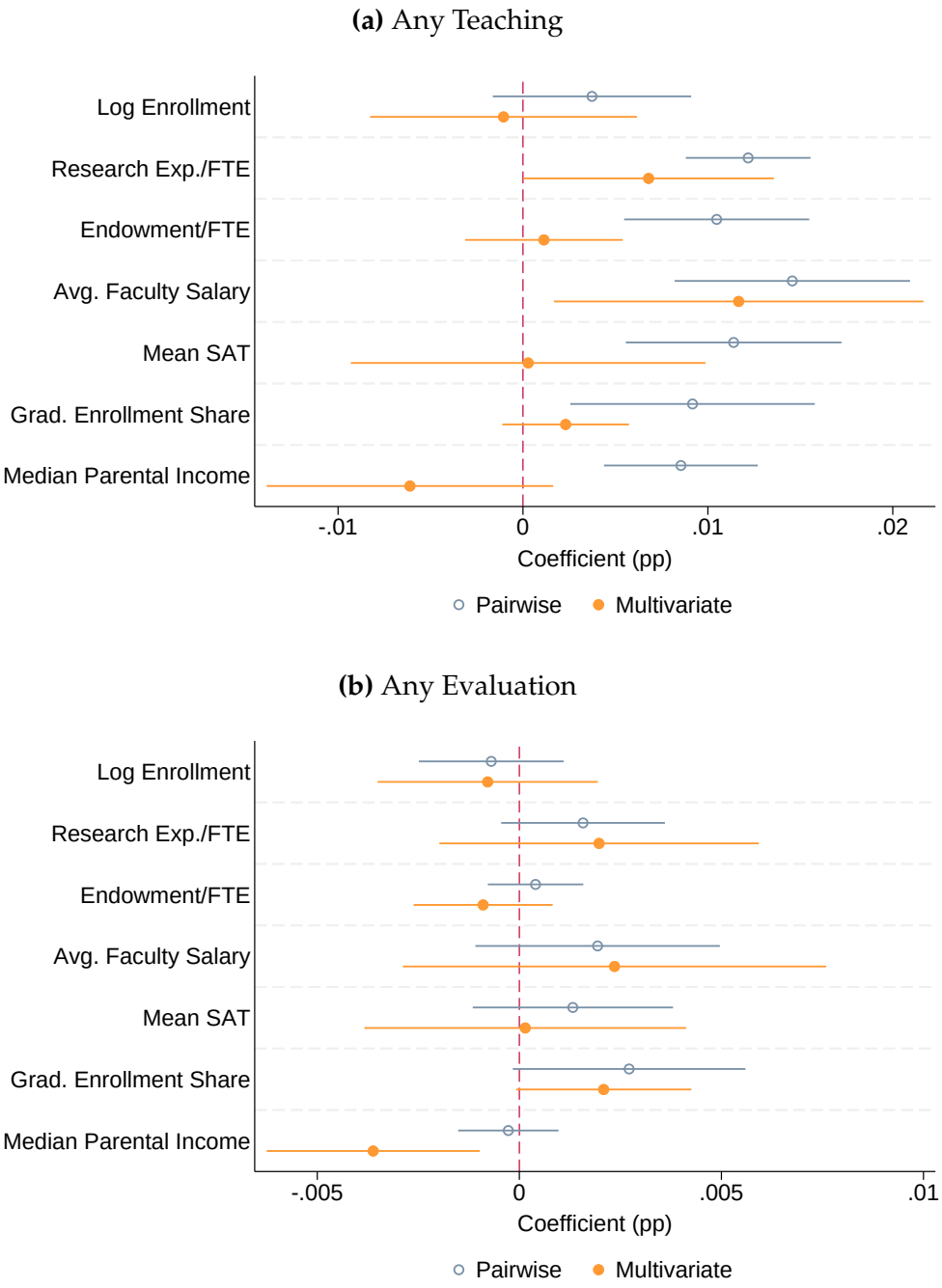


(b) Any Classroom



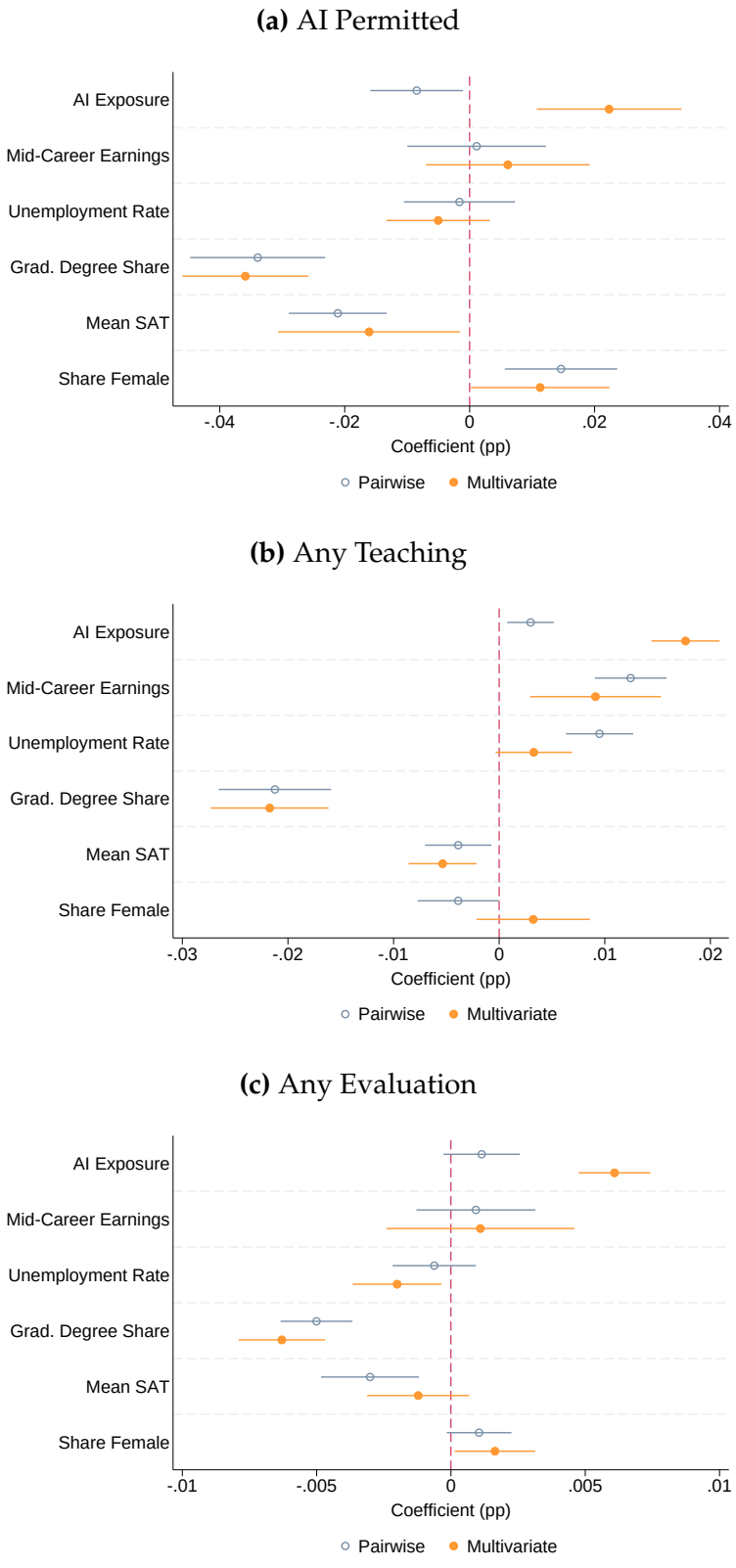
Note: Figure A12 presents university-level scatter plots of AI integration against mean SAT score in Spring 2026. Panel (a) plots AI policy prevalence against mean SAT; panel (b) plots classroom AI integration against mean SAT. Each point represents one of the 25 universities.

Figure A13: University Covariates: Teaching and Evaluation AI Integration



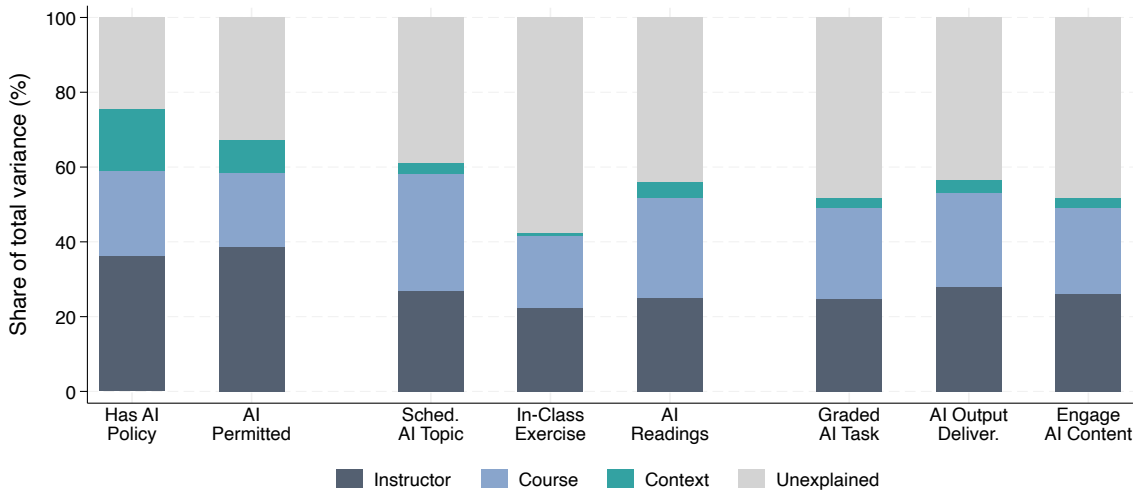
Note: Figure A13 presents coefficient estimates from regressions of AI teaching integration (panel a) and AI evaluation integration (panel b) on seven university-level covariates. Circles report pairwise estimates; diamonds report multivariate estimates including all covariates simultaneously. Bars represent 90% confidence intervals. Fixed effects: major × course level × term. Standard errors clustered at the university level.

Figure A14: Major Covariates: Permitted, Teaching, and Evaluation AI Integration



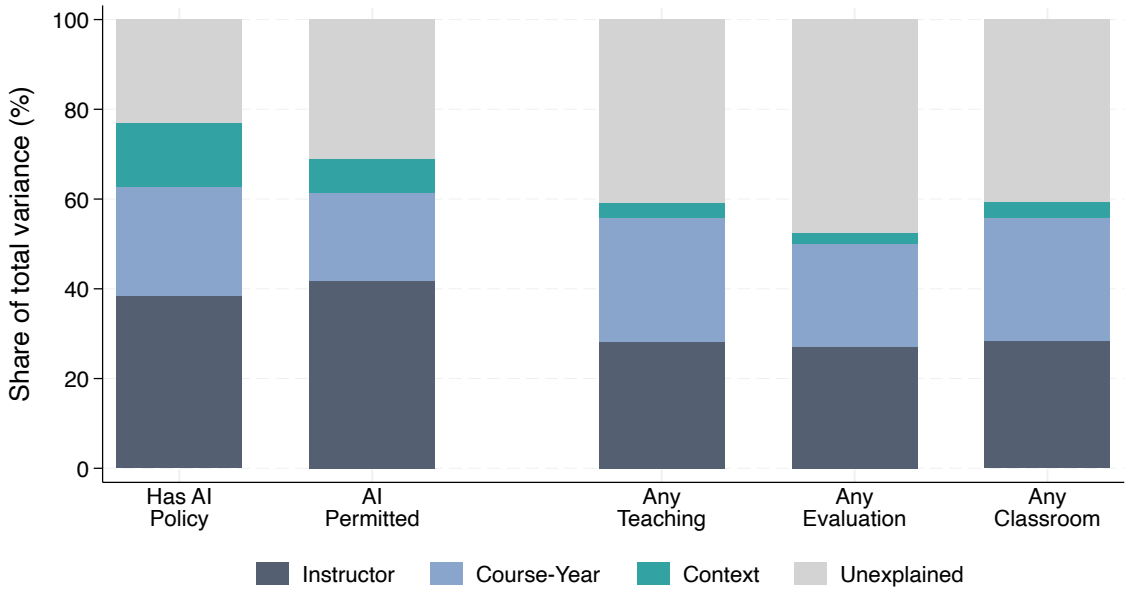
Note: Figure A14 presents coefficient estimates from regressions of AI being permitted (panel a), AI teaching integration (panel b), and AI evaluation integration (panel c) on six major-level covariates. Circles report pairwise estimates; diamonds report multivariate estimates including all covariates simultaneously. Bars represent 90% confidence intervals. Fixed effects: course level × term and university. Standard errors clustered at the university level.

Figure A15: Variance Decomposition for the Specific AI Outcomes



Note: Figure A15 shows the three-way variance decomposition for the eight specific AI outcomes. Each bar reports the shares of total variance attributed to instructor fixed effects, local course fixed effects, university-by-major-by-level-by-term fixed effects, and the unexplained residual.

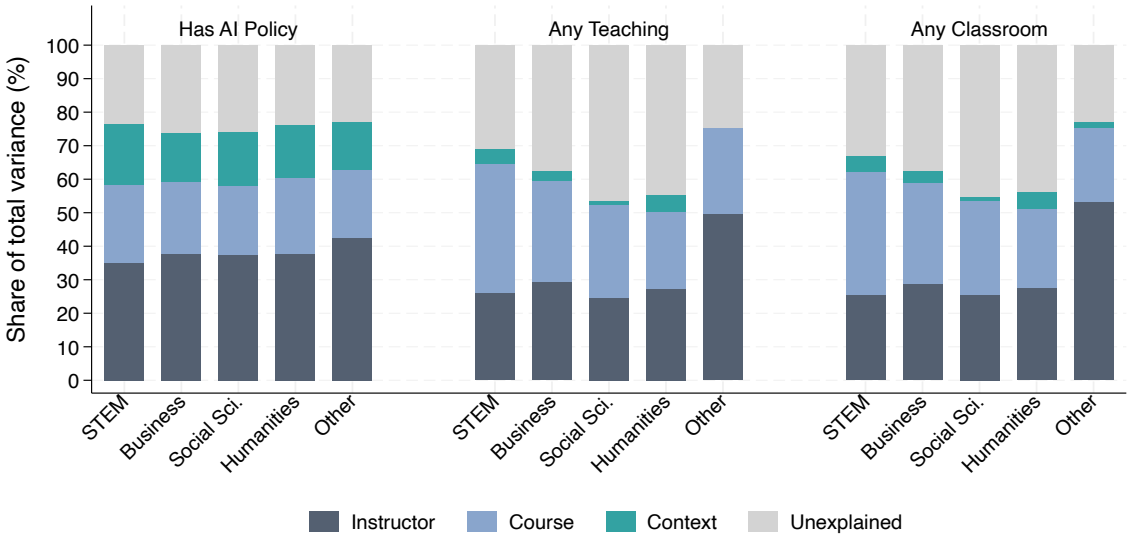
Figure A16: Variance Decomposition with Course-Year Fixed Effects



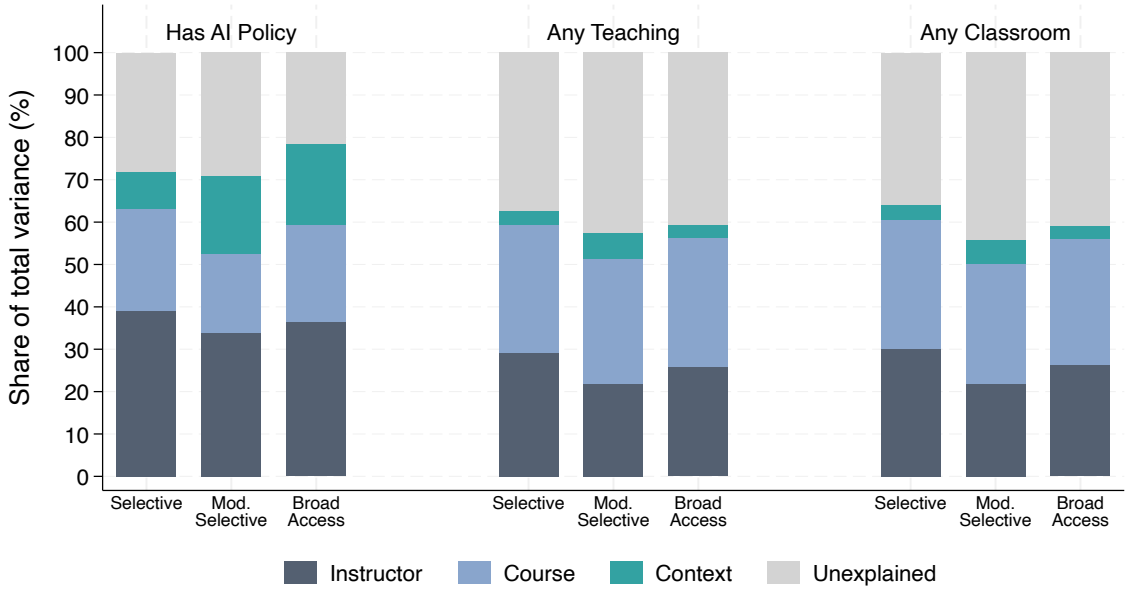
Note: Figure A16 shows the three-way variance decomposition for the five main AI outcomes using a more stringent specification that replaces local course fixed effects with local course-by-year fixed effects. Each bar reports the shares of total variance attributed to instructor fixed effects, course-year fixed effects, the broader context component, and the unexplained residual.

Figure A17: Variance Decomposition by Broad Field and Selectivity

(a) By Broad Major Group

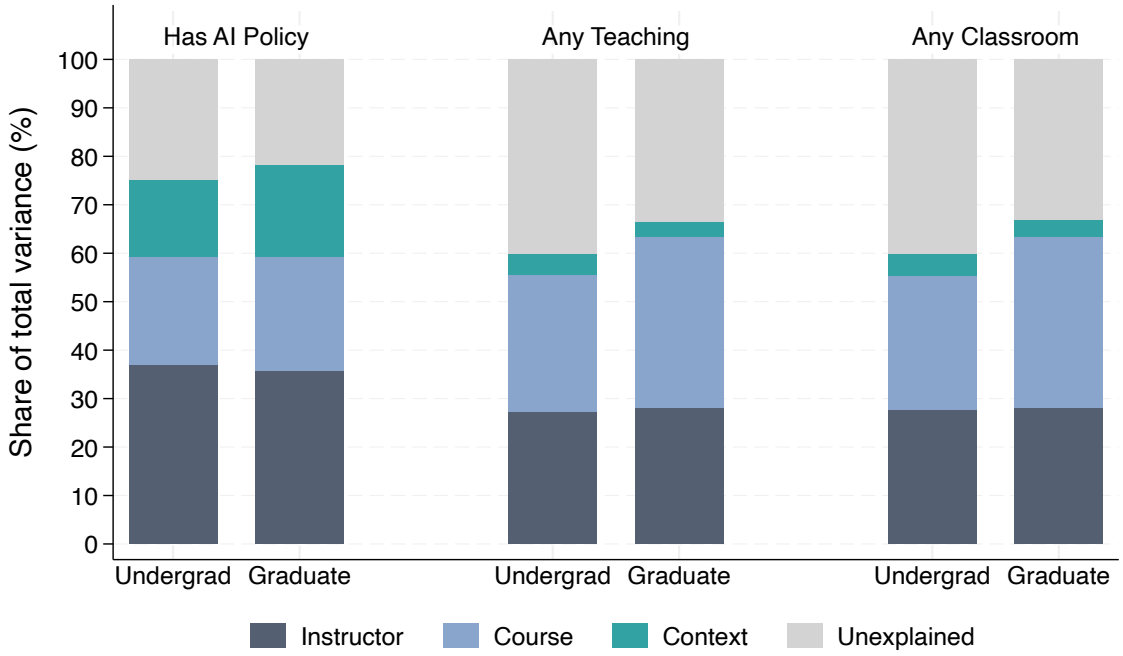


(b) By University Selectivity



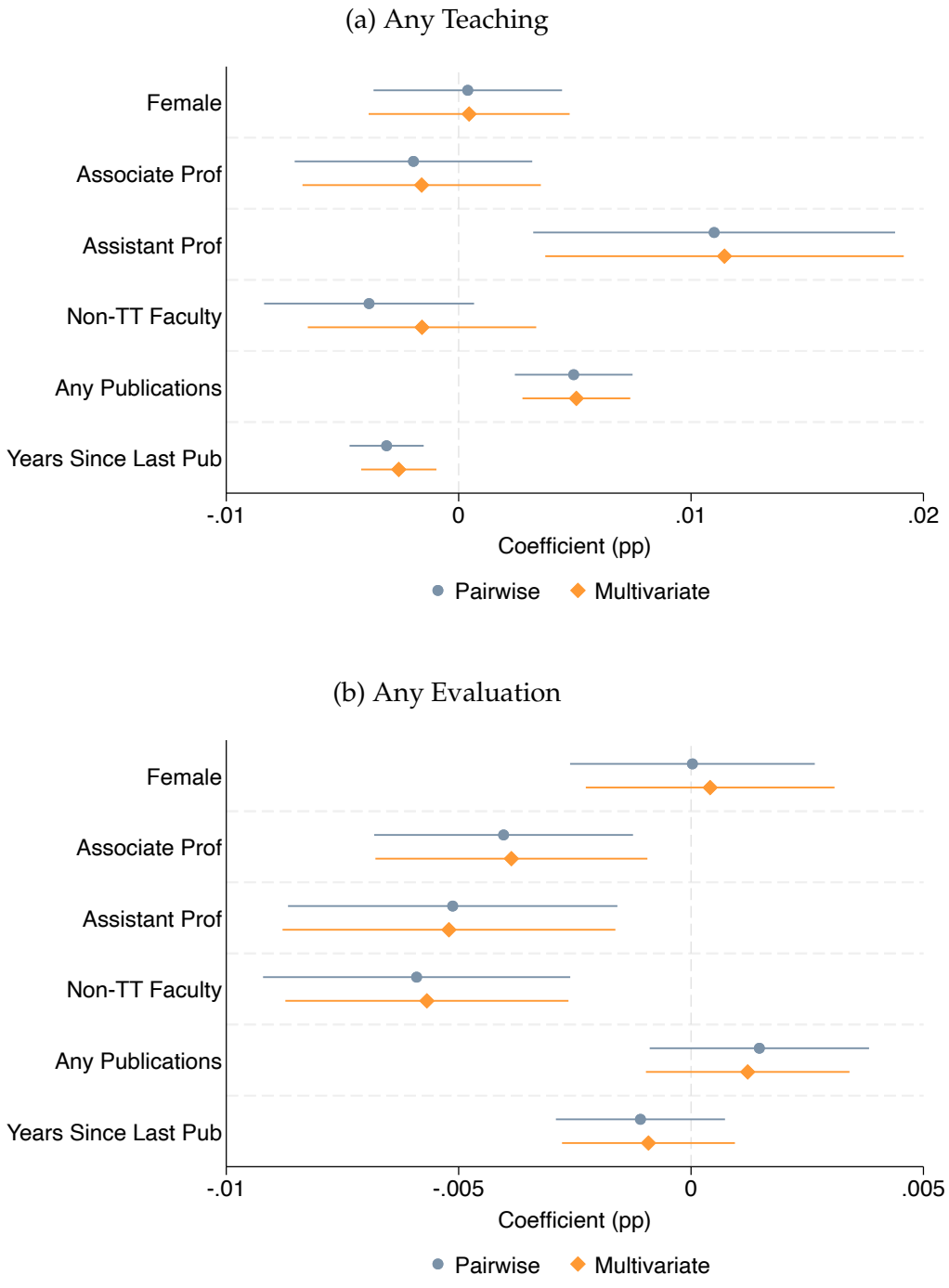
Note: Figure A17 shows the baseline three-way variance decomposition for Has AI Policy, Any Teaching, and Any Classroom across subsamples. Panel (a) reports the decomposition separately across the five broad fields. Panel (b) reports the decomposition separately across the three selectivity tiers. Each bar reports the shares of total variance attributed to instructor fixed effects, local course fixed effects, the broader context component, and the unexplained residual.

Figure A18: Variance Decomposition by Course Level



Note: Figure A18 shows the baseline three-way variance decomposition for Has AI Policy, Any Teaching, and Any Classroom separately for undergraduate and graduate courses. Each bar reports the shares of total variance attributed to instructor fixed effects, local course fixed effects, the broader context component, and the unexplained residual.

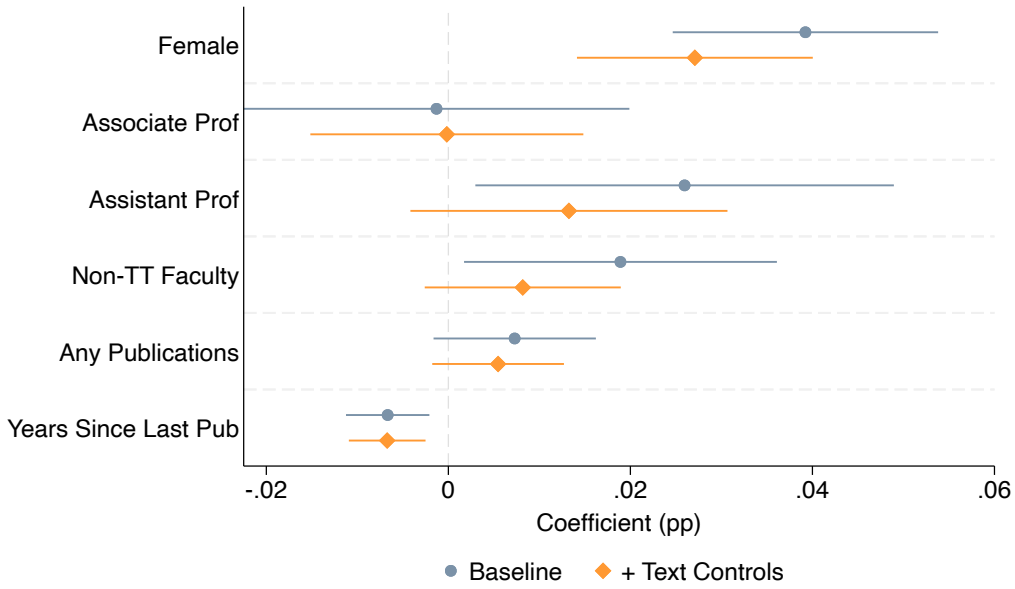
Figure A19: Determinants of Teaching and Evaluation Outcomes



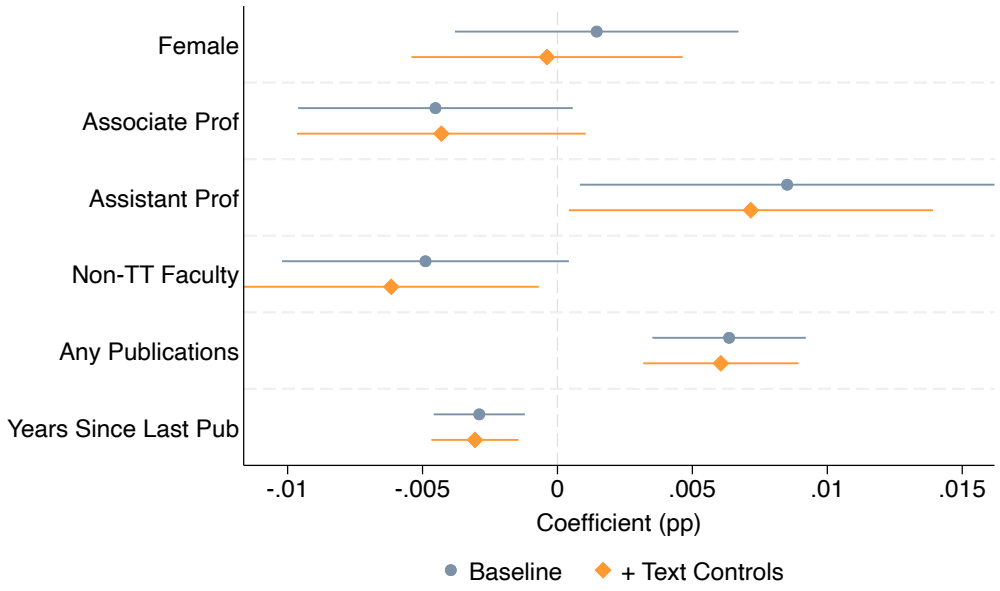
Note: Figure A19 shows coefficient estimates from regressions of AI teaching and evaluation outcomes on instructor-level covariates. Circles (blue) report pairwise estimates; diamonds (orange) report multivariate estimates. Bars represent 90% confidence intervals. The specification absorbs local course and university×major×level×term fixed effects. Standard errors are clustered at the university level. The sample covers the 25 primary universities from Fall 2024 through Spring 2026.

Figure A20: Robustness to Textual Controls

(a) AI Policy



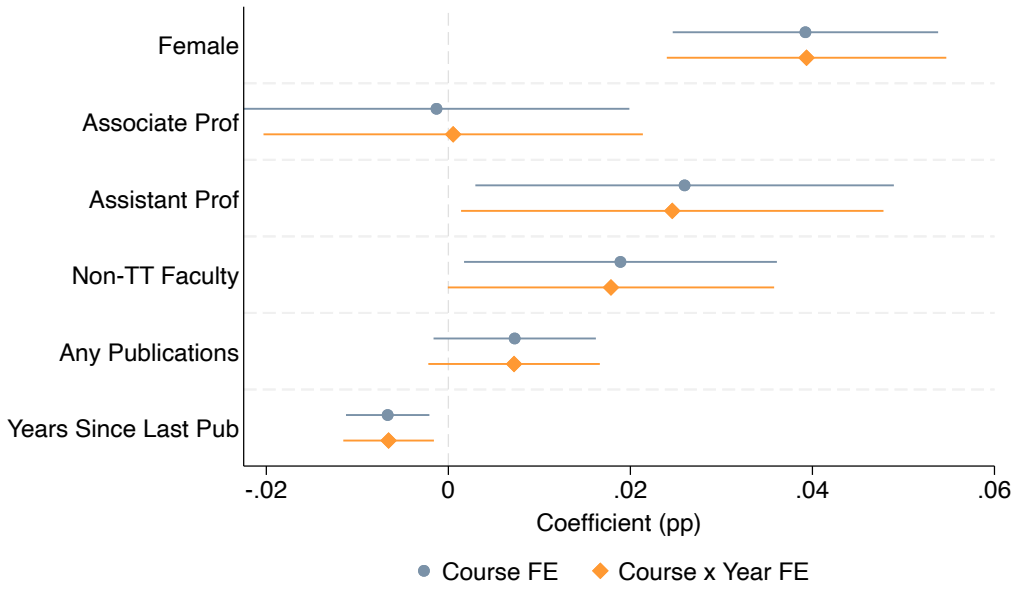
(b) Any Classroom



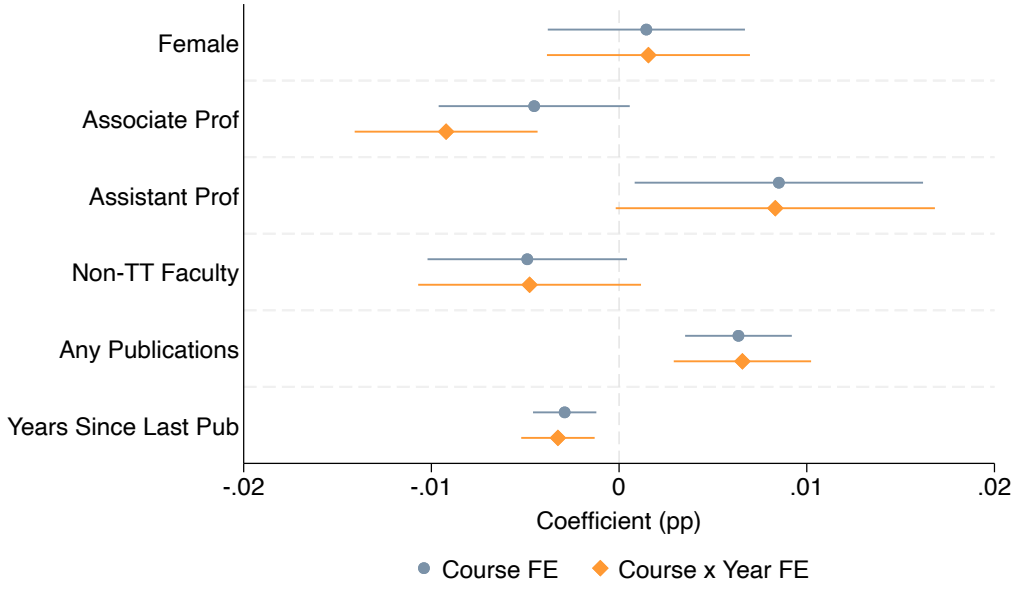
Note: Figure A20 compares the multivariate instructor-characteristic coefficients with and without syllabus text controls. Circles (blue) report the baseline specification; diamonds (orange) add nine text-based measures (Flesch–Kincaid grade level, lexical diversity, average sentence length, modal prescriptiveness, penalty and supportive language density, log word count, late-work and attendance policy indicators). Bars represent 90% confidence intervals. Standard errors are clustered at the university level. The sample covers the 25 primary universities from Fall 2024 through Spring 2026.

Figure A21: Robustness to Course×Year Fixed Effects

(a) AI Policy

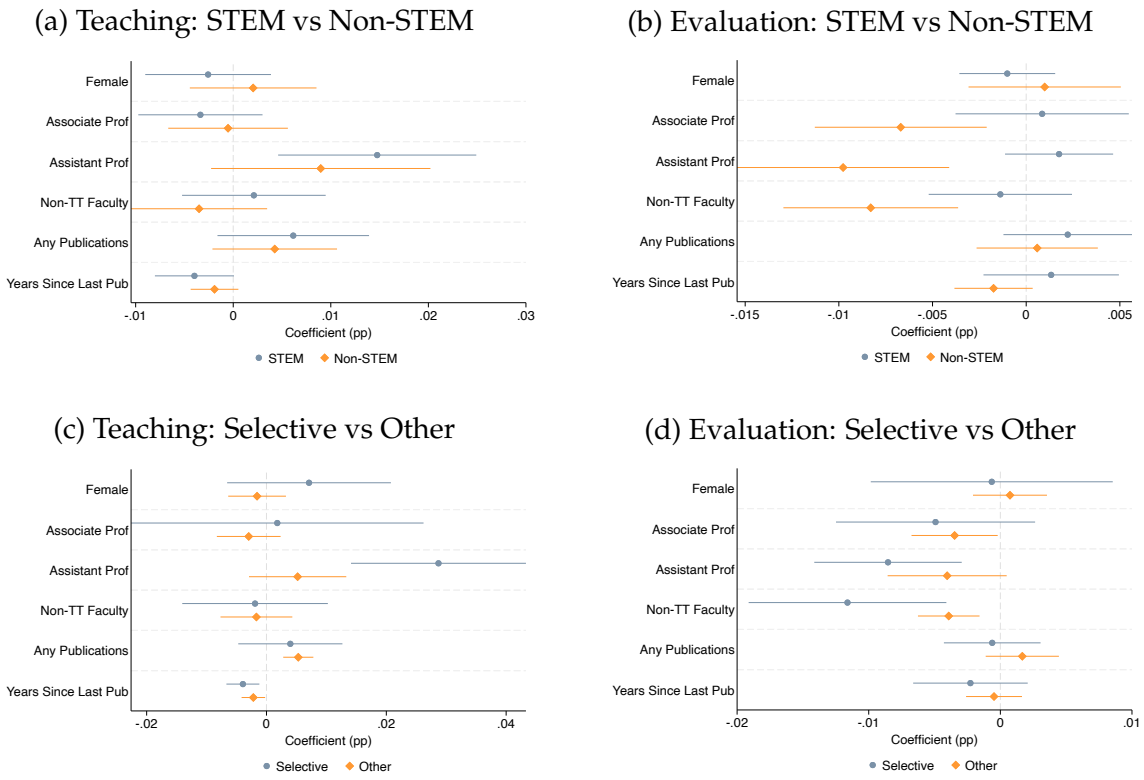


(b) Any Classroom



Note: Figure A21 compares the multivariate instructor-characteristic coefficients under the main specification (course fixed effects) and a more stringent specification replacing course fixed effects with course×year fixed effects. Circles (blue) report the baseline; diamonds (orange) report the course×year specification. Bars represent 90% confidence intervals. Standard errors are clustered at the university level. The sample covers the 25 primary universities from Fall 2024 through Spring 2026.

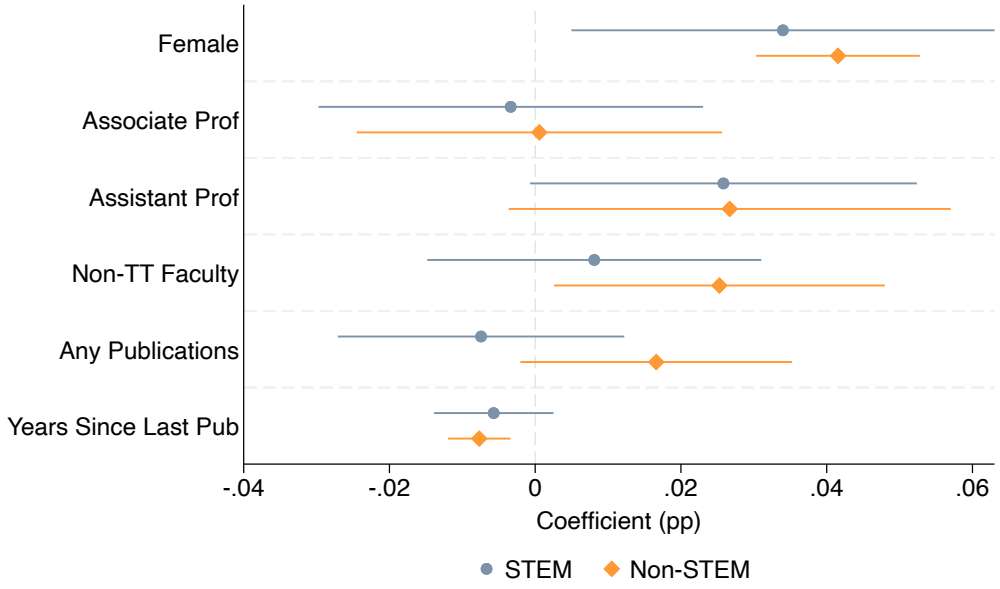
Figure A22: Heterogeneity in Teaching and Evaluation Outcomes



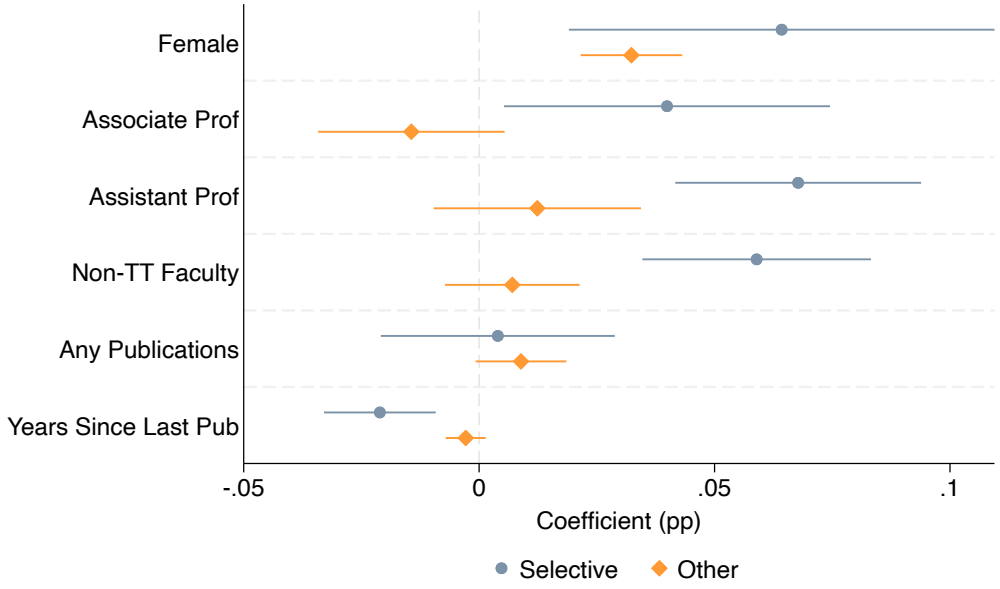
Note: Figure A22 shows heterogeneity in the multivariate instructor-characteristic coefficients for AI teaching and evaluation outcomes. Panels (a)–(b) compare STEM and non-STEM fields; panels (c)–(d) compare selective (Barron's tier 1) and other institutions. Bars represent 90% confidence intervals. Standard errors are clustered at the university level. The sample covers the 25 primary universities from Fall 2024 through Spring 2026.

Figure A23: Heterogeneity in AI Policy

(a) STEM vs Non-STEM



(b) Selective vs Other



Note: Figure A23 shows heterogeneity in the multivariate instructor-characteristic coefficients for AI policy. Panel (a) compares STEM and non-STEM fields; panel (b) compares selective (Barron's tier 1) and other institutions. Bars represent 90% confidence intervals. The specification absorbs local course and university×major×level×term fixed effects. Standard errors are clustered at the university level. The sample covers the 25 primary universities from Fall 2024 through Spring 2026.

Table A1: Number of Syllabi by University and Academic Term

University	**Fall 2024** (1)	**Spring 2025** (2)	**Fall 2025** (3)	**Spring 2026** (4)	**Total** (5)
Texas					
UT Austin	3,344	3,550	3,556	3,382	13,832
UT Dallas	2,001	1,943	2,105	1,901	7,950
UT San Antonio	2,367	2,460	2,427	2,369	9,623
University of Houston	0	0	3,947	3,859	7,806
Texas Tech	1,552	1,325	1,459	1,200	5,536
Texas State	0	0	2,794	2,633	5,427
Stephen F. Austin State	1,344	520	523	1,844	4,231
Texas A&M	0	0	1,211	2,321	3,532
West Texas A&M	616	611	630	614	2,471
Sam Houston State	0	0	488	1,014	1,502
Florida					
University of Central Florida	4,203	4,189	4,456	4,457	17,305
Florida International	1,465	1,613	4,433	4,722	12,233
Florida Atlantic	2,296	1,919	2,312	2,094	8,621
Florida State	0	0	0	4,520	4,520
University of Florida	1,593	1,572	1,821	1,904	6,890
Florida A&M	0	0	1,466	1,577	3,043
University of North Florida	0	0	93	1,761	1,854
Georgia					
Kennesaw State	0	0	2,443	4,516	6,959
University of Georgia	1,153	1,182	1,525	2,377	6,237
Georgia State	0	0	240	1,863	2,103
Georgia College & State University	0	0	637	767	1,404
Georgia Southern	0	0	191	1,157	1,348
Indiana					
IU Bloomington	0	0	3,178	3,084	6,262
Purdue	157	181	2,210	1,747	4,295
IU Indianapolis	0	0	1,667	1,567	3,234
Total	**22,091**	**21,065**	**45,812**	**59,250**	**148,218**

Note: Table A1 shows the number of deduplicated syllabi available for each of the four main terms, separately by university. The counts correspond to the 25-university main sample, with one representative syllabus retained for each course-instructor-term group.

Table A2: Syllabi Text Characteristics by University

University	N	Words	FK Grade	MATTR	Prescriptive	Late Policy	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Texas							
UT Austin	13,832	3,159	16.18	0.732	0.266	0.57	0.57
UT Dallas	7,950	2,359	14.30	0.718	0.295	0.69	0.49
UT San Antonio	9,623	2,812	14.68	0.716	0.254	0.48	0.36
University of Houston	7,806	2,893	13.97	0.721	0.198	0.40	1.00
Texas Tech	5,536	2,832	16.74	0.717	0.287	0.52	0.62
Texas State	5,427	3,116	13.25	0.723	0.280	0.63	0.71
Stephen F. Austin State	4,231	2,799	14.71	0.711	0.390	0.45	0.71
Texas A&M	3,532	3,498	16.42	0.732	0.219	0.99	1.00
West Texas A&M	2,471	3,195	15.15	0.722	0.329	0.66	0.67
Sam Houston State	1,502	2,801	15.83	0.719	0.299	0.97	1.00
Florida							
University of Central Florida	17,305	3,001	15.08	0.722	0.216	0.76	0.40
Florida International	12,233	2,538	15.24	0.712	0.253	0.55	0.35
Florida Atlantic	8,621	3,519	15.91	0.711	0.241	1.00	1.00
Florida State	4,520	2,044	16.69	0.727	0.205	0.20	1.00
University of Florida	6,890	2,841	15.57	0.726	0.298	0.55	0.68
Florida A&M	3,043	3,878	17.93	0.721	0.239	0.38	1.00
University of North Florida	1,854	2,511	14.93	0.716	0.358	0.66	0.40
Georgia							
Kennesaw State	6,959	3,260	16.79	0.699	0.311	0.41	0.88
University of Georgia	6,237	2,345	17.40	0.730	0.287	0.51	0.52
Georgia State	2,103	2,334	103.57	0.713	0.288	0.35	0.82
Georgia College & State University	1,404	1,874	15.57	0.704	0.384	0.32	1.00
Georgia Southern	1,348	1,560	17.52	0.709	0.212	0.26	1.00
Indiana							
IU Bloomington	6,262	1,621	14.97	0.725	0.229	0.33	0.44
Purdue	4,295	2,483	13.88	0.727	0.241	0.51	0.60
IU Indianapolis	3,234	1,256	16.86	0.710	0.230	0.35	0.18

Note: Table A2 shows average syllabus text characteristics by university for the four-term main sample. Words denotes the cleaned document word count, FK Grade is the Flesch-Kincaid readability score, and MATTR is the moving-average type-token ratio. Prescriptive is the share of strong prescriptive modal language in the syllabus text, while Late Policy and Attendance indicate whether the syllabus contains those policies.

Table A3: Instructor Characteristic Coverage by University

University	Instructors (1)	Gender (2)	Research Profile: Rank (3)	Research Profile: OpenAlex (4)	Research Profile: *h*-Index (5)	Research Profile: Any Pub. (6)	Research Profile: Hire Year (7)	Education: PhD Year (8)	Education: PhD Inst. (9)	Education: BA Inst. (10)
Texas										
UT Austin	3,812	93.8	72.1	45.7	42.0	46.1	60.1	21.2	35.2	18.6
UT Dallas	2,122	79.7	67.6	34.1	30.7	34.4	36.4	31.0	36.8	10.8
UT San Antonio	2,511	91.5	84.4	31.5	29.7	31.9	58.3	15.7	30.1	14.7
UH	2,531	91.2	74.6	38.4	35.1	38.8	58.1	25.5	32.0	15.9
Texas State	1,933	92.9	74.2	31.8	28.8	32.2	62.5	14.1	17.8	11.3
Texas A&M	1,888	78.0	64.7	42.2	39.0	42.8	53.3	32.9	36.0	27.0
West Texas A&M	376	94.4	79.5	38.6	35.1	38.6	—	9.8	12.2	7.2
Stephen F. Austin	777	95.2	77.5	11.8	10.6	11.8	—	12.2	12.9	9.7
Sam Houston State	600	87.2	74.3	36.5	32.7	36.5	66.2	15.0	35.5	19.0
Texas Tech	1,657	90.3	71.9	41.0	36.3	41.3	64.1	16.7	23.1	13.4
Florida										
Central Florida	3,956	83.8	74.6	38.1	35.1	38.7	—	16.0	20.2	11.4
Florida Atlantic	1,697	92.2	63.0	39.7	36.2	40.2	—	24.3	28.1	20.3
Florida International	3,327	81.7	62.5	29.6	26.1	29.9	—	17.4	21.7	12.4
North Florida	969	93.5	75.6	38.4	35.1	38.4	—	6.3	9.6	5.5
Florida A&M	722	84.5	81.9	34.3	31.3	34.8	—	22.7	26.0	22.4
Florida	3,387	70.2	46.7	34.6	31.9	34.8	—	12.9	15.9	7.9
FSU	2,557	90.1	63.0	39.6	34.8	39.6	—	18.4	23.0	9.7
Indiana										
IU Bloomington	3,530	79.2	73.1	35.6	31.7	35.6	—	5.9	8.2	3.9
IU Indianapolis	1,774	81.5	74.0	23.4	21.5	23.4	—	4.8	8.4	3.2
Purdue	1,977	88.0	55.2	57.2	52.7	57.2	—	22.2	26.9	16.2
Georgia										
Kennesaw State	3,231	38.8	36.5	15.2	13.6	15.3	—	7.0	7.6	5.2
Georgia	2,138	90.3	80.8	51.4	48.0	52.3	—	28.4	33.0	20.1
Georgia State	1,258	89.1	84.7	34.3	30.4	41.4	—	41.3	47.4	59.4
Ga. Southern	688	89.5	73.0	37.9	33.4	38.2	—	25.1	27.0	24.0
Georgia College	416	94.5	82.2	38.5	33.7	38.7	—	20.9	58.7	56.2
All 25 universities	**49,834**	**83.0**	**67.8**	**36.1**	**32.8**	**36.6**	**19.4**	**18.1**	**23.7**	**14.2**

Note: Table A3 reports the share of instructors (in percent) with non-missing values for each characteristic, restricted to instructors active from Fall 2024 through Spring 2026. Gender combines direct records from Texas salary data with the SSA baby-name proxy. Academic Rank reflects any source (salary database, faculty directory, ORCID employment, or syllabus text). The OpenAlex, *h*-Index, and Any Publication columns all derive from the OpenAlex bibliometric match and are thus highly correlated. Hire Year is available only for Texas institutions with salary databases reporting hire date. PhD Year draws on ORCID education records, faculty directory listings, and first-publication-year proxy. PhD Inst. and BA Inst. report shares with non-missing doctoral and undergraduate institution, respectively. "—" indicates the variable is not available for that institution. The sample includes 49,834 unique instructors across the 25 primary universities.

Table A4: University Characteristics

	Enrollment (1)	Admit Rate (2)	Mean SAT (3)	Grad. Rate (4)	Research/FTE (5)	Endow./FTE (6)	Avg. Salary (7)	Grad. Share (8)	Parent. Income (9)
Panel A: Selective (Barron's Highly Competitive)									
Texas A&M	78,501	0.63	1,260	0.61	13,322	230,934	129,056	0.23	119,400
Florida	62,811	0.24	1,390	0.75	16,135	37,160	121,404	0.39	106,100
Georgia	44,713	0.37	1,275	0.71	12,019	40,500	116,845	0.25	127,400
UT Austin	55,710	0.29	1,360	0.73	17,781	103,056	139,586	0.22	125,100
UT Dallas	34,317	0.65	1,290	0.56	4,141	23,068	133,815	0.33	100,800
Tier mean	*55,010*	*0.44*	*1,315*	*0.67*	*12,680*	*86,944*	*128,141*	*0.28*	*115,760*
Panel B: Moderately Selective (Barron's Very Competitive)									
Florida International	69,591	0.59	1,155	0.59	3,118	4,406	—	0.18	66,700
FSU	49,885	0.25	1,315	0.73	5,398	18,984	123,267	0.27	100,500
Georgia College	7,397	0.88	1,150	0.50	132	8,104	77,902	0.21	95,100
Georgia Southern	29,617	0.90	1,070	0.32	1,384	3,572	82,561	0.15	88,100
Georgia State	42,129	0.62	1,160	0.34	5,516	5,737	—	0.21	82,200
IU Bloomington	61,491	0.80	1,285	0.72	2,376	30,492	—	0.21	99,400†
IU Indianapolis	29,740	0.81	1,135	0.38	8,844	78,770	—	0.33	99,400†
Purdue	56,210	0.50	1,325	0.64	7,364	65,393	128,721	0.26	103,200
Central Florida	80,332	0.40	1,270	0.50	2,472	2,316	103,179	0.15	91,200
North Florida	19,673	0.62	1,100	0.49	928	6,822	85,218	0.14	88,900
Tier mean	*44,607*	*0.64*	*1,197*	*0.52*	*3,753*	*22,460*	*100,141*	*0.21*	*89,970*
Panel C: Broad Access (Barron's Competitive / Less Competitive)									
Florida A&M	10,428	0.21	1,090	0.27	3,138	9,351	86,655	0.17	57,000
Florida Atlantic	36,189	0.73	1,125	0.50	2,441	9,798	95,327	0.19	77,100
Kennesaw State	49,914	0.69	1,120	0.21	360	2,232	84,717	0.11	99,900
Sam Houston State	23,907	0.85	1,040	0.35	590	6,847	74,785	0.16	84,000
SFA State	12,740	0.88	1,070	0.38	316	9,839	74,125	0.14	89,300
Texas State	41,676	0.89	1,085	0.37	3,256	8,605	78,309	0.12	100,900
Texas Tech	43,357	0.71	1,180	0.44	5,565	22,554	103,453	0.21	110,300
Houston	52,007	0.70	1,240	0.45	3,925	23,387	123,437	0.19	65,700†
UT San Antonio	38,728	0.88	1,100	0.32	4,342	7,401	93,359	0.14	74,300
West Texas A&M	10,667	0.97	1,060	0.14	834	11,459	74,711	0.27	74,600
Tier mean	*31,961*	*0.75*	*1,111*	*0.34*	*2,477*	*11,147*	*88,888*	*0.17*	*83,310*
Overall mean	**40,416**	**0.64**	**1,186**	**0.47**	**5,240**	**30,440**	**102,163**	**0.21**	**93,016**

Note: Table A4 reports institutional characteristics for the 25 primary universities, grouped by Barron's selectivity tier. I obtain all variables except parental income from IPEDS (2023). Enrollment is twelve-month unduplicated headcount. Mean SAT is the average of the 25th and 75th percentiles on the verbal and math sections. Graduation rate is the six-year bachelor's completion rate (150% of normal time). Research per FTE, endowment per FTE, and average salary are in US dollars. Parental income is median household income of parents from Chetty et al. (2020). "—" denotes missing data (faculty salary unavailable in 2023 for Florida International, Georgia State, IU Bloomington, and IU Indianapolis). †System-level aggregate from Opportunity Insights (IU System and UH System), assigned where campus-level data are unavailable. Tier means computed over non-missing values.

Table A5: Major Characteristics

Major	Macro-Field	N (1)	AI Exposure (2)	Earnings ($) (3)	UN Rate (4)	Grad. Degree Share (5)	Mean SAT (6)	Female Share (7)
Agriculture	STEM	2,061	0.244	79,500	2.0	29.55	1,096	0.68
Architecture	STEM	2,048	0.707	105,500	4.5	26.53	1,109	0.48
Biology	STEM	6,620	1.164	87,667	3.8	65.14	1,160	0.85
Biomedical Engineering	STEM	773	–	–	–	–	–	–
Chemistry	STEM	2,736	1.121	86,000	4.3	67.48	1,160	0.86
Civil Engineering	STEM	1,063	–	–	–	–	–	–
Computer Science	STEM	3,967	0.969	120,000	7.0	32.69	1,141	0.41
Electrical Engineering	STEM	2,140	–	–	–	–	–	–
Engineering (Other)	STEM	5,876	–	–	–	–	–	–
Environmental Science	STEM	1,068	0.598	80,000	6.3	31.95	1,160	0.31
Geology	STEM	934	1.128	75,000	3.8	45.56	1,160	0.44
Health Sciences	STEM	6,060	1.179	75,500	4.6	49.60	1,026	0.77
Information Systems	STEM	2,709	0.969	104,500	5.5	27.29	1,082	0.33
Mathematics	STEM	7,030	1.253	100,000	5.8	51.34	1,246	0.50
Nursing	STEM	3,206	0.422	87,000	2.1	30.29	1,026	0.95
Physics	STEM	2,527	1.245	93,000	4.6	60.64	1,160	0.40
Public Health	STEM	3,329	1.165	70,000	4.5	50.56	1,026	0.83
Art	Humanities	6,830	1.020	81,000	6.7	27.75	1,071	0.61
Communication	Humanities	7,225	1.207	86,750	4.3	23.27	1,084	0.78
Dance	Humanities	571	0.719	–	–	–	1,071	0.95
English	Humanities	6,778	1.270	76,000	6.1	46.20	1,124	0.88
Film & Theater	Humanities	3,112	0.744	75,000	6.9	38.37	1,071	0.89
Foreign Languages	Humanities	4,886	1.264	77,000	1.6	53.14	1,124	0.68
History	Humanities	2,789	1.132	80,000	4.3	51.39	1,167	0.25
Interdisciplinary Studies	Humanities	6,508	–	81,667	4.2	40.82	1,026	0.61
Music	Humanities	7,233	1.197	–	–	–	1,071	0.32
Philosophy	Humanities	1,821	1.800	73,000	4.1	51.64	1,124	0.00
Anthropology	Social Sciences	1,345	0.913	65,000	7.9	48.20	1,156	0.67
Criminal Justice	Social Sciences	1,798	0.098	80,000	3.6	24.61	1,005	0.55
Economics	Social Sciences	2,195	1.203	115,000	3.5	41.60	1,156	0.43
Education	Social Sciences	9,502	1.065	56,833	2.4	53.30	1,045	0.81
Geography	Social Sciences	1,330	0.810	88,000	1.6	34.71	1,156	0.49
Political Science	Social Sciences	3,745	1.131	95,000	5.3	49.52	1,156	0.72
Psychology	Social Sciences	5,107	0.977	72,000	5.0	51.86	1,089	0.86
Public Administration	Social Sciences	1,659	0.775	80,000	2.2	46.96	1,022	0.58
Social Work	Social Sciences	2,634	1.103	60,000	1.9	53.78	1,022	0.85
Sociology	Social Sciences	1,517	0.940	77,000	3.4	39.24	1,156	0.72
Accounting	Business	1,911	1.054	97,000	2.6	33.67	1,082	0.57
Business Administration	Business	6,363	1.001	86,000	4.1	25.27	1,082	0.58
Finance	Business	1,951	0.955	112,000	2.8	31.48	1,082	0.33
Hospitality	Business	1,301	0.762	70,000	3.5	35.92	1,082	0.75
Marketing	Business	1,706	1.321	100,000	4.4	19.67	1,082	0.59
Other	Other	2,254	–	–	–	–	1,026	0.67

Note: Table A5 shows major-level characteristics for the harmonized major categories in the main sample. AI Exposure is the language modeling occupational exposure index from Felten et al. (2023). Earnings reports median mid-career earnings from the NY Fed. UN Rate is the major-linked unemployment rate. Grad. Degree Share is the share of students in the field who later earn a graduate degree. Mean SAT is the average SAT score from the NPSAS. Female Share is the share of women in the major-level IPEDS completion data.

Table A6: Descriptive Statistics: Full Sample

	Overall (1)	**AI Mentioned** (2)	**No AI Mentioned** (3)
Syllabi Characteristics			
AI Mentioned	0.55	1.00	0.00
Number of Words	2,776	3,208	2,251
Has TA	0.18	0.19	0.17
Instructor Characteristics			
Female	0.47	0.51	0.42
Tenure-Track	0.49	0.49	0.49
Any Publications	0.38	0.38	0.39
University Characteristics			
Selective	0.26	0.24	0.28
Moderately Selective	0.37	0.30	0.45
Broad Access	0.37	0.46	0.27
Major Characteristics			
Business	0.09	0.09	0.08
STEM	0.37	0.33	0.40
Social Sciences	0.21	0.22	0.20
Humanities	0.32	0.34	0.30
Course Characteristics			
Undergraduate	0.77	0.79	0.73
Graduate	0.23	0.20	0.26
Syllabi	148,218	81,251	66,967

Note: Table A6 shows descriptive statistics for the full 2024–2026 paper sample across the 25 universities in the main sample. The table follows the same structure as Table 1, using all four main terms rather than only Spring 2026.

Table A7: Qualitative Categories: AI Integration in Teaching

Category	N	Share	Example 1	Example 2
Scheduled AI Topic: N = 5,953				
Dedicated AI Modules	1,406	23.6%	Week on AI design methods	Multi-week AI sequence
Generative AI & LLMs	1,043	17.5%	Lecture on generative AI	LLM applications in domain
Machine Learning & Deep Learning	573	9.6%	Neural networks module	ML classification lecture
AI Activities in Schedule	550	9.2%	AI discussion board	Named AI assignment in schedule
AI Tools & Exploration	541	9.1%	AI tools overview session	AI for research demo
Domain-Specific AI Topics	459	7.7%	AI in healthcare analytics	AI in business intelligence
AI Ethics & Responsible Use	430	7.2%	AI bias & fairness module	Responsible AI use session
Professional Skills & Careers	411	6.9%	AI in communication & media	AI career readiness
AI Futures & Innovation	349	5.9%	Emerging AI technologies	AI & innovation trends
In-Class AI Exercise: N = 1,671				
Hands-On AI Tool Practice	410	24.5%	Generative AI hands-on session	Prompt engineering practice
Named In-Class AI Activities	244	14.6%	Numbered GenAI activities	AI simulation & role-play
Presentations & Collaborative Projects	196	11.7%	Student-led AI presentation	AI team exercise
Discussions & Debates	181	10.8%	Scheduled AI discussion	AI policy debate
Writing & Research Exercises	148	8.9%	AI writing exercise	AI research activity
AI Lab Sessions	124	7.4%	Supervised AI lab practice	AI software tutorial
AI Workshops	92	5.5%	Single-session AI workshop	Multi-session AI workshop
AI Demonstrations	73	4.4%	AI tool demo in class	Live AI output walkthrough
Professional Skills Exercises	55	3.3%	AI interview practice	AI resume review exercise
AI Readings and Materials: N = 3,161				
Weekly AI Readings	758	24.0%	Dedicated AI week with readings	Multiple numbered AI readings
Scholarly Articles	714	22.6%	Generative AI research paper	AI research paper with citation
AI/ML Textbooks	546	17.3%	Russell & Norvig (AI: Modern Approach)	Goodfellow et al. (Deep Learning)
Tool Tutorials & How-To Guides	301	9.5%	ChatGPT tutorial	LinkedIn Learning AI course
Videos & Documentaries	245	7.8%	YouTube AI video	AI documentary screening
AI Literacy Guides	182	5.8%	MLA-CCCC Student Guide to AI	AI pedagogy handbook
Popular Press & Journalism	177	5.6%	AI news article discussion	AI op-ed as assigned reading
Citation & Integrity Guides	115	3.6%	AI citation style guide	Academic integrity & AI resource
Case Studies	79	2.5%	Business AI case study	Technology AI case analysis

Note: Table A7 reports the supercategories identified from the qualitative evidence behind the three teaching integration measures. For each syllabus classified as integrating AI into teaching, the classification pipeline yields the supporting text, which is then grouped into broader categories through iterative coding. N reports the number of syllabi in each category and Share is the percentage of all positive instances for that measure. Examples describe common ways each category appears in syllabi. Unclassified cases are excluded from this table.

Table A8: Qualitative Categories: AI Integration in Evaluation

Category	N	Share	Example 1	Example 2
Graded AI Assignment: N = 1,826				
AI-Generated Writing Comparisons	285	15.6%	Statement of purpose vs. AI version	Human vs. AI output comparison
AI-Assisted Projects	278	15.2%	AI design project	AI business analytics project
Generic AI Assignments	270	14.8%	Named AI assignment	Compare own work with AI draft
ChatGPT / LLM Interaction	241	13.2%	Nine ChatGPT assignments	AI prompt engineering assignment
Discipline-Specific AI Tasks	164	9.0%	AI in animal behavior study	AI in nursing case study
Career Readiness (Resume & Interview)	144	7.9%	AI resume assignment	AI mock interview
Role-Play & Simulation	95	5.2%	AI sales role-play	AI simulation exercise
AI Discussion Forums	90	4.9%	AI-required discussion post	AI reflection forum
Adaptive Learning Platforms	76	4.2%	AI interactive case study	AI-adaptive homework system
Translation & Language Learning	21	1.2%	AI translation exercise	AI language production task
AI Output Deliverable: N = 1,140				
AI-Generated Written Text	293	25.7%	AI-drafted essay as deliverable	AI-generated summary submission
Interaction Transcripts & Logs	213	18.7%	ChatGPT conversation log	AI chat transcript submission
Screenshots of AI Output	112	9.8%	Screenshot of AI tool results	Annotated AI output screenshot
AI Content in Final Projects	112	9.8%	AI-generated visuals in portfolio	AI component in capstone project
Creative & Multimedia Artifacts	95	8.3%	AI-generated art submission	AI music/video creation
Lesson Plans & Professional Docs	63	5.5%	AI-assisted lesson plan	AI-drafted curriculum document
Input/Output Documentation	55	4.8%	AI prompt & response log	Transparency documentation
AI Tool Feedback Output	51	4.5%	AI writing feedback report	AI grading tool output
Speech & Language Production	10	0.9%	AI speech transcription	AI language production sample
Engage AI Content: N = 2,127				
AI-Specific Named Assignments	267	12.6%	Named AI engagement assignment	AI exploration project
Structured Engagement with Documentation	244	11.5%	AI use with reflection journal	Documented AI interaction
In-Class AI Exercises	221	10.4%	Active learning AI activity	Guided AI tool exercise
Human vs. AI Text Analysis	212	10.0%	Compare human & AI text	AI output evaluation essay
Resume & Interview Prep with AI	178	8.4%	AI resume feedback integration	AI interview preparation
AI in Writing Process	169	7.9%	AI-assisted drafting & revision	AI brainstorming for essays
Critical Evaluation of AI Output	152	7.1%	AI output critique assignment	Assess AI accuracy & bias
Design, Analysis & Problem-Solving	145	6.8%	AI-integrated design project	AI data analysis exercise
AI Interaction with Reflection	117	5.5%	AI chat with written reflection	Structured AI dialogue & analysis
Revision of AI-Generated Content	81	3.8%	Edit AI-generated draft	Revise AI output for accuracy

Note: Table A8 reports the supercategories identified from the qualitative evidence behind the three evaluation integration measures. The table follows the same structure as Table A7. N reports the number of syllabi in each category and Share is the percentage of all positive instances for that measure. Unclassified cases are excluded from this table.

Table A9: Prevalence of AI Integration by Semester

	Fall 2024	Spring 2025	Fall 2025	Spring 2026
Panel A: Full Sample				
Policy				
AI Policy	38.6	42.5	52.3	58.9
Explicitly Permits AI	12.2	15.1	22.0	25.7
Teaching				
Scheduled AI Topic	3.0	3.5	4.2	4.0
In-Class AI Exercise	0.7	1.0	1.1	1.0
AI Readings or Materials	1.7	1.9	2.4	2.4
Evaluation				
Named Graded AI Assignment	1.1	1.1	1.6	1.4
AI Output Deliverable	0.6	0.6	0.8	0.6
Engagement with AI Content	1.1	1.0	1.6	1.3
Unique Syllabi	22,078	21,054	45,799	59,231
Panel B: Balanced Sample				
Policy				
AI Policy	38.7	42.9	58.2	64.4
Explicitly Permits AI	11.9	15.1	25.1	29.8
Teaching				
Scheduled AI Topic	3.0	3.6	4.7	4.8
In-Class AI Exercise	0.7	1.0	1.4	1.2
AI Readings or Materials	1.7	2.0	2.6	3.1
Evaluation				
Named Graded AI Assignment	1.0	1.2	1.8	1.8
AI Output Deliverable	0.5	0.6	0.9	0.7
Engagement with AI Content	0.9	1.1	1.8	1.6
Unique Syllabi	18,372	17,738	18,936	18,634

Note: Table A9 reports the prevalence of the eight headline AI integration measures across the four semesters in the main sample. Panel A uses the full sample of 25 universities. Panel B restricts to courses observed in at least one term in both 2024–2025 and 2025–2026. All values are expressed as percentages.

Table A10: AI Prevalence by Broad Major Field

	Business	Humanities	STEM	Social Sci.	Other
Panel A. Full Sample					
AI Mentioned	57.9	57.7	50.2	57.3	51.7
Policy					
AI Policy	51.8	54.7	46.9	54.7	47.4
AI Permitted	29.9	19.9	19.7	21.4	20.7
Teaching					
Scheduled AI Topic	7.6	3.9	3.7	2.5	4.3
In-Class AI Exercise	2.0	1.1	0.8	0.7	1.4
AI Readings / Materials	4.6	2.3	1.8	1.9	3.1
Any Teaching	9.7	4.8	4.4	3.6	5.6
Evaluation					
Named AI Assignment	3.3	1.4	0.9	1.1	3.2
AI Output Deliverable	1.1	0.8	0.4	0.5	2.7
Engage AI Content	2.7	1.5	0.9	1.1	3.3
Any Evaluation	3.9	1.8	1.3	1.4	3.7
Any Classroom	11.3	5.5	5.0	4.4	7.9
Number of Syllabi	13,232	47,753	54,147	30,832	2,254
Panel B. Spring 2026					
AI Mentioned	64.0	63.9	59.3	65.2	59.0
Policy					
AI Policy	57.8	60.8	55.5	62.5	55.0
AI Permitted	35.2	23.2	25.0	26.9	24.6
Teaching					
Scheduled AI Topic	8.1	3.8	3.9	2.9	4.2
In-Class AI Exercise	2.1	1.0	0.8	0.7	1.8
AI Readings / Materials	4.9	2.6	1.9	2.2	2.6
Any Teaching	10.1	5.0	4.6	4.0	5.4
Evaluation					
Named AI Assignment	3.5	1.3	1.0	1.2	1.9
AI Output Deliverable	1.0	0.7	0.4	0.6	1.0
Engage AI Content	2.7	1.4	1.0	1.2	1.9
Any Evaluation	4.2	1.7	1.4	1.6	2.7
Any Classroom	12.0	5.9	5.3	4.9	6.5
Number of Syllabi	5,096	18,658	21,846	12,759	891

Note: Table A10 shows the prevalence of the main AI outcomes across the five broad major fields. Panel A reports averages pooled from Fall 2024 through Spring 2026; Panel B reports the Spring 2026 cross-section. Each entry reports the share of syllabi exhibiting the row outcome, in percent. Majors are grouped into macro-fields following a harmonized classification. The sample covers the 25 primary universities.

Table A11: AI Prevalence by University Selectivity

	Selective	Mod. Selective	Broad Access	All
Panel A. Full Sample				
AI Mentioned	51.3	44.7	67.3	54.8
Policy				
AI Policy	46.9	41.6	64.4	51.5
AI Permitted	19.4	17.8	25.4	21.0
Teaching				
Scheduled AI Topic	6.5	2.8	3.1	3.9
In-Class AI Exercise	1.6	0.6	0.9	1.0
AI Readings / Materials	3.7	1.8	1.7	2.3
Any Teaching	7.6	3.7	4.0	4.8
Evaluation				
Named AI Assignment	1.9	1.1	1.2	1.4
AI Output Deliverable	1.1	0.5	0.5	0.6
Engage AI Content	1.9	1.1	1.1	1.3
Any Evaluation	2.3	1.5	1.6	1.7
Any Classroom	8.6	4.5	4.7	5.6
Number of Syllabi	38,441	54,558	55,219	148,218
Panel B. Spring 2026				
AI Mentioned	61.7	53.7	73.0	62.4
Policy				
AI Policy	57.4	50.6	69.5	58.9
AI Permitted	24.8	22.2	30.3	25.7
Teaching				
Scheduled AI Topic	7.4	3.3	3.0	4.0
In-Class AI Exercise	1.7	0.7	0.9	1.0
AI Readings / Materials	4.0	2.2	1.8	2.4
Any Teaching	8.6	4.4	4.0	5.1
Evaluation				
Named AI Assignment	1.9	1.2	1.3	1.4
AI Output Deliverable	0.9	0.5	0.5	0.6
Engage AI Content	1.9	1.2	1.2	1.3
Any Evaluation	2.4	1.6	1.7	1.8
Any Classroom	9.7	5.3	4.8	6.0
Number of Syllabi	11,885	25,645	21,720	59,250

Note: Table A11 shows the prevalence of the main AI outcomes across the three university selectivity tiers. Panel A reports averages pooled from Fall 2024 through Spring 2026; Panel B reports the Spring 2026 cross-section. Each entry reports the share of syllabi exhibiting the row outcome, in percent. Selectivity tiers follow the Barron's classification: Selective (5 universities), Moderately Selective (10), and Broad Access (10). The sample covers the 25 primary universities.

Table A12: University-Level AI Prevalence, Spring 2026

University	**AI Policy**	**AI Permitted**	**Any Teaching**	**Any Eval.**	**Any Classroom**
Panel A: Selective (Barron's Highly Competitive)					
Texas A&M	62.3	19.1	8.7	1.6	9.3
Florida	40.5	19.4	10.8	3.4	11.9
Georgia	78.1	34.6	5.8	2.0	7.0
UT Austin	60.8	29.9	7.9	1.9	8.6
UT Dallas	36.3	15.8	11.1	3.6	13.0
Tier mean	*55.6*	*23.8*	*8.9*	*2.5*	*10.0*
Panel B: Moderately Selective (Barron's Very Competitive)					
Florida International	31.4	10.7	6.4	2.0	7.5
Florida State	60.0	30.5	4.7	0.8	5.0
Georgia College	100.0	2.3	1.0	1.3	2.1
Georgia Southern	92.0	15.4	1.4	1.2	2.2
Georgia State	36.6	15.8	2.8	1.8	3.9
IU Bloomington	34.9	14.4	3.5	1.5	4.4
IU Indianapolis	14.2	8.1	2.3	1.0	2.7
Purdue	57.6	35.8	5.8	2.7	7.3
Central Florida	62.8	34.3	4.9	1.8	5.8
North Florida	65.9	34.1	4.4	1.8	5.4
Tier mean	*55.6*	*20.1*	*3.7*	*1.6*	*4.6*
Panel C: Broad Access (Barron's Competitive / Less Competitive)					
Florida A&M	15.3	7.9	4.1	2.2	4.9
Florida Atlantic	100.0	28.8	4.4	1.5	5.0
Kennesaw State	87.2	34.4	0.8	0.5	1.2
Stephen F. Austin	38.9	16.8	3.2	1.7	4.0
Sam Houston State	98.2	41.7	3.0	1.8	3.5
Texas Tech	59.7	29.0	4.9	2.1	5.8
Texas State	74.9	35.9	6.5	2.6	7.7
Houston	39.5	18.7	4.9	1.6	5.7
UT San Antonio	96.8	53.7	5.7	2.8	7.3
West Texas A&M	95.9	46.1	4.9	2.0	5.7
Tier mean	*70.7*	*31.3*	*4.2*	*1.9*	*5.1*
Overall mean	**61.6**	**25.3**	**5.0**	**1.9**	**5.9**

Note: Table A12 shows Spring 2026 prevalence rates for the main AI outcomes at each of the 25 universities, grouped by Barron's selectivity tier. Each entry is in percent. Tier means are unweighted averages across universities within each tier.

Table A13: AI Prevalence by Course Level

	All	Basic UG	Advanced UG	Graduate
Panel A. Full Sample				
AI Mentioned	54.8	58.9	55.4	48.4
Policy				
AI Policy	51.5	56.3	52.1	44.2
AI Permitted	21.0	18.0	22.7	21.5
Teaching				
Scheduled AI Topic	3.9	3.0	3.8	5.0
In-Class AI Exercise	1.0	0.9	1.1	1.0
AI Readings / Materials	2.3	1.8	2.2	3.1
Any Teaching	4.8	3.9	4.8	6.1
Evaluation				
Named AI Assignment	1.4	1.4	1.4	1.2
AI Output Deliverable	0.6	0.8	0.6	0.5
Engage AI Content	1.3	1.5	1.3	1.1
Any Evaluation	1.7	1.8	1.8	1.5
Any Classroom	5.6	4.9	5.6	6.8
Number of Syllabi	148,120	43,887	69,856	34,377
Panel B. Spring 2026				
AI Mentioned	62.4	66.0	64.1	54.9
Policy				
AI Policy	58.9	63.1	60.5	50.6
AI Permitted	25.7	21.4	28.0	26.4
Teaching				
Scheduled AI Topic	4.0	3.0	4.0	5.2
In-Class AI Exercise	1.0	0.9	1.0	0.9
AI Readings / Materials	2.4	2.0	2.3	3.2
Any Teaching	5.1	4.1	5.0	6.4
Evaluation				
Named AI Assignment	1.4	1.4	1.4	1.2
AI Output Deliverable	0.6	0.7	0.6	0.4
Engage AI Content	1.3	1.5	1.4	1.1
Any Evaluation	1.8	1.9	1.9	1.5
Any Classroom	6.0	5.2	6.0	7.0
Number of Syllabi	59,216	17,323	27,421	14,472

Note: Table A13 shows the prevalence of the main AI outcomes across the three course-level groups. Panel A reports averages pooled from Fall 2024 through Spring 2026; Panel B reports the Spring 2026 cross-section. Each entry reports the share of syllabi exhibiting the row outcome, in percent. The sample covers the 25 primary universities.

Table A14: Shapley–Owen Decomposition for the Specific AI Outcomes

	Has AI Policy (1)	**AI Permitted** (2)	**Sched. AI Topic** (3)	**In-Class Exercise** (4)	**AI Readings** (5)	**Graded AI Assign.** (6)	**AI Output Deliv.** (7)	**Engage AI** (8)
Instructor	37.3	39.7	27.3	22.5	24.8	25.3	28.7	26.7
Course	23.5	20.9	32.6	20.0	26.9	25.6	26.8	24.1
University	6.6	1.6	0.3	0.1	0.2	0.1	0.1	0.1
Major	1.2	0.8	1.0	0.2	0.6	0.3	0.2	0.2
Term	2.9	2.0	0.1	0.0	0.1	0.1	0.0	0.1
Unexplained	28.5	35.0	38.6	57.1	47.4	48.7	44.2	48.8
Adjusted R^2	71.5	65.0	61.4	42.9	52.6	51.3	55.8	51.2
Observations	98,146	98,146	98,146	98,146	98,146	98,146	98,146	98,146

Note: Table A14 shows the five-way Shapley–Owen variance decomposition for the eight specific AI outcomes. The five components are instructor, local course, university, major, and term fixed effects. Each entry reports the share of total outcome variance, in percentage points.

Table A15: Univariate Course Characteristics and AI Integration

	Has AI Policy (1)	**AI Permitted** (2)	**Any Teaching** (3)	**Any Evaluation** (4)	**Any Classroom** (5)
Panel A. Course Level					
Freshman	6.77***	0.42	-0.08	0.75***	0.39
	(0.64)	(0.54)	(0.39)	(0.25)	(0.43)
Sophomore	4.36***	-0.32	-0.76**	0.42**	-0.51
	(0.50)	(0.45)	(0.30)	(0.18)	(0.33)
Junior	3.92***	0.87**	-0.89***	0.35**	-0.69**
	(0.43)	(0.39)	(0.28)	(0.15)	(0.29)
Senior	2.69***	0.98***	-0.67***	0.40***	-0.57**
	(0.38)	(0.35)	(0.25)	(0.14)	(0.26)
Panel B. Other Course Characteristics					
Log Enrollment	1.48***	0.45***	-0.19	0.16**	-0.08
	(0.17)	(0.16)	(0.12)	(0.07)	(0.13)
Has TA	5.41***	1.24***	0.41	0.25	0.59*
	(0.55)	(0.48)	(0.30)	(0.18)	(0.33)
Credit Hours	-0.57***	-0.32***	-0.10***	-0.05***	-0.12***
	(0.08)	(0.06)	(0.03)	(0.02)	(0.04)
Observations	131,396	131,396	131,396	131,396	131,396
Adjusted R^2	0.768	0.701	0.527	0.509	0.534

Note: Table A15 shows the univariate course-characteristic coefficients for the five main AI outcomes. Each coefficient is estimated in a separate linear probability model that absorbs instructor fixed effects and university-by-major-by-term fixed effects. Graduate courses are the omitted course-level category, and missing indicators are included for Log Enrollment, Has TA, and Credit Hours. Standard errors, two-way clustered by instructor and local course, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A16: Course Characteristics and the Specific AI Outcomes

	AI Permitted (1)	**Sched. AI Topic** (2)	**In-Class Exercise** (3)	**AI Readings** (4)	**Graded AI Assign.** (5)	**AI Output Deliv.** (6)	**Engage AI** (7)
Panel A. Course Level							
Freshman	-0.22	-0.13	-0.06	-0.43*	0.65***	0.29*	0.52**
	(0.55)	(0.38)	(0.17)	(0.26)	(0.22)	(0.17)	(0.22)
Sophomore	-0.94**	-0.66**	-0.14	-0.53**	0.27	0.08	0.25
	(0.46)	(0.29)	(0.15)	(0.22)	(0.17)	(0.13)	(0.18)
Junior	0.44	-0.86***	-0.22*	-0.50**	0.28**	0.12	0.17
	(0.39)	(0.25)	(0.12)	(0.20)	(0.14)	(0.10)	(0.14)
Senior	0.74**	-0.65***	0.03	-0.48**	0.31**	0.18**	0.28**
	(0.35)	(0.23)	(0.12)	(0.19)	(0.13)	(0.09)	(0.13)
Panel B. Other Course Characteristics							
Log Enrollment	0.46***	-0.10	-0.03	-0.04	0.05	0.07	0.09
	(0.16)	(0.11)	(0.05)	(0.08)	(0.06)	(0.05)	(0.07)
Has TA	1.13**	0.51*	0.05	0.45**	0.09	0.16	0.20
	(0.48)	(0.29)	(0.16)	(0.23)	(0.16)	(0.12)	(0.16)
Credit Hours	-0.31***	-0.08***	-0.02*	-0.07***	-0.04**	-0.01	-0.03**
	(0.05)	(0.03)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Observations	131,396	131,396	131,396	131,396	131,396	131,396	131,396
Adjusted R^2	0.701	0.522	0.451	0.497	0.505	0.536	0.507

Note: Table A16 shows the multivariate course-characteristic coefficients for AI Permitted and the six specific teaching and evaluation outcomes. Each column reports a linear probability model with the full course-characteristic bundle, instructor fixed effects, and university-by-major-by-term fixed effects. Graduate courses are the omitted category and missing indicators are included for Log Enrollment, Has TA, and Credit Hours. Standard errors, two-way clustered by instructor and local course, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: Course Characteristics by Broad Field

	STEM		Business		Social Sciences		Humanities		Other	
	Policy	**Classroom**	**Policy**	**Classroom**	**Policy**	**Classroom**	**Policy**	**Classroom**	**Policy**	**Classroom**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Course Level										
Freshman	5.70***	0.98	-4.60	-0.05	3.21*	-0.38	5.86***	0.10	1.87	3.75
	(1.17)	(0.88)	(3.01)	(3.81)	(1.86)	(1.19)	(0.96)	(0.58)	(5.20)	(5.33)
Sophomore	3.39***	-0.57	1.74	-0.25	2.10*	-0.80	3.70***	-0.21	-2.64	-4.52
	(0.93)	(0.62)	(2.40)	(1.81)	(1.25)	(0.70)	(0.83)	(0.54)	(4.07)	(3.04)
Junior	2.71***	-0.87*	2.36*	-0.14	3.59***	-0.73	3.75***	-0.23	-0.15	-7.00**
	(0.73)	(0.51)	(1.32)	(1.36)	(1.00)	(0.55)	(0.73)	(0.48)	(3.61)	(3.03)
Senior	2.15***	-0.65	0.23	-1.29	1.88**	-0.79	3.00***	-0.26	-0.71	-4.46
	(0.59)	(0.45)	(1.15)	(1.14)	(0.87)	(0.51)	(0.69)	(0.45)	(3.01)	(2.80)
Panel B. Other Course Characteristics										
Log Enrollment	0.11	-0.01	1.15**	-0.94*	0.72*	-0.82***	1.77***	0.47**	5.14	5.29*
	(0.26)	(0.22)	(0.49)	(0.51)	(0.39)	(0.25)	(0.37)	(0.22)	(4.00)	(3.08)
Has TA	5.20***	0.43	3.28	2.93*	6.74***	0.65	2.68***	0.24	10.23*	0.81
	(0.87)	(0.43)	(2.19)	(1.77)	(1.26)	(0.61)	(0.99)	(0.69)	(5.97)	(5.55)
Credit Hours	-0.51***	-0.10**	-0.52	-0.54***	-0.90***	-0.26***	-0.16	-0.07	-0.21	0.02
	(0.07)	(0.04)	(0.38)	(0.20)	(0.24)	(0.08)	(0.39)	(0.12)	(0.13)	(0.04)

Note: Table A17 shows the multivariate course-characteristic coefficients separately by broad field for Has AI Policy and Any Classroom. All columns use the same specification with instructor fixed effects and university-by-major-by-term fixed effects. Graduate courses are the omitted category and missing indicators are included for Log Enrollment, Has TA, and Credit Hours. Standard errors, two-way clustered by instructor and local course, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A18: Course Characteristics by Selectivity Tier

	Selective		Moderately Selective		Broad Access	
	Policy	**Classroom**	**Policy**	**Classroom**	**Policy**	**Classroom**
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Course Level						
Freshman	6.00***	2.01	7.11***	0.00	3.70***	0.11
	(1.73)	(1.49)	(1.21)	(0.66)	(0.85)	(0.65)
Sophomore	3.27***	-0.86	2.90***	0.04	2.56***	-0.89
	(1.12)	(0.80)	(0.90)	(0.51)	(0.74)	(0.55)
Junior	4.63***	-0.89	1.71**	-0.51	2.67***	-0.86*
	(1.03)	(0.86)	(0.69)	(0.42)	(0.62)	(0.47)
Senior	3.86***	-1.18*	1.31**	-0.43	1.29**	-0.51
	(0.80)	(0.61)	(0.62)	(0.38)	(0.55)	(0.46)
Panel B. Other Course Characteristics						
Log Enrollment	0.85***	-0.61***	1.89***	0.23	0.90***	0.51**
	(0.26)	(0.23)	(0.44)	(0.21)	(0.24)	(0.24)
Has TA	4.25***	0.79	7.57***	0.35	1.62*	0.81
	(0.83)	(0.57)	(1.09)	(0.57)	(0.91)	(0.55)
Credit Hours	0.15	-0.03	-0.58***	-0.09***	-0.12	-0.28**
	(0.47)	(0.26)	(0.09)	(0.04)	(0.14)	(0.12)

Note: Table A18 shows the multivariate course-characteristic coefficients separately by selectivity tier for Has AI Policy and Any Classroom. All columns use the same specification with instructor fixed effects and university-by-major-by-term fixed effects. Graduate courses are the omitted category and missing indicators are included for Log Enrollment, Has TA, and Credit Hours. Standard errors, two-way clustered by instructor and local course, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A19: Course Characteristics with Pre-2024 Course Information

	Has AI Policy (1)	**Any Classroom** (2)
Panel A. Course Level		
Freshman	9.02***	4.76*
	(2.68)	(2.54)
Sophomore	3.94**	-0.04
	(1.56)	(1.13)
Junior	3.57**	0.32
	(1.53)	(1.55)
Senior	4.62***	-0.65
	(1.16)	(0.87)
Panel B. Other Course Characteristics		
Log Enrollment	1.13***	-0.04
	(0.38)	(0.34)
Has TA	6.09***	0.95
	(1.04)	(0.73)
Credit Hours	0.90	0.06
	(0.61)	(0.35)
Average GPA	3.37***	1.85**
	(1.24)	(0.90)
AI Susceptible Share	0.02	0.06***
	(0.02)	(0.02)
Observations	20,336	20,336
Adjusted R^2	0.723	0.595

Note: Table A19 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 and AI Susceptible Share 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 A20: Florida Harmonized-Course Decomposition

	Has AI Policy (1)	**AI Permitted** (2)	**Any Teaching** (3)	**Any Evaluation** (4)	**Any Classroom** (5)
Instructor	32.6	35.2	22.9	19.6	22.3
FL Local Course	17.4	12.2	19.2	15.1	19.4
FL Harmonized Course	13.6	11.2	9.4	5.6	7.9
Context	15.5	6.8	5.6	1.1	4.9
Unexplained	20.9	34.6	42.9	58.5	45.5
Adjusted R^2	79.1	65.4	57.1	41.5	54.5
Observations	22,135	22,135	22,135	22,135	22,135

Note: Table A20 shows the Florida harmonized-course decomposition for the five main outcomes. The decomposition separates local course variation from Florida SCNS harmonized-course-by-term variation within the specification that also includes instructor fixed effects and the broader context component. The FL Harmonized Course component captures variation at the harmonized-course-by-term level. Each entry reports the share of total outcome variance, in percentage points, attributed to the corresponding component.

Table A21: Instructor Characteristics: Additional AI Outcomes

	AI Permitted (1)	**Any Teaching** (2)	**Any Evaluation** (3)
Gender			
Female	32.4	9.4	3.9
Male	28.1	9.8	3.5
Academic Rank			
Professor	28.5	9.5	3.4
Associate Professor	30.9	10.2	3.6
Assistant Professor	36.4	13.0	3.7
Non-Tenure-Track	30.4	9.8	4.1
Research Activity			
Any Publications	32.2	10.9	3.8
No Publications	28.6	8.9	3.5
Publication Count Tercile			
Low	32.3	10.1	4.0
Medium	35.3	11.5	4.3
High	29.5	11.3	3.2
H-Index Tercile			
Low	32.5	10.0	4.1
Medium	35.2	12.0	4.1
High	29.5	11.0	3.3
PhD at Highly Selective Institution			
Yes	30.4	13.2	3.4
No	32.4	10.4	3.6
Instructors		47,020	

Note: Table A21 shows the prevalence of additional AI outcomes by instructor characteristics. Each cell reports the share of instructors (in percent) for whom at least one syllabus exhibits the column outcome. Publication count and h-index terciles are computed among instructors with positive values. PhD institution selectivity follows the Barron's classification. The sample covers the 25 primary universities from Fall 2024 through Spring 2026.

Table A22: Instructor Characteristics: Heterogeneity by Field and Selectivity

	STEM		Non-STEM	
Panel A: STEM vs. Non-STEM				
	AI Policy (1)	Any Classroom (2)	AI Policy (3)	Any Classroom (4)
Gender				
Female	61.7	8.6	67.7	11.7
Male	52.1	9.6	59.4	11.9
Academic Rank				
Professor	53.2	8.8	64.8	12.0
Associate Professor	56.1	9.7	66.6	12.3
Assistant Professor	59.8	12.5	68.4	15.6
Non-Tenure-Track	55.9	9.7	60.7	12.0
Research Activity				
Any Publications	56.3	10.5	65.3	13.7
No Publications	55.5	8.0	62.3	11.1
Instructors	17,377		29,643	
Panel B: Selective vs. Other				
	Selective		**Other**	
	AI Policy (1)	Any Classroom (2)	AI Policy (3)	Any Classroom (4)
Gender				
Female	64.8	15.0	66.6	9.6
Male	53.9	15.0	58.0	9.8
Academic Rank				
Professor	53.6	14.1	63.5	9.6
Associate Professor	60.1	15.0	63.9	10.0
Assistant Professor	59.1	17.6	67.8	12.7
Non-Tenure-Track	63.1	17.0	58.3	10.1
Research Activity				
Any Publications	59.3	15.7	62.7	11.2
No Publications	54.6	14.8	62.2	8.8
Instructors	12,707		34,313	

Note: Table A22 shows the prevalence of AI outcomes by instructor characteristics, separately by field and selectivity. Panel A splits by STEM vs. non-STEM fields. Panel B splits by Barron's selectivity tier (Selective = tier 1; Other = tiers 2–3). Each cell reports the share of instructors (in percent) for whom at least one syllabus exhibits the column outcome. The sample covers the 25 primary universities from Fall 2024 through Spring 2026.

Table A23: Instructor Characteristics Across AI Integration Outcomes

	AI Permitted (1)	Scheduled AI Topic (2)	In-Class Exercise (3)	AI Readings (4)	Named Graded Assignment (5)	AI Output Deliverable (6)	Engagement w/ AI Content (7)
Female	0.013**	0.001	-0.001	0.001	0.000	-0.000	-0.000
	(0.005)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Associate Prof	-0.003	-0.001	0.000	-0.002	-0.004**	-0.003**	-0.005**
	(0.007)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
Assistant Prof	0.025***	0.007**	0.001	0.006*	-0.006***	-0.003***	-0.005**
	(0.009)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)	(0.002)
Non-TT Faculty	0.012	-0.002	-0.001	-0.002	-0.006***	-0.003**	-0.004*
	(0.009)	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Any Publications	0.007	0.002	0.001	0.005***	0.001	0.001	0.001
	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Years Since Last Pub	-0.002	-0.002***	-0.001	-0.001	-0.001	0.001	-0.001
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Course FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Univ. × Major × Level × Term FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	114,275	114,275	114,275	114,275	114,275	114,275	114,275
Adjusted R^2	0.405	0.396	0.214	0.336	0.315	0.338	0.295

Table A23 shows the estimated multivariate coefficients from the core instructor specification for each of the seven specific AI integration outcomes. Years Since Last Pub is standardized among publishers and set to zero for non-publishers. The specification absorbs local course and university×major×level×term fixed effects. Standard errors clustered at the university level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A24: Instructor Characteristics and AI Outcomes: Extended Covariates

	AI Policy		Any Classroom	
	(1) Univariate	(2) Multivariate	(3) Univariate	(4) Multivariate
Female	0.040***	0.039***	0.001	0.001
	(0.009)	(0.008)	(0.003)	(0.003)
Associate Prof	0.002	-0.005	-0.005	-0.005*
	(0.013)	(0.012)	(0.003)	(0.003)
Assistant Prof	0.029**	0.020	0.009*	0.007
	(0.014)	(0.013)	(0.005)	(0.005)
Non-TT Faculty	0.021**	0.016*	-0.007**	-0.006**
	(0.011)	(0.009)	(0.003)	(0.003)
Any Publications	0.003	0.003	0.006***	0.006***
	(0.005)	(0.005)	(0.002)	(0.002)
Years Since Last Pub	-0.007***	-0.008***	-0.004***	-0.003***
	(0.003)	(0.003)	(0.001)	(0.001)
H-Index	-0.013***	-0.006	-0.003*	-0.001
	(0.004)	(0.005)	(0.002)	(0.003)
Publication Count	-0.011***	-0.007**	-0.004***	-0.003
	(0.003)	(0.003)	(0.001)	(0.003)
PhD Highly Selective	0.008	0.013	-0.002	-0.003
	(0.014)	(0.014)	(0.005)	(0.005)
Course FE	Yes	Yes	Yes	Yes
Univ. × Major × Level × Term FE	Yes	Yes	Yes	Yes
Observations	115,050	114,275	115,050	114,275
Adjusted R^2	0.555	0.557	0.401	0.402

Table A24 shows the estimated coefficients from regressions of the main AI outcomes on an extended set of instructor covariates. Columns (1) and (3) enter each covariate separately; columns (2) and (4) enter all covariates jointly. Years Since Last Pub is standardized among publishers and set to zero for non-publishers. H-Index and Publication Count are standardized among instructors with positive values and set to zero otherwise. The specification absorbs local course and university×major×level×term fixed effects. Standard errors clustered at the university level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A25: Instructor Characteristics: UT Austin and UT Dallas

	AI Policy		Any Classroom	
	(1) Univariate	(2) Multivariate	(3) Univariate	(4) Multivariate
Female	0.075***	0.069***	0.006	0.007
	(0.009)	(0.010)	(0.006)	(0.006)
Associate Prof	0.064***	0.056***	-0.010	-0.008
	(0.018)	(0.019)	(0.011)	(0.011)
Assistant Prof	0.087***	0.081***	0.022**	0.025**
	(0.017)	(0.019)	(0.011)	(0.012)
Non-TT Faculty	0.078***	0.070***	-0.009	-0.002
	(0.015)	(0.016)	(0.009)	(0.010)
Any Publications	-0.010	-0.001	0.003	0.005
	(0.010)	(0.010)	(0.006)	(0.006)
Years Since Last Pub	-0.018***	-0.020***	-0.007*	-0.005
	(0.006)	(0.007)	(0.004)	(0.004)
H-Index	-0.017***	0.002	0.003	0.007
	(0.006)	(0.008)	(0.004)	(0.005)
Publication Count	-0.011***	-0.008*	-0.005**	-0.006**
	(0.004)	(0.004)	(0.002)	(0.003)
PhD Highly Selective	0.039	0.044	-0.003	-0.005
	(0.029)	(0.029)	(0.020)	(0.020)
Course FE	Yes	Yes	Yes	Yes
Univ. × Major × Level × Term FE	Yes	Yes	Yes	Yes
Observations	19,460	19,312	19,460	19,312
Adjusted R^2	0.459	0.465	0.400	0.402

Table A25 shows the core specification restricted to UT Austin and UT Dallas. Columns (1) and (3) enter each covariate separately; columns (2) and (4) enter all covariates jointly. H-Index and Publication Count are standardized among instructors with positive values and set to zero otherwise. Heteroskedasticity-robust standard errors in parentheses (only two institutions, so clustering is inappropriate). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A26: Instructor Characteristics and Any Classroom by Macro-Field

	Business (1)	**Humanities** (2)	**STEM** (3)	**Social Sci.** (4)	**Other** (5)
Female	0.016	0.003	-0.003	-0.002	0.041**
	(0.015)	(0.004)	(0.004)	(0.005)	(0.017)
Associate Prof	0.004	-0.008	-0.004	-0.005	0.058
	(0.018)	(0.006)	(0.003)	(0.007)	(0.055)
Assistant Prof	0.012	0.002	0.017***	0.003	-0.025
	(0.022)	(0.008)	(0.006)	(0.007)	(0.020)
Non-TT Faculty	-0.002	-0.013**	0.000	-0.002	0.024
	(0.022)	(0.005)	(0.004)	(0.007)	(0.047)
Any Publications	0.000	0.003	0.008*	0.007	0.101***
	(0.013)	(0.006)	(0.005)	(0.005)	(0.037)
Years Since Last Pub	-0.012*	-0.000	-0.003	-0.001	-0.030
	(0.006)	(0.002)	(0.003)	(0.004)	(0.023)
Course FE	Yes	Yes	Yes	Yes	Yes
Univ. × Major × Level × Term FE	Yes	Yes	Yes	Yes	Yes
Observations	11,178	37,602	41,294	23,021	1,180
Adjusted R^2	0.413	0.335	0.502	0.331	0.184

Table A26 shows the core multivariate specification for Any Classroom estimated separately by macro-field. Fixed effects: course and university×major×level×term. Standard errors clustered at the institution level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A27: Instructor Characteristics and AI Policy by Macro-Field

	Business (1)	**Humanities** (2)	**STEM** (3)	**Social Sci.** (4)	**Other** (5)
Female	0.018	0.050***	0.034**	0.036***	0.062*
	(0.016)	(0.010)	(0.017)	(0.010)	(0.036)
Associate Prof	-0.026	0.021	-0.003	-0.019	0.005
	(0.019)	(0.026)	(0.015)	(0.013)	(0.072)
Assistant Prof	0.020	0.025	0.026*	0.034	-0.001
	(0.029)	(0.028)	(0.016)	(0.023)	(0.044)
Non-TT Faculty	0.021	0.030*	0.008	0.014	0.122**
	(0.025)	(0.018)	(0.013)	(0.014)	(0.054)
Any Publications	0.007	0.020*	-0.007	0.014	0.128***
	(0.016)	(0.012)	(0.011)	(0.012)	(0.029)
Years Since Last Pub	0.003	-0.011***	-0.006	-0.009	-0.048
	(0.009)	(0.004)	(0.005)	(0.009)	(0.035)
Course FE	Yes	Yes	Yes	Yes	Yes
Univ. × Major × Level × Term FE	Yes	Yes	Yes	Yes	Yes
Observations	11,178	37,602	41,294	23,021	1,180
Adjusted R^2	0.506	0.555	0.586	0.525	0.493

Table A27 shows the core multivariate specification for AI Policy estimated separately by macro-field. Fixed effects: course and university×major×level×term. Standard errors clustered at the institution level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A28: Instructor Characteristics by Selectivity Tier

	Selective		Moderately Selective		Broad Access	
	(1) Policy	(2) Classroom	(3) Policy	(4) Classroom	(5) Policy	(6) Classroom
Female	0.064***	0.009	0.028***	-0.002	0.036***	0.001
	(0.021)	(0.008)	(0.006)	(0.006)	(0.011)	(0.004)
Associate Prof	0.040**	-0.002	-0.019*	-0.007	-0.010	-0.005
	(0.016)	(0.010)	(0.011)	(0.006)	(0.020)	(0.004)
Assistant Prof	0.068***	0.024***	0.008	-0.007*	0.016	0.009
	(0.012)	(0.006)	(0.006)	(0.003)	(0.021)	(0.007)
Non-TT Faculty	0.059***	-0.006	0.005	-0.009**	0.007	-0.002
	(0.011)	(0.007)	(0.014)	(0.004)	(0.011)	(0.006)
Any Publications	0.004	0.004	0.019***	0.005***	-0.001	0.009**
	(0.012)	(0.004)	(0.006)	(0.002)	(0.008)	(0.003)
Years Since Last Pub	-0.021***	-0.005***	-0.004	-0.003*	-0.002	-0.002
	(0.006)	(0.001)	(0.004)	(0.002)	(0.003)	(0.001)
Course FE	Yes	Yes	Yes	Yes	Yes	Yes
Univ. × Major × Level × Term FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28,812	28,812	40,362	40,362	45,101	45,101
Adjusted R^2	0.479	0.426	0.510	0.378	0.595	0.369

Table A28 shows the core multivariate specification for AI Policy and Any Classroom estimated separately by Barron's selectivity tier: Selective, Moderately Selective, and Broad Access. Fixed effects: course and university×major×level×term. Standard errors clustered at the institution level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B: Data Construction

B.1 Syllabi Data

B.1.1 Data Sources

I collect course syllabi from 25 public universities across Texas, Florida, Georgia, and Indiana, all of which have state-level policies requiring their institutions to make course syllabi publicly accessible. These policies generally require that syllabi be posted on a publicly accessible institutional website within a specified period of the start of classes. The scope of these requirements has expanded in recent years as additional states have adopted similar provisions.[1] Altogether, the sample spans ten universities in Texas, seven in Florida, five in Georgia, and three in Indiana. These range from selective research universities such as UT-Austin and the University of Georgia to broad-access institutions such as Florida A&M and West Texas A&M. For each university, I draw on the publicly accessible syllabi repository to collect documents covering the four academic terms between Fall 2024 and Spring 2026, and Table B1 summarizes the data source and collection approach for each institution.

How universities make syllabi publicly available varies considerably across the 25 institutions in the sample. Some universities maintain dedicated portals where syllabi are stored as standalone PDF documents, while others use course management platforms that record structured metadata alongside each syllabus, including the instructor name, course title, and department. Several universities instead provide syllabi through searchable online systems organized by department and academic term. This variation determines what information I can recover directly from the data source versus what I need to extract from the syllabus text itself. I describe the extraction approach I use to handle this variation in Section B.1.2.

For UT-Austin and UT-Dallas, I additionally recover syllabi dating back to Fall 2015. This provides a longitudinal window of more than ten years that I use to document the evolution of AI-related content before and after the rollout of ChatGPT.[2] Within the main estimation window, UT-Austin contributes approximately 11,700 syllabi and UT-Dallas 6,300, making them among the largest individual contributors to the sample.

Table B1 lists each of the 25 universities along with its state, selectivity tier, source type, the academic terms covered, and both the raw and final document counts after all processing steps. The initial corpus comprises more than 190,000 parsed syllabi across all 25 institutions. Section B.1.3 describes the quality filtering and deduplication steps that produce the final estimation sample.

[1] The disclosure requirements vary in scope and historical coverage across the four states. Texas and Florida have maintained these mandates for a longer period, while Indiana and Georgia implemented their requirements more recently, with full compliance across their public universities by 2024.

[2] Altogether, I recover approximately 55,800 syllabi for UT-Austin and 29,400 for UT-Dallas between Fall 2015 and Spring 2024, in addition to the syllabi from these two institutions that are included in the main estimation window.

Table B1: University Collection Summary

University	Selectivity	Source Type	Terms	Syllabi (Final)
Texas (10 universities)				
UT Austin	Selective	PDF portal	F24–S26	13,832
UT Dallas	Selective	PDF portal	F24–S26	7,950
UT San Antonio	Broad Access	SimpleSyllabus	F24–S26	9,623
Houston	Broad Access	PDF portal	F24–S26	7,806
Texas State	Broad Access	SimpleSyllabus	F24–S26	5,427
Texas A&M	Selective	SimpleSyllabus	F24–S26	3,532
West Texas A&M	Broad Access	PDF portal	F24–S26	2,471
Stephen F. Austin	Broad Access	PDF portal	F24–S26	4,231
Sam Houston State	Broad Access	SimpleSyllabus	F24–S26	1,502
Texas Tech	Broad Access	SimpleSyllabus	F24–S26	5,536
Florida (7 universities)				
Central Florida	Moderate	PDF portal	F24–S26	17,305
Florida State	Moderate	SimpleSyllabus	F24–S26	4,520
Florida Atlantic	Broad Access	SimpleSyllabus	F24–S26	8,621
Florida International	Moderate	SimpleSyllabus	F24–S26	12,233
North Florida	Moderate	SimpleSyllabus	F24–S26	1,854
Florida A&M	Broad Access	SimpleSyllabus	F24–S26	3,043
Florida	Selective	SimpleSyllabus	S25–S26	6,890
Indiana (3 universities)				
IU Bloomington	Moderate	CourseLeaf API	F24–S26	6,262
IU Indianapolis	Moderate	CourseLeaf API	F24–S26	3,234
Purdue	Moderate	PDF portal	F24–S26	4,295
Georgia (5 universities)				
Kennesaw State	Broad Access	SimpleSyllabus	F24–S26	6,959
Georgia	Selective	PDF portal	F24–S26	6,237
Georgia State	Moderate	SimpleSyllabus	F24–S26	2,103
Georgia Southern	Moderate	SimpleSyllabus	F24–S26	1,348
Georgia College	Moderate	SimpleSyllabus	F24–S26	1,404
Total (25)				**148,218**

Note: Table B1 reports the data-collection summary for the 25 primary universities. "Selectivity" follows the three-tier Barron's classification (Selective, Moderate, Broad Access). "Source Type" indicates the platform from which syllabi were collected (PDF portal, SimpleSyllabus, or CourseLeaf API). "Terms" reports the academic terms included in the analysis window. "Syllabi (Final)" is the section-deduplicated count entering the estimation sample. The sample covers 148,218 section-deduplicated syllabi across the 25 primary universities from Fall 2024 through Spring 2026.

B.1.2 Metadata 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, and they form the basis of the fixed-effect structures used throughout the paper. Table 1 reports the coverage rates for each of these fields across the 25 universities in the sample.

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 a 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.[3]

Instructor names and course titles require a more involved extraction approach, as their format and placement vary widely across institutions and even across individual syllabi within the same university. For instructor names, I search the opening pages of each syllabus using a series of text-matching rules ordered by precision, beginning with explicitly labeled fields and progressing to less structured patterns such as academic honorifics and university email addresses. I then standardize each extracted name to ensure that the same instructor is consistently identified across documents. For course titles, I apply an analogous approach that targets common formatting conventions, supplemented by institution-specific rules where the 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. Table B2 presents the extraction prompt verbatim; I apply it to the first 5,000 characters of the syllabus text.

I assign each syllabus a unique identifier, which I use to track individual documents through every stage of the data construction. This identifier ensures that the classification results described in later sections can be linked back to the original syllabus text at any point.

B.1.3 Sample Construction

I focus the sample on the four academic terms between Fall 2024 and Spring 2026, which yields 191,757 parsed records across the 25 universities.[4] I then apply a set of quality filters designed to remove documents that do not constitute genuine course syllabi. These filters exclude records with fewer than 500 characters of extractable text, course listing pages identified by structural markers, preliminary course descriptions that lack substantive content, and empty course management system shells. These filters remove 10,731 records, or about six percent of the corpus, yielding 181,026 documents that I classify as genuine syllabi. The filtering is concentrated at a small number of institutions, most notably Indiana University, where empty CourseLeaf shells account for 17 percent of the records removed at that university.

Since many instructors teach multiple sections of the same course using an identical syllabus, I deduplicate the corpus 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. This step reduces the corpus from 181,026 post-filter records to 148,218 unique course-section observations, a reduction

[3] At Purdue, the syllabus portal does not record the academic term in its metadata, affecting roughly 2,000 records. In these cases, I recover the term from date references and season-year patterns in the document text.

[4] I exclude summer terms because summer course offerings are substantially smaller and differ systematically from regular-semester courses in both instructor composition and course selection.

Table B2: Metadata Extraction Prompt

SYSTEM PROMPT

You extract course metadata from American university course syllabi. You identify the course title, primary instructor, department code, course number, section number, semester, and year. Be precise and return values exactly as they appear in the syllabus.

USER MESSAGE TEMPLATE

Extract ALL course metadata from this syllabus excerpt.

RULES:

1. `title`: The full course title (e.g., *"Principles of Marketing"*). Do NOT include the dept/number prefix.
2. `instructor`: The PRIMARY instructor's full name. Strip titles (Dr., Prof.) and credentials (Ph.D., MBA). Do NOT return TAs or graders.
3. `dept`: The department/subject code (e.g., *"CS"*, *"MKT"*). Use the standard 2–4 letter abbreviation.
4. `course_number`: The numeric course number (e.g., *"3300"*, *"6V99"*). Include any letter suffixes.
5. `section`: The section number (e.g., *"001"*). If not found, return `''''`.
6. `semester`: One of *"Fall"*, *"Spring"*, or *"Summer"*.
7. `year`: The 4-digit year (e.g., 2023).

CONFIDENCE GUIDE:

- `high`: All fields clearly labeled and unambiguous
- `medium`: Most fields found but some inferred from context
- `low`: Key fields missing or ambiguous
- `none`: Cannot extract meaningful metadata

`{JSON output format shown}`

SYLLABUS EXCERPT:

—

`{text}`

—

Note: This table presents the prompt used for LLM-based metadata extraction, applied to records where regex-based extraction failed to recover the course title, instructor name, or department code. The prompt is applied to the first 5,000 characters of the syllabus text.

of 18 percent. The deduplication rate varies considerably across institutions, from under one percent at universities where few courses run multiple sections to nearly 30 percent at universities with large multi-section introductory sequences.[5]

The final estimation sample contains 148,218 unique course-section observations covering approximately 42,000 instructors across the 25 universities and 42 harmonized college majors. I recover 59,250 syllabi in Spring 2026 and 45,812 in Fall 2025, while the two earlier terms account for about 21,000 observations each. The higher counts in the more recent terms reflect expanding compliance with public disclosure mandates, particularly in Indiana and Georgia, where the disclosure requirements took effect more recently than in Texas and Florida. Table B3 reports the per-university sample funnel, documenting the number

[5] I keep a crosswalk that maps each non-representative syllabus to its group representative, so that classification results obtained for one section propagate to all sections sharing the same course and instructor in a given term.

of records at each stage of the filtering and deduplication process for each of the 25 universities.

Table B3: Per-University Sample Funnel

		Available		*After Filtering*		*Unique Syllabi*	
University	**Source**	**All**	**4 Terms**	**Cluster**	**Filter %**	**Final**	**Dedup. %**
Texas (10 universities)							
UT Austin	PDF portal	18,113	18,113	14,914	17.7	13,832	7.3
UT Dallas	PDF portal	15,227	15,227	11,206	26.4	7,950	29.1
UT San Antonio	SimpleSyllabus	12,843	12,843	12,838	0.0	9,623	25.0
Houston	PDF portal	9,872	9,872	9,872	0.0	7,806	20.9
Texas State	SimpleSyllabus	6,733	6,733	6,661	1.1	5,427	18.5
Texas A&M	SimpleSyllabus	4,956	4,956	4,956	0.0	3,532	28.7
West Texas A&M	PDF portal	3,376	3,376	2,754	18.4	2,471	10.3
Stephen F. Austin	PDF portal	6,341	6,341	6,318	0.4	4,231	33.0
Sam Houston State	SimpleSyllabus	1,925	1,925	1,925	0.0	1,502	22.0
Texas Tech	SimpleSyllabus	6,339	6,339	6,314	0.4	5,536	12.3
TX subtotal		85,725	85,725	77,758	9.3	61,910	20.4
Florida (7 universities)							
Central Florida	PDF portal	22,445	22,445	22,445	0.0	17,305	22.9
Florida State	SimpleSyllabus	5,742	5,742	5,742	0.0	4,520	21.3
Florida Atlantic	SimpleSyllabus	11,389	11,389	11,389	0.0	8,621	24.3
Florida International	SimpleSyllabus	16,130	16,130	15,968	1.0	12,233	23.4
North Florida	SimpleSyllabus	2,548	2,548	2,548	0.0	1,854	27.2
Florida A&M	SimpleSyllabus	3,431	3,431	3,431	0.0	3,043	11.3
Florida	SimpleSyllabus	9,262	7,021	6,905	1.7	6,890	0.2
FL subtotal		70,947	68,706	68,428	0.4	54,466	20.4
Indiana (3 universities)							
IU Bloomington	CourseLeaf API	8,131	8,131	6,753	17.0	6,262	7.3
IU Indianapolis	CourseLeaf API	4,303	4,303	3,525	18.1	3,234	8.3
Purdue	PDF portal	4,394	4,394	4,295	2.3	4,295	0.0
IN subtotal		16,828	16,828	14,573	13.4	13,791	5.4
Georgia (5 universities)							
Kennesaw State	SimpleSyllabus	8,215	8,215	8,215	0.0	6,959	15.3
Georgia	PDF portal	6,591	6,286	6,237	0.8	6,237	0.0
Georgia State	SimpleSyllabus	2,782	2,781	2,599	6.5	2,103	19.1
Georgia Southern	SimpleSyllabus	1,614	1,614	1,614	0.0	1,348	16.5
Georgia College	SimpleSyllabus	1,602	1,602	1,602	0.0	1,404	12.4
GA subtotal		20,804	20,498	20,267	1.1	18,051	10.9
Total (25)		**194,304**	**191,757**	**181,026**	**5.6**	**148,218**	**18.1**

Note: Table B3 presents the per-university sample funnel from raw records to the final estimation sample for the 25 primary universities. "Source" indicates the data collection platform. "All" reports total records in the enriched corpus across all terms. "4 Terms" restricts to the four analysis terms (Fall 2024, Spring 2025, Fall 2025, Spring 2026). "Cluster" reports records remaining after quality filtering (removal of non-syllabus documents). "Unique Syllabi" is the final count after section deduplication, which retains one representative per (institution, department, course number, term, instructor) group. "Dedup. %" is the share of post-filter records removed by deduplication. For UT Austin, UT Dallas, and West Texas A&M, the cluster draws from the longitudinal fulltext pipeline, which accounts for the larger filtering loss.

B.2 Extracting Information from Syllabi

B.2.1 Syllabus Text Characteristics

Beyond the course metadata described in the previous subsection, I extract a set of textual characteristics from each syllabus that serve as control variables in the regression specifications. Syllabi vary considerably in their length, writing complexity, and the degree to which instructors use prescriptive or encouraging language. These dimensions of course design and instructor style may correlate with AI policy adoption independently of the observable course and instructor characteristics. I construct nine text-based controls organized around three broad dimensions of syllabus heterogeneity. These measures capture the complexity of the document's prose, the prescriptiveness of the instructor's tone, and the presence of explicit course policies. I compute all nine measures directly from the syllabus text using pattern-matching rules and standard readability formulas, without using language models.

I begin with three measures of text complexity that capture how accessible and varied the syllabus prose is. The Flesch-Kincaid Grade Level is a widely used readability index that combines average sentence length and syllable complexity to approximate the U.S. grade level required to comprehend the document (Flesch, 1948; Kincaid et al., 1975). I complement this with the Moving-Average Type-Token Ratio (MATTR), a length-independent measure of lexical diversity that estimates vocabulary richness through a 100-word sliding window over the text (Covington and McFall, 2010), and average sentence length in words. 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, including AI-related mentions.[6]

I also capture several dimensions of instructor tone that reflect baseline differences in how prescriptive or encouraging an instructor's written language is. I compute the prescriptive modal share as the fraction of strong directive modals such as "must," "shall," and "required" among all modal verbs in the document. I also measure penalty language density and supportive language density as the frequency of sanction-related and encouraging terms per 1,000 words, respectively. I further recover binary indicators for whether the syllabus contains an explicit late-work policy and an explicit attendance policy, which together capture the degree of policy prescriptiveness in the course design. All continuous text controls are standardized to mean zero and unit standard deviation in the estimation sample. I recover all nine measures for 99.7 percent of syllabi in the main sample.

In addition to the nine core text controls, I identify whether a teaching assistant is mentioned in each syllabus, which I use in extended specifications as a proxy for course size and instructional support. I combine a two-step text search with platform metadata where available, searching first for explicit role labels such as "Teaching Assistant" and "Gradu-

[6] I winsorize Flesch-Kincaid grade level and average sentence length at the 99th percentile before standardization, as documents with minimal sentence-ending punctuation can produce artificially high values for both measures.

ate Assistant," and then for contextual references such as "your TA" or "TA office hours." For the subset of universities whose syllabi are sourced from the SimpleSyllabus platform, I supplement the text-based detection with structured metadata fields that record TA assignments. I combine both sources so that a course is coded as having a TA if either the text or the metadata indicates one.[7] This approach identifies a teaching assistant in approximately 19 percent of syllabi in the estimation sample, with rates ranging from under 3 percent at smaller teaching-focused institutions to nearly 50 percent at large research universities.

Table B4 reports definitions, summary statistics, and coverage rates for each of the nine core text control variables. The table shows that the text controls exhibit meaningful variation across the sample, fitting in with the substantial heterogeneity in course design and instructor style across universities and disciplines documented in Table A2.

Table B4: Text Control Variables: Summary Statistics

Variable	Description	Mean	SD
Flesch–Kincaid Grade	Readability index (higher = more complex)	16.99	79.90
Lexical Diversity (MATTR)	Moving-average type–token ratio (50-word window)	0.72	0.03
Avg. Sentence Length	Mean words per sentence	25.49	201.05
Modal Prescriptive Share	Share of modal verbs that are prescriptive	0.26	0.14
Penalty Language Density	Penalty-related terms per 1,000 words	2.61	2.47
Supportive Language Density	Supportive/encouraging terms per 1,000 words	2.60	2.09
Has Late-Work Policy	Binary: syllabus mentions a late-work policy	0.57	0.49
Has Attendance Policy	Binary: syllabus mentions an attendance policy	0.62	0.49
Log Word Count	Natural log of full-text word count	7.74	0.68

Note: Table B4 reports 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 148,218 section-deduplicated syllabi across the 25 primary universities from Fall 2024 through Spring 2026.

B.2.2 AI-Related Content

To identify syllabi that contain AI-related content, I use a curated set of ten keyword patterns organized into two tiers that capture distinct dimensions of AI engagement in the classroom. The first tier covers five policy-oriented terms that are directly tied to the post-ChatGPT period, including ChatGPT, GPT-4, GPT-3, GPT, and OpenAI, all of which are essentially absent from syllabi before the 2022–2023 academic year. The second tier captures five broader patterns that reflect general engagement with artificial intelligence as a topic or tool, including "artificial intelligence," "AI," "generative AI," "GenAI," and "large language model." Together, these two tiers allow me to distinguish between syllabi that reference specific generative AI products and those that engage with AI as a broader concept in their course content or policies. I deliberately exclude terms with high false-positive rates, such as "Claude" and "Bard," which frequently appear as personal names or liter-

[7] Four universities in the sample (Texas A&M, Houston, Florida Atlantic, and Florida International) populate the SimpleSyllabus TA metadata fields. The text-based and metadata-based sources are largely complementary, and the combined approach identifies substantially more TAs than either source alone.

ary references in course materials. I also exclude terms like "Copilot" and "DALL-E" that provide limited incremental signal relative to the ten core patterns.

The standalone acronym "AI" requires special treatment because it frequently refers to unrelated concepts in university syllabi, most notably "Associate Instructor," a widespread title at Indiana University and UT Austin, as well as building codes and room numbers. I address this through a two-step procedure that I develop using an analysis of over 88,000 "AI" matches across five universities in the sample. In the first step, I check whether any known false-positive pattern appears in the immediate vicinity of the match, including common title formats such as "Instructor, AI" and room codes, and I block the match if one is found. In the second step, I require at least one AI-related term to appear within five content words of the "AI" match. The reference list contains 130 terms covering AI tools and concepts (e.g., "generative," "algorithm"), academic integrity (e.g., "plagiarism," "integrity"), and course design (e.g., "assignment," "submission").[8] This two-step approach allows me to keep the large majority of genuine AI references while removing the false positives that would otherwise inflate the measured prevalence of AI-related content in the sample.

For each syllabus with at least one keyword match, I extract the surrounding text using a window of 500 characters on each side of every mention, which typically captures one to two paragraphs of context around each reference. When multiple keyword matches occur within the same syllabus, I merge overlapping windows into contiguous passages, so that a syllabus with several nearby AI references produces a single coherent excerpt rather than redundant fragments. This extraction produces a focused text field that captures the AI-relevant portions of each syllabus while discarding unrelated course content such as grading scales and office hours. This extraction substantially reduces the volume of text that enters the classification stage, allowing me to classify the full corpus of syllabi with AI mentions efficiently.

An important concern in this setting is that some of the observed AI-related text may reflect institution-level or department-level mandates rather than individual instructor choices. Many universities now require instructors to include standardized language about academic integrity and AI use in their syllabi. To distinguish between these two sources, I develop a decomposition procedure that segments each syllabus's AI-related text into sentences and groups similar sentences within institution-department-term cells. I define cells at this level to allow for the fact that boilerplate language often varies across departments within the same university and across academic terms. Sentence groups that appear in 70 percent or more of the syllabi in a given cell are classified as institutional boilerplate and subtracted from the text.[9] The resulting residual text captures instructor-specific content and is available for supplemental analyses that aim to isolate individual pedagogical choices from institutional mandates.

[8] I construct this reference list from the patterns observed in the calibration sample of five universities. The full list is documented in the replication package.

[9] The 70 percent threshold balances the need to capture genuinely repeated institutional language against the risk of misclassifying common but independently authored phrases.

B.3 Instructor Characteristics

I complement the syllabus-level analysis with a rich set of instructor characteristics that I recover from multiple administrative and bibliometric data sources. These characteristics allow me to examine how faculty attributes, covering academic rank, gender, research output, and career experience, relate to the adoption of AI-related policies in the classroom. This is particularly important given that the instructor workforce at these institutions ranges from research-active faculty who appear in bibliometric databases to lecturers, adjuncts, and instructors of practice who are largely absent from these sources and appear primarily in administrative records.

I construct the instructor dataset in several steps. I first build a roster of all unique instructors from the syllabus metadata, identifying 49,834 individuals active between Fall 2024 and Spring 2026 across the 25 primary universities. I then link each instructor to six families of external data sources, described in turn below, covering OpenAlex bibliometric records (Section B.3.1.1), ORCID education records (Section B.3.1.2), publicly available CVs and faculty profiles (Section B.3.1.3), department websites (Section B.3.1.4), state salary databases (Section B.3.1.5), and additional text-based and name-based sources (Section B.3.1.6). Because several data sources may provide the same information for a given instructor, I consolidate overlapping information using a priority ordering that favors official administrative records over self-reported data, as described in Section B.3.2.

Altogether, I recover academic rank for 68 percent of the instructors in the sample, gender for 83 percent, research output indicators for 36 percent, and years since PhD for 18 percent. I find that these rates vary substantially across universities, reflecting differences in the availability of state salary databases and in the representation of non-research-intensive faculty in bibliometric databases such as OpenAlex. Table A3 presents the full set of instructor characteristics and their availability across the 25 universities, showing how recovery rates for each characteristic vary by state and institution.

B.3.1 Main Data Sources

B.3.1.1 OpenAlex

OpenAlex is an open-access bibliometric database that indexes over 250 million scholarly works and their associated author profiles (Priem et al., 2022). I draw on OpenAlex to recover information on instructors' research productivity and career trajectories, which allow me to distinguish instructors who are engaged in frontier research from those whose primary role is teaching. For each university in the sample, I download the complete set of authors affiliated with the institution using its Research Organization Registry (ROR) identifier, yielding roughly 100,000 author records across all 25 universities. I run all matching on this pre-downloaded data.

I match the downloaded author pool to the instructor roster using a name-matching approach that proceeds from exact matches to progressively more flexible comparisons, allowing for alternative spellings, common nickname variants such as "Bob" for "Robert,"

and approximate string similarity. When multiple candidates match a single instructor, I accept the match automatically if one candidate's publication count exceeds twice that of the next-best candidate, and I flag the remaining cases for manual review. Fewer than 5 percent of matches require this manual review. I obtain an OpenAlex match for 35 percent of instructors across the 25 universities, with match rates ranging from 12 percent at Stephen F. Austin to 57 percent at Purdue. Table B5 presents the match rates for each university, showing substantial variation that reflects the composition of the instructor workforce at these institutions, which includes large numbers of lecturers, adjuncts, and instructors of practice who do not maintain indexed research profiles.

From the matched profiles, I recover the h-index, publication count, total citations, two-year mean citedness, and research topics for each instructor, which I use to construct measures of research productivity and impact in the analysis. I also extract the years since first and last publication as career-stage proxies, where the first-publication year serves as a lower bound on career length when direct PhD information is unavailable. Additionally, 66 percent of matched instructors have ORCID identifiers linked through their OpenAlex profiles, which I use to query ORCID education records as described below.

B.3.1.2 ORCID

ORCID is a nonprofit organization that maintains persistent digital identifiers for researchers, who self-report their education and employment histories on the platform. Approximately one in five instructors in the roster have ORCID identifiers linked through their OpenAlex profiles. For these instructors, I query the ORCID database to extract education records, including degree type, granting institution, field, and graduation year, along with employment records that provide supplementary information on institutional affiliations and academic titles. For instructors without a direct ORCID link, I additionally conduct name-based searches with institutional affiliation filters as a supplementary approach, though this yields substantially fewer matches.

ORCID education records are the highest-priority source for PhD information in the project because the data are researcher-verified, which makes them substantially more reliable than text-extracted records from CVs or faculty profiles. Of the instructors with ORCID identifiers, 61 percent have education records on file, from which I recover 5,235 PhD completion years, representing 58 percent of all PhD year information recovered across all sources. I also recover the PhD institution for these instructors, along with the PhD field for 93 percent of those with education records and the degree-granting country for all of them. Table B6 presents the ORCID match rates and the availability of education records by university, showing how the recovery of PhD information varies with institutional research intensity.

Table B5: OpenAlex Match Rates and Bibliometric Characteristics by University

				Conditional on Match			
University	**Instructors**	**Matched**	**OA (%)**	**ORCID (%)**	**Mean h**	**Mean Works**	**Mean Cites**
Florida A&M	722	248	34.3	48.0	7.9	34	888
Florida Atlantic	1,697	673	39.7	65.8	9.7	42	1,003
Florida International	3,327	984	29.6	68.6	10.7	55	1,351
FSU	2,557	1,012	39.6	74.6	11.2	51	1,450
Georgia College	416	160	38.5	56.9	5.2	24	462
Georgia Southern	688	261	37.9	61.7	7.0	52	1,205
Georgia State	1,258	431	34.3	51.3	8.6	39	1,137
IU Bloomington	3,530	1,257	35.6	53.7	11.7	60	1,627
IU Indianapolis	1,774	416	23.4	43.8	10.1	50	1,419
Kennesaw State	3,231	492	15.2	58.3	7.0	33	663
Purdue	1,977	1,130	57.2	81.2	15.1	76	1,980
Stephen F. Austin	777	92	11.8	53.3	5.4	27	349
Sam Houston State	600	219	36.5	59.4	5.9	27	493
Texas A&M	1,888	796	42.2	79.0	13.9	69	1,834
Texas Tech	1,657	680	41.0	68.5	10.1	49	1,388
Texas State	1,933	614	31.8	59.6	8.4	36	1,028
Central Florida	3,956	1,507	38.1	70.1	11.4	55	1,250
Florida	4,284	1,354	31.6	71.0	14.0	73	2,158
Georgia	2,181	1,122	51.4	78.1	14.6	62	1,876
UH	2,531	972	38.4	79.1	13.4	66	1,736
North Florida	969	372	38.4	53.2	8.3	37	645
UT Austin	7,960	2,912	36.6	58.7	14.0	69	2,376
UT Dallas	4,638	1,254	27.0	62.7	13.9	70	2,141
UT San Antonio	2,511	790	31.5	76.2	11.9	56	1,332
West Texas A&M	795	267	33.6	48.3	6.3	28	842
All	**57,857**	**20,015**	**34.6**	**66.2**	**11.9**	**58**	**1,632**

Note: Table B5 reports OpenAlex match rates and mean bibliometric characteristics for each of the 25 primary universities. The sample includes all unique instructors appearing in the syllabus data from Fall 2024 through Spring 2026 ($N = 57{,}857$). "OA" is the share of the full roster matched to an OpenAlex author profile. "ORCID" is the share of matched instructors for whom OpenAlex reports an ORCID identifier. Mean h-index, works, and citations are computed conditional on having an accepted match.

B.3.1.3 Instructors' CVs

To complement the information from OpenAlex and ORCID, I additionally recover instructors' educational backgrounds from publicly available CVs and faculty profile pages. This is particularly valuable for PhD institution and PhD year, which are available from ORCID for only a subset of instructors, and allows me to extend the recovery of educational backgrounds to a broader set of the instructor workforce. I extract PhD information by searching biography text for degree-related phrases and use a language model as a fallback for cases where the text does not follow standard formatting, yielding 3,462 CV-derived PhD year estimates.

In this context, I assign quality tiers to PhD information based on the underlying source, which informs the priority ordering I apply when consolidating information from multiple sources in Section B.3.2. ORCID education records represent the highest quality because they are researcher-verified. Faculty directory entries represent high quality because they are typically maintained by institutional staff. CV-derived records are of moderate quality because formatting and completeness vary across profiles. The first-publication-year

Table B6: ORCID Coverage and Covariate Recovery by University

University	Instructors (1)	ORCID Linked (2)	PhD Year (3)	PhD Field (4)	BA Year (5)
Texas					
UT Austin	3,812	27.8	14.1	13.5	8.4
UT Dallas	2,122	21.6	12.1	11.4	7.1
UH	2,531	24.1	14.9	13.9	8.5
Texas State	1,933	16.3	9.2	8.6	5.3
Texas A&M	1,888	28.2	17.3	16.3	12.6
West Texas A&M	376	21.3	9.0	8.2	5.1
Stephen F. Austin	777	4.8	6.2	5.5	4.2
Sam Houston State	600	16.7	7.7	7.0	5.7
Texas Tech	1,657	22.7	11.5	10.7	7.7
UT San Antonio	2,511	19.9	13.2	11.8	6.7
Florida					
Central Florida	3,956	22.7	10.1	9.3	5.4
Florida Atlantic	1,697	22.1	10.1	9.2	7.4
Florida International	3,327	16.3	8.8	7.9	5.7
North Florida	969	20.4	4.5	4.4	2.6
Florida A&M	722	14.8	6.5	6.4	3.7
Florida	3,387	23.4	11.1	10.6	5.7
Indiana					
IU Bloomington	3,530	18.2	5.8	5.1	3.1
IU Indianapolis	1,774	9.6	2.8	2.6	2.1
Purdue	1,977	38.4	20.4	19.5	11.9
Georgia					
Kennesaw State	3,231	7.2	4.1	3.7	2.8
Georgia	2,138	34.9	15.4	14.4	10.2
Georgia State	1,258	17.6	4.6	4.1	3.9
Georgia Southern	688	18.8	9.6	8.6	7.3
Georgia College	416	17.5	6.2	5.8	4.6
FSU	2,557	23.6	12.2	11.4	7.2
Total	**49,834**	**21.2**	**10.5**	**9.8**	**6.4**

Note: Table B6 reports the share of instructors (in percent) for whom I recover each covariate from ORCID education records, for the 25 primary universities in the estimation sample, covering Fall 2024 through Spring 2026. Column (1) reports the number of unique instructors active in the four main terms. "ORCID Linked" is the share with a linked ORCID identifier via their OpenAlex author profile. "PhD Year" is the doctoral completion year. I recover the PhD-granting institution for all instructors with a PhD Year. "PhD Field" is the department or discipline listed on the ORCID profile. "BA Year" is the bachelor's completion year. The roster includes 49,834 unique instructors.

proxy from OpenAlex provides a lower bound on career length when no direct PhD record is available. Table B7 presents the CV-derived PhD recovery by university, showing the number of matched profiles and the share with PhD year and PhD institution information.

B.3.1.4 Department Website

I collect faculty directory information from 14 of the 25 primary universities by systematically downloading profile pages from departmental and institutional websites.[10] This yields approximately 8,160 scraped profiles, of which approximately 3,780 match to the in-

[10]The specific collection method varies by university depending on how each institution organizes its faculty directory. Some universities maintain centralized searchable databases, while others provide faculty listings at the departmental level.

Table B7: Instructor CV Recovery by University

			Share of instructors with (%)				
University	**Syllabi**	**Instructors**	**Gender**	**Rank**	**Pubs**	***h*-idx**	**PhD yr**
Texas							
UT Austin	13,832	3,771	93.8	72.4	46.1	45.8	15.8
UT Dallas	7,950	2,107	79.7	67.7	34.6	34.2	28.5
UT San Antonio	9,623	2,510	91.5	84.4	32.0	31.5	15.1
Houston	7,806	2,483	92.9	76.0	39.5	39.1	16.5
Texas State	5,427	1,928	92.8	74.1	32.2	31.7	12.7
Texas A&M	3,532	1,782	82.6	68.4	45.2	44.5	18.3
Texas Tech	5,536	1,651	90.2	71.8	41.2	40.9	15.9
Stephen F. Austin	4,231	777	95.2	77.5	11.8	11.8	6.2
Sam Houston State	1,502	583	89.7	76.5	37.6	37.6	7.9
West Texas A&M	2,471	376	94.4	79.5	38.6	38.6	9.0
Florida							
Central Florida	17,305	3,858	85.7	76.2	39.3	38.8	10.5
Florida International	12,233	3,233	84.0	64.1	30.7	30.4	9.2
Florida	6,890	2,943	68.7	47.6	35.1	34.9	12.3
Florida State	4,520	2,558	90.0	63.0	39.6	39.6	16.2
Florida Atlantic	8,621	1,697	92.2	63.0	40.2	39.7	10.1
North Florida	1,854	969	93.5	75.6	38.4	38.4	4.5
Florida A&M	3,043	721	84.5	81.8	34.7	34.3	6.4
Georgia							
Georgia	6,237	2,126	90.3	80.8	52.4	51.5	27.0
Kennesaw State	6,959	2,321	54.1	50.8	21.2	21.2	5.7
Georgia State	2,103	1,205	89.1	85.5	42.1	34.7	36.9
Georgia Southern	1,348	668	91.9	74.9	39.2	38.9	9.9
Georgia College	1,404	415	94.5	82.2	38.8	38.6	6.3
Indiana							
IU Bloomington	6,262	2,956	79.1	73.0	35.6	35.6	5.9
Purdue	4,295	1,951	88.1	55.3	56.9	56.9	20.3
IU Indianapolis	3,234	1,457	81.7	74.3	23.5	23.5	3.0
Total	**148,218**	**47,046**	**85.3**	**69.8**	**37.7**	**37.2**	**13.9**

Note: Table B7 presents instructor CV covariate coverage for the 25 primary universities from Fall 2024 through Spring 2026. "Instructors" counts unique instructor names per university. "Gender" combines direct observation from salary records with SSA baby-name proxy (95% threshold). "Rank" aggregates salary databases, faculty directories, ORCID employment, and syllabus text. "Pubs" indicates any indexed scholarly work. "h-idx" indicates an OpenAlex match yielding the Hirsch index. "PhD yr" reflects the strict measure (directly observed PhD completion year only). Kennesaw State's low coverage reflects its use of login usernames rather than real names. The sample covers 148,218 section-deduplicated syllabi taught by 47,046 unique instructors across the 25 primary universities from Fall 2024 through Spring 2026.

structor roster, for an overall match rate of 47 percent.

From matched faculty profiles, I recover academic rank, PhD year, PhD institution, and research area. Table B8 presents the results for the 14 scraped universities, showing the number of profiles collected, the match rates, and the marginal contribution of each variable beyond what I recover from the other sources. In aggregate, faculty directories provide 643 additional rank classifications, 792 PhD years, and 808 PhD institutions that I do not recover from salary records, OpenAlex, or ORCID. Match rates vary across the 14 universities, with Houston, Central Florida, and UT Dallas yielding the highest rates among the

scraped institutions. Although the aggregate contribution is modest, faculty directories fill important gaps for institutions where salary data provide limited information on academic rank. This is particularly relevant for the three Indiana universities, where salary records include generic titles with no tenure-track distinction.

Table B8: Instructor Enrichment Coverage: 14 Universities with Faculty Directory Scraping

		Faculty Dir.	PhD (Any Source)	
University	**Instructors**	**Matched (%)**	**Year (%)**	**Inst. (%)**
Central Florida	3,956	26.1	16.0	20.2
UT Dallas	2,122	38.9	31.0	36.8
Florida State	2,557	23.0	18.4	23.0
Florida International	3,327	9.3	17.4	21.7
Sam Houston State	600	50.0	15.0	35.5
Houston	2,531	9.5	25.5	32.0
Georgia	2,138	7.8	28.4	33.0
Florida	3,387	3.6	12.9	15.9
McMaster	537	16.9	7.3	9.1
UT San Antonio	2,511	3.5	15.7	30.1
UT Austin[a]	3,812	0.0	21.2	35.2
Texas Tech	1,657	0.8	16.7	23.1
Texas A&M	1,888	0.1	32.9	36.0
South Florida	344	0.3	8.1	10.5
Subtotal (14)	**31,367**	**12.1**	**20.0**	**26.8**
Other 11[b]	**15,736**	—	**11.8**	**14.5**
All 25	**47,103**	**8.0**	**17.3**	**22.7**

Note: Table B8 reports enrichment coverage for the 14 universities with active faculty directory scrapers, sorted by number of roster matches. "Instructors" is the count of unique instructors in the Fall 2024 through Spring 2026 analysis roster. "Matched" is the share linked to a scraped faculty directory record. "PhD Year" and "PhD Inst." report coverage from *any* enrichment source. [a]UT Austin uses a separate departmental page scraper (rank only); the 0.0% reflects that results are stored in a different table. [b]West Texas A&M, Florida Atlantic, North Florida, Florida A&M, Georgia Tech, Georgia Southern, Kennesaw State, IU Bloomington, IU Indianapolis, Purdue, and North Dakota.

B.3.1.5 State Salary Databases

All four states in the sample mandate public disclosure of employee compensation at state universities. I draw on these administrative records to recover instructor characteristics that are difficult to obtain from other sources, particularly academic rank and institutional tenure status. I match instructors to public salary records from four sources, covering the Texas College Salaries database for the 10 Texas universities, OpenTheBooks for the 7 Florida universities, Open Georgia for the 5 Georgia universities, and the Indiana Gateway portal for the 3 Indiana universities. Table B9 summarizes the fields available from each state, including which states provide rank, gender, hire date, and salary information.

I match instructors to salary records by name, using exact matching as the primary strategy and approximate string matching as a fallback for minor spelling variations. Altogether, I obtain a salary match for 57 percent of instructors in the roster. Match rates range from 81 percent at Florida A&M, where nearly all instructors are full-time state employees, to

17 percent at Purdue, where Indiana salary data are less comprehensive for graduate instructors and adjuncts. These differences in match rates reflect variation in the composition of the instructor workforce across institutions rather than limitations of the matching approach itself.

I recover academic rank from all four states by normalizing raw job titles to a canonical four-level hierarchy that distinguishes among Professor, Associate Professor, Assistant Professor, and non-tenure-track faculty, following the same classification I describe in Section 2.2 of the main text. The Texas College Salaries database additionally records gender as a demographic field, making it the primary source for gender information at the 10 Texas institutions. It also provides hire date, which I use to construct a measure of years at the current institution.[11] For the other three states, gender information is not available from salary records, and I instead rely on the name-based approach described in Section B.3.1.6.

B.3.1.6 Other Data Sources

For instructors whose academic rank I do not recover from salary databases or faculty directories, I extract rank information directly from the syllabus text by searching for references to academic titles in instructor information blocks and signature lines. I take a conservative approach to this text-based extraction, accepting only short text snippets that clearly indicate rank. I reject common false matches such as references to departmental staff, office administrators, or course policies that mention academic titles without referring to the instructor of record. I treat this as the lowest-priority source for rank in the consolidation, given that self-reported titles in syllabi are occasionally informal or incomplete.

I recover direct gender information only for instructors matched to Texas salary records, which record gender as a demographic field, and for a small number of records from faculty directories. For the remaining instructors, I impute gender using U.S. Social Security Administration baby-name frequency data, which report the number of male and female births for each first name across all historical years. I apply a two-step procedure to this imputation. I first check whether other instructors at the same institution who share the same first name have a known gender from salary records, and if so, I assign the majority gender from this peer-based inference. For the remaining instructors, I assign the majority gender from the SSA data when the name prevalence exceeds 75 percent. I record the imputation source and confidence level for each assignment, distinguishing among tiers above 99 percent, 95 to 99 percent, 90 to 94 percent, and 75 to 89 percent. Altogether, this approach yields gender information for 83 percent of instructors in the sample.

For UT Austin and UT Dallas, where I additionally recover syllabi dating back to 2015, I construct experience measures from each instructor's pre-2024 course histories. These include years at the university, the number of distinct courses the instructor has taught, and the number of times the instructor has taught the current course, which provide additional

[11] Indiana salary records provide salary amounts but essentially no usable rank information, as they record generic titles with no tenure-track distinction. I address this gap using faculty directory information for the three Indiana universities (Section B.3.1.4).

Table B9: Salary Database Recovery by University

University	Instructors	Salary (%)	Rank (%)	Hire Year (%)
Texas				
Stephen F. Austin	777	76.4	36.4	—
Sam Houston State	582	69.4	30.6	68.0
Texas A&M	1,780	59.3	34.4	56.5
Texas Tech	1,651	66.4	36.1	64.0
Texas State	1,928	26.3	2.6	21.3
Houston	2,476	59.2	36.7	59.2
UT Austin	3,771	62.7	42.0	60.4
UT Dallas	2,098	54.8	26.2	36.4
UT San Antonio	2,500	75.9	31.9	58.4
West Texas A&M	376	74.2	35.4	—
Subtotal	*17,939*	*60.3*	*31.7*	*49.3*
Florida				
Florida A&M	721	80.7	54.8	—
Florida Atlantic	1,697	55.6	33.2	—
Florida International	3,225	57.7	25.0	—
Florida State	2,558	34.7	34.5	—
Central Florida	3,837	58.7	29.8	—
Florida	2,937	38.7	30.6	—
North Florida	969	74.2	30.9	—
Subtotal	*15,944*	*52.6*	*31.3*	
Georgia				
Georgia College	415	67.7	44.1	—
Georgia Southern	667	64.0	42.9	—
Georgia State	1,205	71.8	38.1	—
Kennesaw State	2,321	48.6	36.0	—
Georgia	2,126	74.2	54.3	—
Subtotal	*6,734*	*63.6*	*43.3*	
Indiana				
IU Bloomington	2,956	66.9	—	—
IU Indianapolis	1,457	65.9	—	—
Purdue	1,951	16.5	—	—
Subtotal	*6,364*	*51.2*		
Total	**46,981**	**56.9**	**29.1**	**18.8**

Note: Table B9 reports what the salary databases specifically contribute for each university. Instructors is the number of unique instructors appearing in syllabi from the four target terms (Fall 2024, Spring 2025, Fall 2025, Spring 2026). Salary is the share of roster instructors successfully matched to a public salary record. Rank is the share for whom the salary database yields an informative ladder-rank classification (Professor, Associate Professor, or Assistant Professor). The remaining matched instructors receive a non-tenure-track classification from their job title. Hire Year is the share with a recovered hire date, available only from the Texas College Salaries database. "—" indicates the field is not available or not informative from the relevant state source. The Indiana Gateway reports generic job titles that do not distinguish among academic ranks, so rank recovery for the three Indiana universities relies on other enrichment sources.

controls for instructor experience in supplementary specifications.

B.3.2 Instructor Enrichment

Because several data sources may provide the same characteristic for a given instructor, I prioritize sources according to their reliability when they disagree. Under this ordering, official administrative records take precedence over self-reported data, which in turn take precedence over information extracted from syllabi and from web-based sources. For academic rank, the ordering draws primarily on salary database titles, which record each instructor's official institutional position, and fills remaining gaps with faculty directory listings, ORCID employment records, and syllabus text.[12] For PhD year and PhD institution, I draw first on ORCID education records, which are researcher-verified, then on faculty directory entries, and then on information extracted from publicly available CVs and faculty profile pages. For gender, direct records from the Texas salary database take precedence, and for the remaining states I rely on the peer-based inference and SSA name proxy described in Section B.3.1.6.

Applying these priority rules, I recover gender for 83 percent of instructors, drawing primarily on direct records from Texas salary databases and supplementing with name-based probabilistic assignment for the remaining states. I classify instructors into four rank categories, distinguishing among full professors, associate professors, assistant professors, and non-tenure-track faculty, which includes lecturers, adjuncts, and instructors of practice. I classify 68 percent of instructors across the 25 universities, with salary database titles serving as the primary source for most institutions. For research activity, I construct two measures from OpenAlex profiles: a binary indicator for whether the instructor has any indexed publications, available for 37 percent of the roster, and the log h-index as a continuous measure of research impact available for all matched instructors.[13]

Beyond these cross-sectional characteristics, I measure career experience as years since doctoral degree completion, drawing on the PhD year recovered through the priority ordering described above. I recover this measure for 18 percent of instructors from high-quality sources, predominantly ORCID education records and faculty directories. I additionally store the year of each instructor's earliest indexed publication on OpenAlex as a separate career-stage proxy, which provides a lower bound on career length for instructors without a direct PhD record. For Texas instructors, I also recover hire year from salary records, which allows me to construct a measure of institutional tenure. For each covariate with incomplete coverage, I include a missing-data indicator alongside the imputed value in the regression specifications, following Jones (1996), so that observations with incomplete instructor information remain in the estimation sample.

Table A3 presents the coverage rates for each instructor characteristic across the 25 universities in the sample, broken down by gender, rank, research indicators, and educational background. I find that these rates vary substantially across institutions, reflecting differ-

[12] For Indiana institutions, salary records provide generic titles with no tenure-track distinction, so the effective primary source for rank at these universities is faculty directory profiles. This is the main contribution of the directory scraping described in Section B.3.1.4.

[13] I set both research measures to zero for unmatched instructors, reflecting the large share of lecturers and adjuncts who do not maintain indexed research profiles.

ences in the composition of the instructor workforce and in the availability of state salary databases rather than limitations of the extraction approach.[14] The distribution of PhD year sources across these methods shows how ORCID education records, faculty directories, and CV-derived extraction each contribute to the overall recovery of doctoral information.

B.4 University Characteristics

To understand how AI adoption varies across institutions, I complement the syllabus-level data with university-level covariates that capture institutional size, selectivity, resources, and the socioeconomic composition of the student body. I draw on the Integrated Postsecondary Education Data System (IPEDS), the Opportunity Insights database of Chetty et al. (2020), and the Barron's Admissions Competitiveness Index. I recover these characteristics for all 25 primary universities in the estimation sample.

IPEDS Data. I obtain institutional characteristics from the Integrated Postsecondary Education Data System, administered by the National Center for Education Statistics.[15] From the Enrollment survey I compute log enrollment as a measure of institutional size and the share of graduate enrollment as a proxy for research orientation. The Admissions survey provides SAT and ACT score distributions for admitted students, from which I construct the mean SAT composite score as a selectivity measure.[16] I also extract average faculty salary from the Salaries survey, and research expenditures and endowment value from the Finance surveys, which I normalize by full-time equivalent enrollment to obtain per-student measures of research intensity and accumulated wealth. IPEDS data are available for all 25 universities, though faculty salary figures are missing for four institutions in the most recent year.

Chetty et al. (2020). I supplement the IPEDS data with measures from the Opportunity Insights project that capture dimensions of the student body's socioeconomic composition that IPEDS variables alone do not measure. I merge median parental household income and the share of students from the bottom income quintile from the Mobility Report Cards of Chetty et al. (2020), matched at the institution level.[17] Median parental income serves as a summary measure of the economic backgrounds of students at each university, and I use it as the primary socioeconomic covariate throughout the analysis.

[14] For instance, Kennesaw State University has low coverage across most characteristics because its syllabus platform stores login identifiers rather than instructor names, preventing both salary matching and name-based gender imputation.

[15] IPEDS data are publicly available at `https://nces.ed.gov/ipeds`. I download complete data files for six survey components covering the 2016–2023 academic years.

[16] For institutions that report ACT scores but not SAT, I convert ACT composites to SAT equivalents using the College Board concordance tables. I compute the mean SAT as the average of the 25th and 75th percentile scores summed across the verbal and math sections.

[17] For four universities in multi-campus systems (IU Bloomington, IU Indianapolis, Purdue, and Houston), the Opportunity Insights data report at the system level, and I assign the system-level aggregate where campus-specific values are unavailable.

Barron's Selectivity. I classify each university by admissions selectivity using the Barron's Admissions Competitiveness Index (2009 vintage), as reported in the Opportunity Insights college-level data files (Chetty et al., 2020). I collapse the original five tiers into a three-tier grouping (Selective, Moderately Selective, and Broad Access) that is widely used in the economics of education literature (Biasi and Ma, 2025). This grouping aligns with the substantial variation in resources and student composition across the 25 universities.[18] I use this classification throughout the paper to examine whether AI adoption patterns differ systematically by institutional selectivity.

Altogether, I carry forward seven standardized university-level covariates for the main specifications: log enrollment, research expenditures per student, endowment per student, average faculty salary, mean SAT, graduate enrollment share, and median parental income. Table A4 reports these characteristics for each of the 25 primary universities, grouped by selectivity tier. Selective universities have mean SAT scores roughly 200 points higher than Broad Access institutions, spend more than five times as much per student on research, and draw students from households with approximately $33,000 higher median income.

B.5 Major Characteristics

B.5.1 Main Variables

I complement the syllabus-level data with a set of 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, conditional on university and term fixed effects, and they serve as controls in specifications that isolate the instructor-level component of AI adoption. I construct these measures for the 42 harmonized majors described in Section B.5.2, drawing on five external data sources.

I draw on the Federal Reserve Bank of New York's College Labor Market data, which uses the American Community Survey to report labor market outcomes for 74 major categories (Abel et al., 2014). From this source, I extract unemployment rates, underemployment rates, early- and mid-career median wages, and the share of graduates who go on to earn a graduate degree. I map these to my 42 harmonized fields through a hand-coded crosswalk. Among these variables, the graduate-degree share turns out to be the strongest negative predictor of AI classroom integration across the outcomes I examine. This fits in with the view that fields oriented toward professional credentialing tend to have more rigid curricular structures that are slower to accommodate new pedagogical tools.[19]

I measure each major's exposure to language modeling advances using the LMOE index from Felten et al. (2023), which maps the capabilities of language models to the tasks performed in specific occupations. I assign scores from occupations to majors through a CIP-

[18] The Selective tier comprises 5 universities (Texas A&M, Florida, Georgia, UT Austin, and UT Dallas), the Moderately Selective tier 10 universities, and the Broad Access tier 10 universities.

[19] For example, Biology, Chemistry, and Philosophy have among the highest graduate-degree shares in the sample, with more than half of graduates in each field going on to earn an advanced degree.

to-SOC crosswalk, weighting each occupation by BLS national employment counts so that each major's score reflects the occupational distribution of its graduates. The index captures intuitive differences across fields, with Philosophy, Marketing, and English ranking among the most exposed and Criminal Justice, Agriculture, and Nursing among the least. This ordering fits with the expectation that fields whose graduates enter writing-intensive and analytical occupations face greater exposure to language modeling advances.

From the 2019–20 National Postsecondary Student Aid Study, a nationally representative survey of approximately 95,000 postsecondary students, I recover mean SAT composite scores, mean ACT scores, and mean college GPA for each of 23 broad major-field categories. I map these to my 42 harmonized majors, with 13 fields matching directly and 27 assigned through approximate parent-field mappings.[20] Unlike the preceding national-level sources, the IPEDS Completions Survey and the College Scorecard provide data at the institution-by-major level. From these sources, I extract degree counts at the bachelor's, master's, and doctoral levels, the share of female graduates in each institution-by-major cell, and post-graduation median earnings. I am able to match IPEDS and Scorecard records for more than 90 percent of the institution-by-major combinations across the 25 universities.

In the main heterogeneity specifications, I focus on six covariates that span four conceptual channels through which major characteristics could predict AI classroom integration. The LMOE score captures the occupational relevance of AI advances for each field's graduates and is the key variable of interest in the heterogeneity analysis. I complement this with two labor market indicators from the NY Fed data. The mid-career median wage proxies for the long-run economic returns to the major, and the unemployment rate captures labor market tightness that could independently shape how urgently instructors prepare students for technologically evolving workplaces. I also include the graduate-degree share, which I interpret as a measure of the academic orientation of the field, since majors with high graduate-degree shares tend to have more structured curricula that may be slower to accommodate new tools. The mean SAT composite score from the NPSAS captures the baseline academic preparation of students selecting into each field, and the share of female graduates from IPEDS Completions is the only covariate that varies at the institution-by-major level. Table B10 reports these six characteristics for each of the 42 harmonized majors.[21]

B.5.2 Harmonization of College Majors

Across the 25 universities in the estimation sample, I observe more than 1,500 distinct department codes, reflecting heterogeneous administrative nomenclature across states and institutions. A code such as "BIO" may appear at one university while another uses

[20] For approximate mappings, closely related majors share the same NPSAS values. For instance, Economics and Political Science both receive the values reported for the broader Social Sciences category.

[21] The NY Fed data cover 39 of the 42 majors, as Dance, Music, and Biomedical Engineering lack separate categories. The LMOE index covers 41 majors, with Interdisciplinary Studies excluded due to the absence of a reliable CIP-to-SOC mapping.

Table B10: Major Characteristics: 42 Harmonized Majors

Major	Mid-Career Earnings ($)	Unemp. Rate (%)	Grad. Degree (%)	Mean SAT	Share Female (%)
STEM					
Agriculture	80,000	2.0	30	1,100	67
Architecture	106,000	4.5	27	1,110	54
Biology	88,000	3.8	65	1,160	67
Biomedical Eng.	–	–	–	1,170	53
Chemistry	86,000	4.3	68	1,160	53
Civil Eng.	115,000	2.3	37	1,170	28
Computer Science	120,000	7.0	33	1,140	25
Electrical Eng.	127,000	5.5	44	1,170	16
Engineering (Other)	112,000	3.6	38	1,170	25
Environmental Sci.	80,000	6.3	32	1,160	63
Geology	75,000	3.8	46	1,160	46
Health Sciences	76,000	4.6	50	1,030	77
Information Systems	105,000	5.5	27	1,080	38
Mathematics	100,000	5.8	51	1,250	44
Nursing	87,000	2.1	30	1,030	87
Physics	93,000	4.6	61	1,160	21
Public Health	70,000	4.5	51	1,030	77
Humanities					
Art	81,000	6.7	28	1,070	71
Communication	87,000	4.3	23	1,080	69
Dance	–	–	–	1,070	92
English	76,000	6.1	46	1,120	74
Film & Theater	75,000	7.0	38	1,070	66
Foreign Languages	77,000	1.6	53	1,120	70
History	80,000	4.3	51	1,170	43
Interdisciplinary St.	82,000	4.2	41	1,030	62
Music	–	–	–	1,070	45
Philosophy	73,000	4.1	52	1,120	44
Social Sciences					
Anthropology	65,000	7.9	48	1,160	65
Criminal Justice	80,000	3.6	25	1,010	61
Economics	115,000	3.5	42	1,160	35
Education	57,000	2.4	53	1,040	82
Geography	88,000	1.6	35	1,160	44
Political Science	95,000	5.3	50	1,160	56
Psychology	72,000	5.0	52	1,090	82
Public Administration	80,000	2.2	47	1,020	62
Social Work	60,000	1.9	54	1,020	89
Sociology	77,000	3.4	39	1,160	77
Business					
Accounting	97,000	2.6	34	1,080	48
Business Admin.	86,000	4.1	25	1,080	47
Finance	112,000	2.8	32	1,080	30
Hospitality	70,000	3.5	36	1,080	73
Marketing	100,000	4.4	20	1,080	61

Note: Table B10 reports major-level characteristics for the 42 harmonized majors in the analysis. Mid-Career Earnings are mid-career median annual wages from the NY Fed College Labor Market data (American Community Survey, ages 35–45). Unemployment Rate and Graduate Degree Share are from the NY Fed. Mean SAT is from the NPSAS:20 public tables (NCES DataLab), rounded to the nearest ten. Share Female is from IPEDS Completions, averaged across universities in the sample. "–" indicates the source does not report a separate category for that major. The sample covers the 25 primary universities from Fall 2024 through Spring 2026 ($N \approx 148{,}500$).

"BIOL" or separates the same field into "BIMM," "BILD," and "BIPN," making direct cross-institutional comparisons infeasible without a systematic mapping. To understand how the prevalence of AI in the classroom differs across college majors, I harmonize these codes into 42 named categories. I follow the classification framework of Biasi and Ma (2025), who group academic departments into discipline categories aligned with two-digit CIP codes.

I assign the large majority of department codes through direct, unambiguous mappings that are consistent across institutions, such as "ECON" to Economics, "MATH" to Mathematics, and "HIST" to History. I then apply per-university override dictionaries to resolve codes that denote different fields at different institutions, as a given departmental prefix may refer to Chemistry at one university but to Chemical Engineering at another. For composite or cross-listed course codes, I assign the course to the first-listed department. Altogether, I assign a harmonized major for 98 percent of the syllabi in the estimation sample. The remaining 2 percent of syllabi, covering 381 distinct codes that do not map cleanly to one of the 42 categories, are assigned to a residual "Other" category.[22]

For broader comparisons across disciplines, I further aggregate the 42 harmonized majors into five macro-fields covering STEM, Humanities, Social Sciences, Business, and Other. The mapping follows disciplinary boundaries that are largely unambiguous, as all engineering subfields map to STEM, all language and literature fields map to Humanities, and all economics and political science fields map to Social Sciences. Three boundary cases deserve mention, each of which I resolve following CIP conventions and the classification in Biasi and Ma (2025).[23] Table B11 reports the full mapping from department codes to the 42 harmonized majors, grouped by macro-field, along with sample sizes and illustrative sub-fields for each category.

STEM and Humanities are the two largest macro-fields, together accounting for roughly two-thirds of the estimation sample, followed by Social Sciences and Business. The distribution reflects the broad disciplinary coverage of the 25 universities in the sample, spanning large STEM-intensive research universities as well as institutions with substantial liberal arts and professional programs. Table B12 summarizes the macro-field distribution, reporting the number of harmonized majors, the number of syllabi, and the sample share for each of the five groups.

B.6 Course Characteristics

B.6.1 Course Levels

I classify each course into one of three academic levels using the first digit of its course number, following the approach in Biasi and Ma (2022). Under this convention, courses numbered 100 through 299 are classified as basic undergraduate, those numbered 300 through 499 as advanced undergraduate, and those numbered 500 and above as graduate. This heuristic reflects a longstanding convention across U.S. public universities, in which the

[22] The 3,006 syllabi in the residual category come disproportionately from interdepartmental programs and administrative course codes that do not correspond to a single academic discipline. I exclude the residual category from specifications that include major fixed effects, though the main results are not sensitive to this choice.

[23] 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 B11: Harmonized Majors: Sample Distribution and Illustrative Sub-Fields

		Illustrative Sub-Fields		
Harmonized Major	N	(1)	(2)	(3)
Panel A: Business (N = 13,224)				
Business Administration	6,355	Management	Supply Chain	Entrepreneurship
Finance	1,951	Real Estate	Risk Management	Financial Planning
Accounting	1,911	Auditing	Managerial Accounting	
Marketing	1,706	Sales	Digital Marketing	
Hospitality	1,301	Tourism	Hotel Management	Culinary Arts
Panel B: Humanities (N = 47,553)				
Music	7,233	Music Education	Music Theory	Music Performance
Communication	7,225	Journalism	Advertising	Public Relations
English	6,778	Creative Writing	Literature	Comparative Literature
Art	6,772	Art History	Graphic Design	Digital Media
Interdisciplinary Studies	6,472	Honors	Gender Studies	Global Studies
Foreign Languages	4,864	Spanish	French	Linguistics
Film & Theater	3,028	Theater	Film Studies	Drama
History	2,789	American History	European History	
Philosophy	1,821	Religious Studies	Ethics	
Dance	571			
Panel C: STEM (N = 53,731)				
Mathematics	6,941	Statistics	Actuarial Science	Applied Mathematics
Biology	6,588	Microbiology	Neuroscience	Biochemistry
Health Sciences	5,981	Kinesiology	Speech Pathology	Occupational Therapy
Engineering (Other)	5,822	Mechanical Eng.	Aerospace Eng.	Industrial Eng.
Computer Science	3,967	Data Science	Cybersecurity	Software Engineering
Public Health	3,324	Pharmacy	Nutrition	Epidemiology
Nursing	3,206	Graduate Nursing		
Chemistry	2,732	Biochemistry	Medicinal Chemistry	
Information Systems	2,665	Information Technology	Library Science	
Physics	2,513	Astronomy	Space Science	
Electrical Engineering	2,140	Computer Engineering	Electronics	
Architecture	2,048	Urban Planning	Interior Design	Landscape Architecture
Agriculture	2,024	Agronomy	Animal Science	Food Science
Civil Engineering	1,063	Construction	Water Resources	
Environmental Science	1,010	Sustainability	Oceanography	Marine Science
Geology	934	Earth Science	Meteorology	Atmospheric Science
Biomedical Engineering	773	Bioengineering		
Panel D: Social Sciences (N = 30,704)				
Education	9,374	Curriculum & Instruction	Special Education	Reading Education
Psychology	5,107	Cognitive Science	Clinical Psychology	Behavioral Analysis
Political Science	3,745	International Relations	Public Policy	Government
Social Work	2,634	Counseling	Human Development	Family Studies
Economics	2,195	Agricultural Economics	Real Estate	
Criminal Justice	1,798	Criminology	Forensic Science	Law Enforcement
Public Administration	1,659	Public Policy	Urban Planning	
Sociology	1,517	Demography	Social Stratification	
Anthropology	1,345	Archaeology	Biological Anthropology	
Geography	1,330	Geographic Info. Systems	Urban Geography	
Panel E: Other (N = 3,006)				
Other	3,006	Writing & Composition	Kinesiology	Religion

Note: Table B11 presents the distribution of the estimation sample across 42 harmonized majors, along with illustrative sub-fields that map to each category. The sample includes 148,218 section-deduplicated syllabi across the 25 primary universities from Fall 2024 through Spring 2026. Majors are sorted by macro-field, then by descending count within each field.

leading digit of the course number indicates the intended year of study, and it achieves coverage for over 99.9 percent of the syllabi in the sample. Table B13 reports the resulting three-tier distribution for each of the 25 universities.

For the main specifications, I refine this classification into five tiers that distinguish

Table B12: Harmonized Major Categories: 42 Disciplines and 4 Macro-Fields

STEM (17 majors)		
Agriculture	Architecture	Biology
Biomedical Engineering	Chemistry	Civil Engineering
Computer Science	Electrical Engineering	Engineering (Other)
Environmental Science	Geology	Health Sciences
Information Systems	Mathematics	Nursing
Physics	Public Health	
Humanities (10 majors)		
Art	Communication	Dance
English	Film & Theater	Foreign Languages
History	Interdisciplinary Studies	Music
Philosophy		
Social Sciences (10 majors)		
Anthropology	Criminal Justice	Economics
Education	Geography	Political Science
Psychology	Public Administration	Social Work
Sociology		
Business (5 majors)		
Accounting	Business Administration	Finance
Hospitality	Marketing	

Note: Table B12 reports the 42 harmonized majors grouped by macro-field. The STEM category combines the separate Science and Engineering groups from Biasi and Ma (2025). CIP codes follow the NCES 2020 Classification of Instructional Programs (National Center for Education Statistics, 2020). A residual "Other" category (2.0% of the sample) is excluded from the four main macro-fields.

freshman, sophomore, junior, senior, and graduate courses, again using the first digit of the course number. This finer classification is valuable because it allows me to separate introductory courses from more advanced offerings within the undergraduate curriculum, which matters given that AI adoption may differ substantially between a freshman survey course and a senior capstone.[24] The five-tier classification is available for 22 of the 25 universities, all of which follow the standard first-digit convention. The remaining three cases require institution-specific adjustments, which I describe below.

UT Austin uses a three-digit course numbering system in which the first digit encodes the number of credit hours rather than the academic level. I instead use the second and third digits to recover the course level, mapping values 01 through 19 to lower-division undergraduate, 20 through 79 to upper-division undergraduate, and 80 through 99 to graduate.[25] For the three Texas universities that participate in the Texas Common Course Numbering System (UT Dallas, West Texas A&M, and Stephen F. Austin), the four-digit course number additionally encodes credit hours in its second digit. I extract this information to construct a credit-hours variable for these institutions. Similarly, for the seven Florida universities,

[24]The five-tier classification produces approximately 1,050 major $\times$ course-level $\times$ term fixed-effect cells, compared with 774 under the three-tier approach. I use the three-tier classification as a robustness check throughout the analysis.

[25]In the five-tier classification, I assign UT Austin's lower-division courses to the sophomore category and upper-division courses to the senior category, since this system does not distinguish freshman from sophomore or junior from senior.

Table B13: Course-Level Distribution: Three-Tier Classification

		Course Level (%)		
	Syllabi	Basic UG	Adv. UG	Graduate
Texas				
UT Austin	13,832	32	56	11
UT San Antonio	9,623	35	52	14
UT Dallas	7,950	25	46	29
Houston	7,806	23	41	37
Texas Tech	5,536	37	63	0
Texas State	5,427	43	57	0
Stephen F. Austin	4,231	35	46	18
Texas A&M	3,532	22	44	34
West Texas A&M	2,464	37	45	18
Sam Houston State	1,502	36	53	12
Florida				
Central Florida	17,305	17	56	27
Florida International	12,233	31	43	26
Florida Atlantic	8,621	27	48	25
Florida	6,805	22	48	30
Florida State	4,520	24	42	34
Florida A&M	3,043	28	41	30
North Florida	1,854	26	53	20
Georgia				
Kennesaw State	6,959	40	41	19
Georgia	6,237	18	43	39
Georgia State	2,103	68	11	21
Georgia College	1,404	48	31	21
Georgia Southern	1,342	67	14	18
Indiana				
IU Bloomington	6,262	34	40	27
Purdue	4,295	28	43	29
IU Indianapolis	3,234	31	31	38
All universities	**148,120**	**30**	**47**	**23**

Note: Table B13 presents the share of syllabi in each of three course-level categories, by university. I classify courses based on the first digit of the course number, following Biasi and Ma (2025). Basic undergraduate combines courses with first digit 0–2, advanced undergraduate 3–4, and graduate 5+. The sample covers the 25 primary universities from Fall 2024 through Spring 2026, section-deduplicated (N = 148,120).

I take advantage of the Statewide Course Numbering System, which assigns standardized course identifiers that are comparable across institutions within the state. This Florida system is particularly useful because it allows me to compare how different instructors and institutions handle the same course. Altogether, 41 percent of Florida syllabi share an identifier that appears at two or more Florida universities in the sample.

Across the full sample, 30 percent of syllabi correspond to basic undergraduate courses,

47 percent to advanced undergraduate, and the remaining 23 percent to graduate courses. The five-tier breakdown further reveals that freshman and sophomore courses account for 13 and 17 percent of the sample, respectively, while junior and senior courses represent 21 and 26 percent. There is substantial variation across universities, driven by differences in institutional mission and the composition of their course offerings. The graduate share ranges from zero at Texas Tech and Texas State, where course numbers are capped at 499, to 39 percent at the University of Georgia, which reflects UGA's large graduate and professional programs.[26] This cross-university variation in the distribution of course levels underscores the importance of controlling for course level in the analysis, which I do through major $\times$ course-level $\times$ term fixed effects in the main specifications.

B.6.2 Course Characteristics

For four universities in the sample, I recover section-level grade distributions from publicly available registrar records, which provide an objective measure of grading norms and course difficulty that is comparable across institutions. For UT Austin and UT Dallas, I download grade records from each university's public grade distribution repository, covering Spring 2020 through Summer 2025 and Fall 2012 through Spring 2025, respectively. For Texas A&M, I extract grade counts from registrar PDF reports covering Spring 2023 through Spring 2025. For Purdue, I collect letter-grade percentages from the university's public repository for Fall 2021 through Spring 2025.[27] From these records, I compute three summary measures at the course-term level: mean GPA on a 4.0 scale, the share of students receiving a D, F, or Withdrawal, and the share receiving an A.

Section enrollment is a natural control variable in this setting, since course size may be correlated with both the type of instruction offered and the likelihood that an instructor addresses AI in the syllabus. For seven additional universities, I recover section-level enrollment or capacity from publicly available registrar and schedule databases. At IU Bloomington and IU Indianapolis, I collect enrollment, total seats, credit hours, and mode of instruction from the university's public course search system. At Florida, I draw on the Schedule of Courses. At Houston, I derive enrollment from publicly available grade distributions that report the total number of students graded in each section. At Florida Atlantic, Florida International, and Georgia, I collect enrollment and capacity data from each university's public course registration system. Because enrollment data come from heterogeneous sources with varying coverage and precision, I construct a harmonized enrollment measure by assigning each observation the best available source. This ordering prioritizes current observed enrollment over inferred enrollment from seat counts, current-term capacity, his-

[26] Similarly, the share of freshman-level courses is particularly high at Georgia State and Georgia Southern, both of which are broad-access institutions with large introductory course offerings. Table B14 reports the full five-tier distribution by university.

[27] At UT Austin, UT Dallas, and Texas A&M, the grade records report the number of students receiving each letter grade, which also serves as a measure of section enrollment. Purdue's repository reports only grade percentages and does not include enrollment counts.

Table B14: Course-Level Distribution: Five-Tier Classification

		Undergraduate (%)				Grad. (%)
	Syllabi	Freshman	Sophomore	Junior	Senior	Graduate
Texas						
UT Austin	13,832	0	32	0	56	11
UT San Antonio	9,623	19	16	30	22	14
UT Dallas	7,950	10	15	24	22	29
Houston	7,806	11	12	21	20	37
Texas Tech	5,536	16	20	36	27	0
Texas State	5,427	27	17	33	24	0
Stephen F. Austin	4,231	21	15	22	24	18
Texas A&M	3,532	9	13	22	22	34
West Texas A&M	2,464	21	16	26	19	18
Sam Houston State	1,502	22	14	29	24	12
Florida						
Central Florida	17,305	6	12	27	29	27
Florida International	12,233	14	16	23	20	26
Florida Atlantic	8,621	11	16	18	30	25
Florida	6,805	4	17	21	27	30
Florida State	4,520	7	17	19	23	34
Florida A&M	3,043	10	18	20	21	30
North Florida	1,854	8	18	26	27	20
Georgia						
Kennesaw State	6,959	25	16	24	18	19
Georgia	6,237	8	9	16	27	39
Georgia State	2,103	45	23	4	7	21
Georgia College	1,404	28	20	16	15	21
Georgia Southern	1,342	44	23	8	7	18
Indiana						
IU Bloomington	6,262	18	16	23	16	27
Purdue	4,295	10	18	25	19	29
IU Indianapolis	3,234	18	13	15	16	38
All universities	**148,120**	**13**	**17**	**21**	**26**	**23**

Note: Table B14 presents the share of syllabi in each of five course-level categories, by university. I classify courses based on the first digit of the course number, following Biasi and Ma (2025). Freshman includes developmental (pre-collegiate) courses. Graduate combines masters, doctoral, and professional courses. UT Austin records receive mapped labels: lower-division courses are assigned to sophomore, upper-division to senior, and graduate courses to the graduate category (see text). The sample covers the 25 primary universities from Fall 2024 through Spring 2026, section-deduplicated (N = 148,120).

torical averages across prior terms, and enrollment derived from grade distributions.[28] Table B15 reports which course-level covariates are available at each of the 25 universities, and Table B16 lists the underlying data sources along with the terms covered and the specific variables collected.

I also recover semester credit hours, which capture the intensity of each course and vary across both disciplines and course levels. For the Texas universities in the sample, I extract credit hours directly from course numbering conventions. At UT Austin, the first digit of

[28] When current enrollment is not directly observed, I assign the next-best available measure in this order: (i) inferred enrollment from total seats minus open seats, (ii) current-term capacity, (iii) historical enrollment averaged across prior terms, and (iv) enrollment derived from grade distributions.

Table B15: Course Covariate Availability by University

	Course-Level Covariate Families		
University	**Grade Dist.**	**Enrollment**	**Credit Hrs**
Texas (10 universities)			
UT Austin	✓	✓[b]	✓
UT Dallas	✓	✓[b]	✓
UT San Antonio			✓
Houston	✓[c]	✓[c]	✓
Texas State			✓
Texas A&M	✓	✓[b]	✓
West Texas A&M			✓
Stephen F. Austin			✓
Sam Houston State			✓
Texas Tech			✓
Florida (7 universities)			
Central Florida			
Florida State			
Florida Atlantic		✓	✓
Florida International		✓	✓
North Florida			
Florida A&M			
Florida		✓	✓
Indiana (3 universities)			
IU Bloomington		✓	✓
IU Indianapolis		✓	✓
Purdue	✓		
Georgia (5 universities)			
Kennesaw State		✓[d]	
Georgia		✓	✓
Georgia State		✓[d]	
Georgia Southern		✓[d]	
Georgia College			
Total (of 25)	**5**	**13**	**16**

Notes: Table B15 reports the availability of section-level course covariates across the 25 primary universities. All 25 universities have complete coverage for universal covariates not shown here (course level, harmonized major, macro field, institution-level IPEDS characteristics). [b] Enrollment inferred from grade data (sum of graded students). [c] UH grades and enrollment from CougarGrades. [d] Program-level (CIP) enrollment from USG report, not section-level. The sample includes 148,218 section-deduplicated syllabi across the 25 primary universities from Fall 2024 through Spring 2026.

the three-digit course number encodes the number of credit hours. At the remaining Texas universities, I draw on the Texas Common Course Numbering System, where the second digit of the four-digit course number serves the same purpose.[29] At IU Bloomington, I obtain credit hours from the university's public course search system. For the seven Florida universities, I draw on the Statewide Course Numbering System catalog, which provides standardized credit-hour information across all public institutions in the state. Altogether, credit hours are available for 11 of the 25 universities, covering roughly half the sections in

[29]The Texas Common Course Numbering System (TCCNS) standardizes lower-division course identifiers across Texas public institutions. The second digit of each four-digit TCCNS number records the number of semester credit hours for the course.

Table B16: Enrollment and Schedule Data Sources

University	Source	Terms	Variables
Panel A: Section-level enrollment			
IU Bloomington	Course search API	F25–S26	Enrollment, capacity, waitlist, credit hours
IU Indianapolis	Course search API	F25–S26	Enrollment, capacity, waitlist, credit hours
Florida	Schedule of Courses	F25–S26	Sections offered, capacity, credit hours
Houston	CougarGrades	F23–S25	Enrollment (derived from grade counts)
Florida Atlantic	Banner SIS	F25–S26	Enrollment, capacity, credit hours
Florida International	Banner SIS	F25–S26	Enrollment, capacity, credit hours
Georgia	Banner SIS	F25–S26	Enrollment, capacity, credit hours
Panel B: Program-level enrollment			
5 GA universities	USG enrollment report	F21–F25	Headcount by CIP and degree level
6 FL universities	FL Board of Governors	F24	Headcount by CIP
Panel C: Grade distributions			
UT Austin	Public grade data	S20–S25	Letter-grade counts, section enrollment
UT Dallas	Public grade data	F12–S25	Letter-grade counts, instructor weights
Texas A&M	Registrar grade PDFs	S23–S25	Letter-grade counts by section
Purdue	Public grade data	F21–S25	Grade percentages only (no enrollment)

Note: Table B16 reports external data sources used to construct enrollment, credit hours, and grading covariates. F = Fall, S = Spring. Panel A lists sources with section-level enrollment or capacity. Panel B lists program-level benchmarks used in the representativeness analysis. Panel C lists grade distribution sources from which enrollment is also derived where letter-grade counts are available.

the sample. I include a missing indicator for the remaining universities in all specifications that use this variable.

B.6.3 Course Offerings and Enrollment Information

The enrollment and schedule data described above also allow me to assess the representativeness of the syllabus corpus by comparing the distribution of syllabi across departments against the full universe of sections offered at each institution. I supplement these section-level comparisons with program-level enrollment data from state higher-education agencies and degree-completion records from IPEDS, which extend the representativeness assessment to all 25 primary universities.

At the section level, I draw on the registrar databases, scheduling APIs, and grade-distribution portals described in Section B.6.2 to recover the total number of sections offered by department at nine universities in the sample (Table B18). These data cover universities across all four states in the sample. The scheduling systems at IU Bloomington and IU Indianapolis provide section counts directly, while the course registration databases at the University of Georgia and the University of Houston serve the same role. For the remaining five universities, I derive section counts from grade-distribution and schedule-of-courses records at UT Austin, UT Dallas, Texas A&M, Purdue, and the University of Florida. For each of these nine universities, I then compare the department-level distribution of syllabi against the department-level distribution of all sections in these external records, providing

a direct test of whether the collected syllabi are proportionally representative across fields.[30]

In addition to these section-level comparisons, I obtain program-level enrollment data from two state higher-education agencies. The Florida Board of Governors publishes headcount enrollment by CIP code for each institution in the State University System, which I use as a benchmark for the seven Florida universities in the sample. The University System of Georgia publishes credit-hour enrollment by program through its institutional reporting system, covering the five Georgia universities from Fall 2021 through Fall 2025. I also obtain degree completions by CIP code from IPEDS for all 25 primary universities, covering academic years 2016 through 2023. These records provide the broadest benchmark, though they measure degree demand rather than course supply.

Taken together, these data provide external benchmarks at the section level for nine universities and at the CIP level for all 25 primary universities, forming the evidence base for the representativeness analysis in Appendix B.8. Table B16 reports the full set of enrollment and schedule data sources with terms covered and variables available.

B.7 Main Sample

In the preceding sections, I describe the construction of a linked dataset that combines course syllabi with instructor, university, major, and course characteristics drawn from multiple administrative, bibliometric, and survey sources. Altogether, I assemble an estimation sample of 148,218 unique course-section observations across 25 universities in four states, spanning four academic terms between Fall 2024 and Spring 2026, 42 harmonized college majors, and approximately 50,000 unique instructors. I keep exactly one syllabus per institution-department-course-term-instructor combination after the section-level deduplication described in Section B.1.[31]

I recover the largest share of the sample from Texas, with 62,000 syllabi across 10 universities, and from Florida, with 54,000 syllabi across 7 universities. I additionally recover 18,000 syllabi from 5 universities in Georgia and 14,000 from 3 universities in Indiana. Across these institutions, 36 percent of syllabi are in STEM fields, 32 percent in Humanities, 21 percent in Social Sciences, and 9 percent in Business. Table B12 reports the full macro-field distribution. I recover 59,250 syllabi in Spring 2026, 45,812 in Fall 2025, and 21,000 in each of the two earlier terms.[32]

I recover department code, course number, academic term, and course level for effectively all syllabi, and I additionally assign a harmonized major for 98 percent and a course title for 94 percent of the sample. I merge university-level characteristics from IPEDS and Opportunity Insights and major-level characteristics from College Scorecard for all obser-

[30] Because the syllabi are recovered from university portals rather than randomly sampled, this comparison is important for establishing that the corpus does not systematically over- or under-represent particular departments. Table B18 presents these results.

[31] I exclude summer terms from the estimation sample. I keep summer syllabi in the longitudinal archive for the time-series analysis of AI prevalence over time.

[32] Earlier terms have narrower university coverage because several institutions did not yet post syllabi for all terms at the time of collection. Table A1 reports the term-by-university breakdown.

vations, since these covariates are defined at the institution or field level and do not require individual-level linkage. For the instructor-level characteristics described in Section B.3, I recover gender for 83 percent of instructors and academic rank for 68 percent, and I include missing-data indicators for all covariates with incomplete coverage, as described in Section B.3.2. I also recover syllabus text controls for 99.7 percent of the sample. I assess the representativeness of the estimation sample relative to the universe of course offerings at each university in the next section.

Table B17: Sample Summary

Dimension	Count	Detail
Syllabi (sections)	148,218	Section-deduplicated, Fall 2024–Spring 2026
Unique instructors	57,857	Linked to at least one syllabus
Primary universities	25	10 TX, 7 FL, 3 IN, 5 GA
Extension universities	15	Robustness and external validity
US states	4	Texas, Florida, Indiana, Georgia
Academic terms	4	Fall 2024, Spring 2025, Fall 2025, Spring 2026
Harmonized majors	42	Mapped from department codes
Macro-fields	6	STEM, Social Science, Humanities, Business, Health, Other

Note: Table B17 summarizes the estimation sample for the cross-sectional AI-openness analysis. Syllabi counts reflect section-level deduplication, which retains one representative per (institution, department, course number, term, instructor) group. Instructor counts include all individuals appearing on at least one syllabus in the analysis window. The sample includes 148,218 section-deduplicated syllabi across the 25 primary universities from Fall 2024 through Spring 2026.

B.8 Representativeness of Sample

I validate the representativeness of the syllabus corpus by comparing the distribution of syllabi across academic fields against external benchmarks drawn from registrar data, state enrollment records, and federal degree-completion statistics. For the nine universities where I have access to direct section-level data, the department-level distribution of syllabi closely tracks the distribution of actual course sections, confirming that the corpus reflects the underlying composition of course offerings at each institution.

I construct these comparisons by matching the number of syllabi in each harmonized major against the total number of course sections reported by each university's registrar or an equivalent institutional source.[33] The field alignment between the syllabus corpus and the registrar data is strong at nearly all nine universities, ranging from 0.96 at Purdue to 0.77 at UT Austin and Georgia (Table B18). The alignment is highest at universities with centralized syllabus repositories, such as Purdue and IU Bloomington, where the corpus captures a broad cross-section of departments. Figure B1 presents scatter plots for six of these universities, showing that departments with more course sections contribute proportionally more syllabi to the corpus. The main exception is Florida, where data access restrictions limit

[33]The external benchmarks include grade distribution records for Purdue, UT Austin, Texas A&M, and UT Dallas, scheduled class counts for IU Bloomington and IU Indianapolis, CougarGrades for Houston, Banner schedule data for Georgia, and the Schedule of Courses for Florida.

coverage to roughly 11 percent of offered sections, attenuating the measured alignment.

Table B18: Section-level coverage and cross-field alignment

University	Benchmark	External Sections	Syllabi Sections	Coverage (%)	Field Alignment
UH	CougarGrades	6,827	4,936	72.3	0.89
IU Indianapolis	Scheduled classes	5,043	3,234	64.1	0.82
Georgia	Banner schedule	15,326	6,237	40.7	0.77
UT Austin	Grade distributions	60,665	20,366	33.6	0.77
Purdue	Grade distributions	14,216	4,295	30.2	0.96
IU Bloomington	Scheduled classes	21,386	5,789	25.0	0.90
Texas A&M	Grade distributions	21,119	3,532	16.7	0.67
UT Dallas	Grade distributions	82,126	11,777	14.3	0.87
Florida	Schedule of Courses	46,749	4,975	10.6	0.36

Note: Table B18 reports section-level coverage and cross-field alignment for universities with external benchmarks. Coverage is syllabi sections divided by external sections.

I also compare the field-level distribution of syllabi against IPEDS degree completions for all 25 universities, as well as against enrollment data from the Florida Board of Governors and the University System of Georgia.[34] The correlations against degree completions are naturally lower than the section-level results, because syllabi capture the courses that universities offer while IPEDS records the degrees that students complete, and these reflect fundamentally different quantities. Service departments such as Mathematics, English, and Chemistry teach large volumes of introductory courses required by students across many majors, and therefore account for a larger share of the syllabus corpus than of degrees awarded. Conversely, professional programs such as Engineering and Health Sciences produce many degrees relative to their course offerings and account for a smaller share of syllabi than of degree completions. These patterns are consistent across states and universities, and they are exactly what one would expect given the structural difference between course supply and degree demand.

Taken together, these results confirm that the syllabus corpus is representative of the courses offered at each university. This is the relevant benchmark for studying instructional practices, which are determined at the course-offering level rather than at the degree-completion level. The strong section-level alignment across nine universities supports the approach of treating each syllabus as one course section, without reweighting to match degree-completion shares.

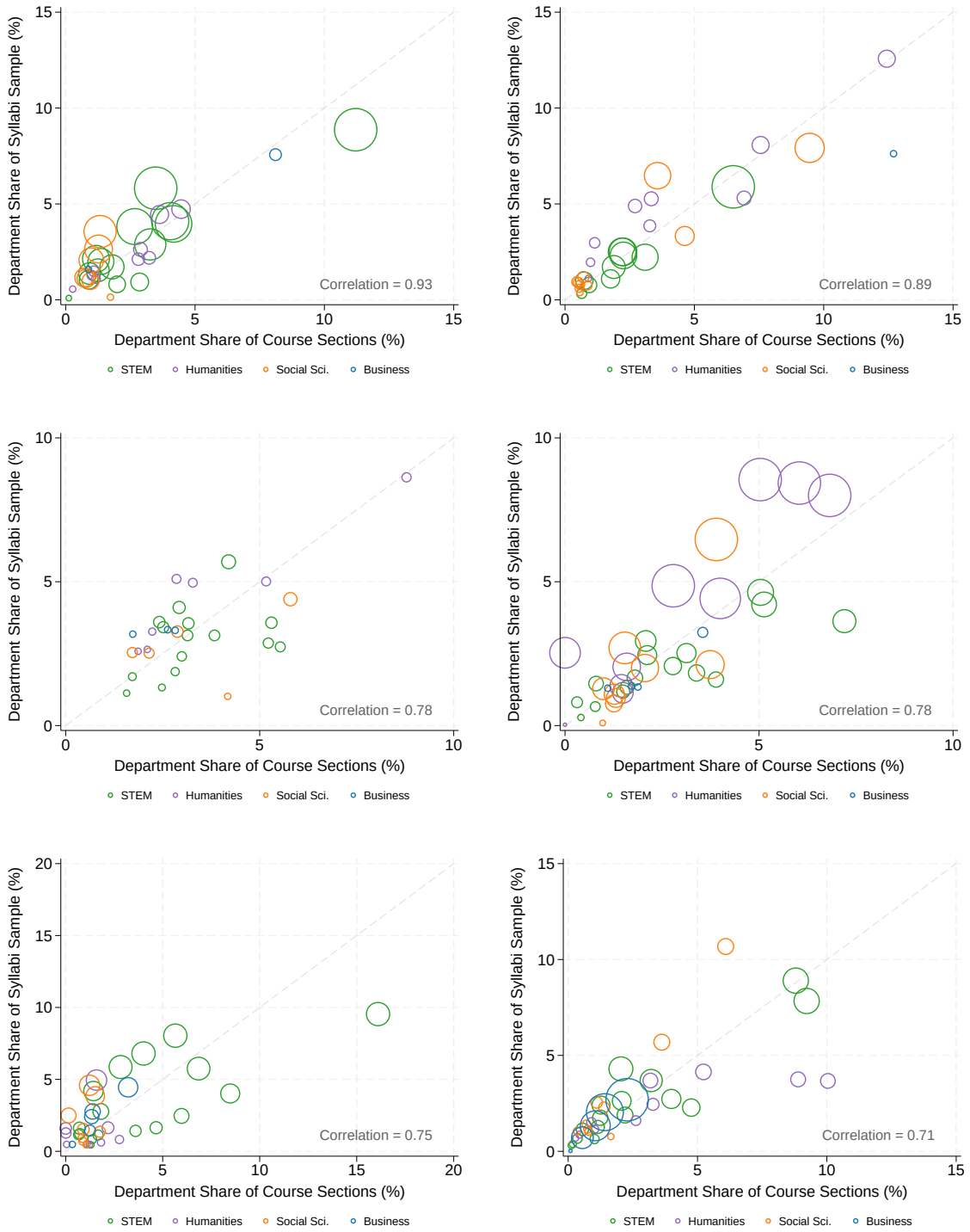
[34] I exclude the Indiana University campuses from this comparison because IPEDS groups IU Bloomington and IU Indianapolis under a single system-level identifier, making campus-level field-share comparisons unreliable.

Table B19: Representativeness of the Syllabus Corpus

Benchmark	**Universities**	**Mean Pearson r**	**Mean Coverage**
Section-level external schedules	8	0.71	0.31
Section-level external schedules, excluding Florida	7	0.78	0.34
IPEDS degree completions	25	0.21	–
Florida system enrollment	7	0.16	–
Georgia system enrollment	5	0.13	–

Table B19 shows the main representativeness benchmarks used to assess the syllabus corpus. The first two rows summarize section-level comparisons against registrar, schedule, and grade-distribution benchmarks for the universities where these sources are available. The remaining rows summarize benchmark correlations between syllabus shares and broader field-level enrollment or completion measures. Coverage is reported only for the section-level benchmarks because the enrollment and completion comparisons do not have a directly comparable section denominator.

Figure B1: Section-Level Representativeness: Department Share Comparisons



Note: Figure B1 presents scatter plots comparing each department's share of syllabi in the corpus (vertical axis) against that department's share of course sections from external registrar data (horizontal axis) for six universities with direct section-level benchmarks. Each circle represents one harmonized major (excluding the residual "Other" category), with circle size proportional to the number of syllabi and color indicating the broad disciplinary group (STEM, Humanities, Social Sciences, or Business). Proximity to the 45-degree dashed line indicates that departments contributing more course sections also contribute proportionally more syllabi to the corpus. External benchmarks are grade distribution records (Purdue, UT Austin, Texas A&M), scheduled classes (IU Bloomington), CougarGrades (Houston), and Banner schedule data (Georgia). The sample covers Fall 2024 through Spring 2026.

Appendix C: Classification Pipeline Details

Section 3.2 presents the eight classification measures and their abbreviated coding guidance. This appendix provides the complete guidance for each measure, including all worked examples of affirmative and negative cases drawn from actual syllabus text. I iteratively refined the guidance through multiple rounds of human review, following standard practice in content analysis (Lombard et al., 2002; O'Connor and Joffe, 2020). Tables C1–C3 present the complete coding guidance for the policy, teaching, and evaluation questions, organized by domain.

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 and adoption questions. I classify these questions from the full syllabus text using two primary open-weight models (GPT-Oss 120B and Qwen 3.5-35B), with a third model (GLM-4.7-Flash) serving as judge to adjudicate disagreements, following the LLM-as-judge framework described in Section 3.3. The abbreviated guidance in Table 2 of the main text omits some examples for concision.

Table C2: Teaching Questions: Full Classification Guidance

Scheduled AI Topic

Question. Does the course schedule, calendar, or module list dedicate at least one class session, week, or module to an AI-related topic?

Yes if the course schedule, calendar, weekly outline, or module list includes at least one entry, associated with a date, week number, module label, or session number, that names an AI-related topic.

(i) "Week 9: AI and Collaborative Composing"
(ii) "Lecture 11: Generative Artificial Intelligence"
(iii) "Module 7: AI Ethics & Bias"
(iv) "Week 6: AI in Healthcare"
(v) "Week 2: Exploring Generative AI Tools like ChatGPT"

No if the session topic is not about AI. An AI assignment or due date in the schedule does not qualify if the session topic is something else.

(i) "Week 5: Professional Ethics. Due: AI Reflection Paper" (topic is ethics, not AI)
(ii) AI mentioned only in a policy section without any AI topic in the schedule

---

In-Class AI Exercise

Question. Does the syllabus describe an in-class activity, demonstration, or exercise involving AI that occurs during scheduled class time (whether graded or ungraded)?

Yes if the syllabus describes a specific activity, exercise, demonstration, lab, or workshop involving AI tools that takes place during scheduled class time. There must be a described activity, not merely a permission or policy about AI use.

(i) "In-class: critique your AI-generated draft in small groups"
(ii) "Student Group Live AI Presentations" during class
(iii) "Writing Workshop: AI activity" as a scheduled lab
(iv) Instructor demonstrates ChatGPT or another AI tool during class

No if the reference is a policy about whether AI may be used during class (permission, not a planned activity), or an AI assignment done outside class.

(i) "During class, we may use AI tools" (permission, not an activity)
(ii) AI mentioned only in a policy section

AI Readings and Materials

Question. Does the syllabus assign readings, videos, tutorials, or other instructional materials specifically about AI?

Yes if the syllabus assigns specific readings, articles, videos, tutorials, podcasts, or other instructional materials about AI, generative AI, or AI tools. The material must appear in a reading list, schedule, or assignment description as assigned or recommended coursework.

(i) "Read: How is AI used in human resources? (Makenzie, 2023)"
(ii) "Video: OpenAI, What Is ChatGPT?" assigned for a specific week
(iii) "Required Text: Artificial Intelligence: A Modern Approach"
(iv) "LinkedIn Learning, Prompt Engineering with ChatGPT"

No if the material is not specifically assigned but merely referenced within a policy statement.

(i) "How to Cite ChatGPT (APA)" appearing only in a policy section
(ii) "Week 12: Artificial Intelligence" as a module title without a named reading
(iii) AI mentioned only in a policy section without assigned AI readings

---

Note: Table C2 presents the complete revised coding guidance for the three teaching questions. These questions use the revised system prompt (Table C4) that explicitly warns that AI use policies are not evidence of course content. I iteratively refined the guidance through multiple rounds of human review to reduce false positives from policy language. I determine the canonical response through the LLM-as-judge framework described in Section 3.3.

C.1 Prompt Structure and Model Configuration

I submit each of the eight 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). The system prompt instructs the model that it is an expert analyst of university course syllabi, that it has access to the full syllabus text, and that it must base its response only on what is explicitly stated in the document. For the six teaching and evaluation questions, the system prompt includes an additional paragraph instructing

Table C3: Evaluation Questions: Full Classification Guidance

Named Graded AI Assignment

Question. Does the syllabus describe at least one specific, named assignment that explicitly requires students to use an AI tool, where the AI use is part of the graded deliverable?

Yes if the syllabus describes at least one specific, named assignment, project, lab, or graded activity that requires students to use an AI tool, where the assignment cannot be completed without AI and the AI use contributes to the student's grade.

(i) "AI Travel Agency project (29.5%)"
(ii) "ChatGPT Assignments (20% of the course grade)"
(iii) "AI Case Study Assignment (30 points)"
(iv) An assignment where AI self-editing is a required graded step
(v) "Write a narrative, then have AI write the same topic. Compare. (30 points)"

No if the syllabus permits AI on existing assignments without requiring it for any specific deliverable. A blanket policy permitting AI is not a named assignment requiring AI.

(i) "You are encouraged to use AI for your assignments" (permission, not a requirement)
(ii) AI mentioned only in a policy section without a named required AI assignment

AI Output Deliverable

Question. Does any assignment require students to submit AI-generated output (transcripts, screenshots, generated text, images, or media) as part of the graded submission?

Critical distinction. A conditional disclosure policy is not an assignment requiring AI output. The test is whether the syllabus mandates AI use and output submission as a designed assignment, or merely requires documentation if AI happens to be used.

Yes if a specific assignment requires students to submit AI-generated output as part of the graded deliverable, including AI-generated text, conversation transcripts, screenshots, images, or code.

(i) "Upload the transcript of your ChatGPT conversation about [topic]"
(ii) "Submit AI prompt and 3-page AI-generated draft"
(iii) "Capture a screenshot of your GPT exchange and embed it in your submission"
(iv) "AI-generated rhetorical banner image required for every discussion post"
(v) AI-generated analysis or content required as a component of the final project

No if the AI output requirement is conditional on whether AI was used, rather than part of a designed assignment.

(i) "If you use AI, you must submit transcripts of your interactions" (conditional, not a designed assignment)
(ii) AI mentioned only in a policy section

Engagement with AI Content

Question. Does any assignment require students to engage with AI-generated content as a structured part of the assignment workflow, where students must produce, evaluate, compare, critique, revise, or respond to AI output as a designed step in the assignment?

Yes if any assignment requires students to engage with AI-generated content as a structured part of the assignment workflow. The student must interact with what the AI produces by evaluating, comparing, critiquing, revising, building upon, or responding to AI output.

(i) "Write a narrative, prompt AI to write the same, compare the two versions"
(ii) "Have AI do the assignment first, improve upon its response, reflect on strengths and weaknesses"
(iii) "Practice motivational interviewing with ChatGPT" (students engage in back-and-forth)
(iv) "Draft, then AI Self-Editing, then Final Paper" (students revise based on AI feedback)
(v) "In groups, identify how the AI draft does or does not fulfill the assignment"

No if the syllabus teaches about AI or schedules an AI topic without requiring students to engage with AI output, or permits AI use without a designed workflow step involving AI output.

(i) "Week 5: Introduction to AI in Healthcare" with no student activity involving AI output
(ii) "You may use AI to brainstorm" (general suggestion, not a designed workflow)

Note: Table C3 presents the complete revised coding guidance for the three evaluation questions. As with the teaching questions, these use the revised system prompt and the LLM-as-judge framework for canonical classification.

the model that AI use policies are not evidence of AI integration into course content or assignments. This additional instruction is important because many syllabi contain policy language about AI that could otherwise be confused with evidence of teaching or evaluation integration.[1] Table C4 presents the system prompt verbatim.

[1]For the two policy questions, where the presence of AI policy language is itself the outcome of interest, I omit the additional instruction. The system prompt is otherwise identical across all eight questions.

Table C4: Classification System Prompt

SYSTEM PROMPT

You are an expert analyst of university course syllabi. You classify whether specific classroom practices related to AI integration are present in the syllabus.

You have access to the FULL syllabus text. Read it carefully. Answer based ONLY on what is explicitly stated in the text. Do not infer intent or read between the lines.

Important: Many syllabi contain AI use policies (permissions, prohibitions, disclosure rules). An AI policy is NOT by itself evidence of AI in the course content, schedule, or assignments. Ensure your evidence comes from course content, not rules about AI.

You must respond with a complete, well-formed JSON object. Provide thorough justifications.

Note: Table C4 presents the system prompt used for the six teaching and evaluation questions. The "Important" paragraph reduces false positives from AI policy language being confused with evidence of classroom integration. The two policy questions (has AI policy and permits AI) use a version of this prompt that omits the "Important" paragraph, since for these questions AI policy language is the evidence being sought.

For each syllabus-question pair, I construct a user message that inserts the question text, the corresponding coding guidance from Tables C1–C3, and the full syllabus text. In 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 (Haaland et al., 2025). Table C5 presents the user message template verbatim.

I use GPT-Oss-120B and Qwen 3.5-35B as the two primary classifiers and GLM-4.7-Flash as the judge throughout the classification pipeline. For all three models, I set the temperature to zero and fix the random seed, aiming to maximize deterministic outputs across repeated runs.[2] All questions receive the full syllabus text. The adoption measure is applied to instructor-authored text remaining after the boilerplate decomposition procedure described in Appendix B.2.2, but the classification models themselves see the full text for every question.

C.2 LLM-as-Judge Implementation

Section 3.3 describes the LLM-as-judge framework I use to resolve disagreements between the two primary classifiers and to select evidence passages when both classifiers agree that a measure is present. This appendix presents the judge prompts verbatim. When both primary models agree that a teaching or evaluation measure is present, I use the judge to select or refine the single most compelling evidence passage from the syllabus for use in the qualitative analysis described in Section 4. Table C6 presents this evidence selection prompt. When the two primary models disagree, I use the judge to adjudicate. The adju-

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

Table C5: User Message Template for Syllabus Classification

USER MESSAGE TEMPLATE

QUESTION

```
{question text}
```

CODING GUIDANCE

```
{guidance text}
```

SYLLABUS TEXT

—

```
{full_text, capped at 60,000 characters}
```

—

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 C5 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 process each syllabus-question pair independently through a separate API call, so that a model's response to one question does not influence its evaluation of the remaining questions.

dication prompt instructs the judge that AI policy language is not evidence of classroom integration, fitting in with the design of the teaching and evaluation questions.[3] Table C7 presents the teaching and evaluation adjudication prompt.

Table C8 reports the distribution of agreement patterns and judge actions across all eight questions. The two primary models agree in the vast majority of cases, with the judge adjudicating between 1 and 6 percent of classifications depending on the question. In approximately half of adjudication cases the judge rules the measure present and in the remainder

[3]For the two policy questions, the adjudication prompt is identical except that it replaces this instruction with a generic rule to classify the measure as absent only when the affirmative rater's evidence is taken out of context or does not satisfy the coding guidance.

Table C6: LLM-as-Judge Prompt: Evidence Selection Mode

SYSTEM PROMPT

You are a research assistant reviewing university course syllabi. Two independent raters both classified the following as YES for the question below. Your task is to select the single best evidence quote from the syllabus text that most clearly supports this classification. You may refine the evidence passages suggested by the raters, combine them, or find a better passage entirely.

USER MESSAGE TEMPLATE

QUESTION: {`question text`}
CODING GUIDANCE: {`guidance text`}

RATER SUGGESTIONS (both agreed YES):
- Rater A evidence: "{*rater_a_evidence*}"
- Rater B evidence: "{*rater_b_evidence*}"

SYLLABUS TEXT:
—
{`full_text`}
—

Select the single most compelling evidence passage from the syllabus. Quote it exactly as it appears in the text. If one rater's evidence is clearly better, use it. If both are partial, combine or find a better passage.

Return ONLY a JSON object:
`{"evidence": "exact quote from syllabus", "evidence_source": "rater_a" or "rater_b" or "independent"}`

Note: Table C6 presents the evidence selection prompt used for the six teaching and evaluation questions when both primary classifiers agree that a measure is present. The judge reads the full syllabus and both models' evidence quotations, then selects or refines the single most compelling evidence passage. I anonymize rater identities as "Rater A" and "Rater B" and randomize their ordering across cases to prevent positional bias (Zheng et al., 2024).

absent, indicating that these cases reflect genuine ambiguity in the syllabus text rather than systematic differences between the two models.

C.3 Cross-Model Agreement

I assess the reliability of the multi-model classification by computing pairwise agreement between the two primary models across the eight questions. Table C9 reports Cohen's kappa and percent agreement for each question. The two models agree in 98 percent of cases on average, with agreement highest for questions with clear textual markers and somewhat

Table C7: LLM-as-Judge Prompt: Teaching and Evaluation Adjudication

SYSTEM PROMPT

You are a research assistant adjudicating a disagreement between two independent raters who classified a university course syllabus. One rater said YES and the other said NO for the question below. Read the full syllabus text and decide which rater is correct. Base your decision ONLY on what is explicitly stated in the text.

USER MESSAGE TEMPLATE

QUESTION: {question text}
CODING GUIDANCE: {guidance text}

RATER A's assessment:
- Answer: {rater_a_answer}
- Reasoning: "{*rater_a_justification*}"
- Evidence: "{*rater_a_evidence*}"

RATER B's assessment:
- Answer: {rater_b_answer}
- Reasoning: "{*rater_b_justification*}"
- Evidence: "{*rater_b_evidence*}"

SYLLABUS TEXT:
—
{full_text}
—

ADJUDICATION INSTRUCTIONS:

Start by examining the YES rater's evidence. If the evidence points to concrete course content—a scheduled topic, a described activity, an assigned reading, a named assignment, or a required deliverable—then the answer is YES. Take the evidence at face value unless you find a specific reason it does not meet the coding guidance.

The answer is NO only if the YES rater's evidence comes exclusively from an AI USE POLICY (permissions, prohibitions, disclosure rules) rather than from course content, schedule, or assignments. Policies about AI are NOT evidence of AI integration into teaching or evaluation.

If your verdict is YES, find and quote the single most compelling evidence passage from the syllabus text. You may use the YES rater's quoted passage, refine it, or find a better one in the syllabus. Quote exactly as it appears in the text.

Return ONLY a JSON object:

```
{"verdict": "yes" or "no", "confidence": "high" or "medium" or "low",
"reasoning": "one sentence", "evidence": "best evidence quote from
syllabus if yes, empty string if no"}
```

Note: Table C7 presents the adjudication prompt used for the six teaching and evaluation questions when the two primary classifiers disagree. The judge receives both raters' answers, justifications, and evidence, then makes an independent determination. The adjudication instructions instruct the judge that AI policy language is not evidence of classroom integration, reflecting the design intent of these questions. The corresponding prompt for the two policy questions is identical except that it replaces the policy-exclusion rule with a generic instruction to rule NO only when the YES rater's evidence is taken out of context or does not satisfy the coding guidance.

Table C8: LLM-as-Judge Usage Statistics

		Primary Classifier Agreement (%)			Of Disagree:
Measure	N	Both Absent	Both Present	Disagree	Judge → Present (%)
Has AI Policy	84,242	5.6	91.7	2.7	72.1
Permits AI	84,242	60.4	33.8	5.8	74.0
Scheduled AI Topic	84,242	91.7	6.2	1.9	53.7
In-Class AI Exercise	84,242	97.2	1.2	1.4	49.3
AI Readings/Materials	84,242	94.4	3.0	2.5	54.8
Graded AI Assignment	84,242	96.6	1.8	1.4	55.6
AI Output Deliverable	84,242	98.1	0.7	0.9	46.9
Engagement with AI	84,242	96.6	1.9	1.4	43.4

Note: Table C8 reports the distribution of primary classifier agreement patterns and judge outcomes for all eight headline questions. Both Absent and Both Present indicate the share where both primary classifiers (GPT-Oss and Qwen) agree. Disagree indicates the share where the two primaries reach different conclusions and the judge (GLM-4.7-Flash) adjudicates. Judge → Present reports the share of disagreement cases where the judge determined the measure was present. The two policy questions (Has AI Policy and Permits AI) use the TE2 judge batch. The six teaching and evaluation questions use the TE6 batch. The sample includes all classified syllabi across the 25 primary universities from Fall 2024 through Spring 2026.

lower for questions requiring more interpretive judgment.[4]

Table C9: Primary Classifier Agreement (Detailed)

	Prevalence (%)		Agreement		Distribution (%)		
Measure	GPT-Oss	Qwen	κ	(%)	Both NO	Both YES	Disagree
Has AI Policy	93.2	92.9	0.788	97.3	5.6	91.7	2.7
Permits AI	36.4	37.0	0.875	94.2	60.4	33.8	5.8
Scheduled AI Topic	6.7	7.8	0.856	98.1	91.8	6.2	1.9
In-Class AI Exercise	2.2	1.6	0.616	98.6	97.4	1.2	1.4
AI Readings/Materials	3.7	4.7	0.689	97.5	94.5	3.0	2.5
Graded AI Assignment	2.4	2.6	0.714	98.6	96.8	1.8	1.4
AI Output Deliverable	1.2	1.3	0.604	99.0	98.3	0.7	1.0
Engagement with AI	2.7	2.4	0.725	98.6	96.8	1.9	1.4

Note: Table reports pairwise agreement between the two primary open-weight classifiers (GPT-Oss 120B and Qwen 3.5-35B) for the eight headline measures. Prevalence indicates the share of syllabi each model classifies as positive. $N = 84,242$ classified syllabi (25 universities, 2024–2026).

C.4 Human Validation Details

I recruited four research assistants to independently classify a sample of 60 syllabi across the full set of eight measures. I drew the syllabi using stratified sampling by AI mention status, institution, and classification profile to ensure representation across the distribution of instructor responses. An additional 10 syllabi served as calibration examples, reviewed jointly by all four raters and the author before the independent coding phase, following the

[4] Among the full estimation sample, 47 percent of syllabi do not mention AI and are coded as absent on all measures without invoking the classification models. The agreement statistics reported here are computed over the remaining syllabi that are sent to the classifiers.

recommended calibration protocol in Lombard et al. (2002).[5] Each rater received the complete coding guidance (Tables C1–C3), a brief orientation document describing the classification task, and the full text of each syllabus in PDF format. Raters worked independently and were not aware of the LLM classifications during the rating process.

Section 3.4 reports the headline validation results. Tables C10–C12 present the per-question breakdowns underlying these results. Table C10 reports the human inter-rater reliability for each of the eight questions. Table C11 compares the accuracy of each classification method against the human modal response, showing that the judge-based canonical classification, GPT-Oss alone, and Qwen alone all exceed 90 percent accuracy on every question. Table C12 reports the full Alternative Annotator Test results, following Calderon et al. (2025).

Table C10: Human Inter-Rater Reliability

Question	**Agr. (%)**
Has AI Policy	92
Permits AI	94
Scheduled AI Topic	87
In-Class AI Exercise	97
AI Readings	91
Graded AI Assignment	99
AI Output Deliverable	99
Engagement with AI	100
Average	**95**
Syllabi = 60	

Note: Table C10 presents inter-rater reliability statistics for four human raters who classified a sample of 60 syllabi according to eight binary questions. Agr. = pairwise agreement rate.

[5]The calibration syllabi are excluded from all reliability and validation calculations. The rating design assigns 20 core syllabi to all four raters and rotates the remaining 40 so that each rater classifies 50 syllabi total.

Table C11: LLM Classification vs. Human Modal Response (%)

	LLM-Judge	GPT-Oss	Qwen
Question	Acc.	Acc.	Acc.
Has AI Policy	93	95	92
Permits AI	93	95	93
Scheduled AI Topic	91	93	91
In-Class AI Exercise	93	90	95
AI Readings	92	92	95
Graded AI Assignment	98	98	98
AI Output Deliverable	98	97	100
Engagement with AI	98	97	98

Note: Table C11 compares three classification approaches against the modal human response. LLM-Judge is the judge-adjudicated canonical classification. GPT-Oss and Qwen are the two primary raters' raw classifications before judge adjudication. Acc. = accuracy (%).

Table C12: Alternative Annotator Test: LLM-Judge

Question	N	Skip	Wins	Losses	Ties	Win (%)	W+T (%)
Has AI Policy	60	0	5	3	52	8.3	95.0
Permits AI	60	0	3	3	54	5.0	95.0
Scheduled AI Topic	60	0	5	4	51	8.3	93.3
In-Class AI Exercise	60	0	0	3	57	0.0	95.0
AI Readings	60	0	4	5	51	6.7	91.7
Graded AI Assignment	60	0	0	0	60	0.0	100.0
AI Output Deliverable	60	0	1	1	58	1.7	98.3
Engagement with AI	60	0	0	1	59	0.0	98.3
Overall			**18**	**20**	**442**		**95.8**

Note: Table C12 reports results from the Alternative Annotator Test (Calderon et al., 2025). For each syllabus, one human rater is held out and the remaining raters' majority response serves as the benchmark. Win = LLM-Judge agrees with majority, held-out rater does not. Tie = both agree or both disagree. W+T rate above 50% indicates human-level performance.