Mapping the Pathways: A Comparative Analysis of AI/ML/DS Prerequisite Structures in R1 Institutions in the United States

Rose Niousha
Electrical Engineering & Computer Science
University of California, Berkeley
Berkeley, USA
rose.n@berkeley.edu

Dev Ahluwalia*
Electrical Engineering & Computer Science
University of California, Berkeley
Berkeley, USA
dahluwalia@berkeley.edu

Michael Wu*
Electrical Engineering & Computer Science
University of California, Berkeley
Berkeley, USA
michael.wu1@berkeley.edu

Lisa Zhang†
Mathematical & Computational Sciences
University of Toronto Mississauga
Toronto, Canada
lczhang@cs.toronto.edu

Narges Norouzi†
Electrical Engineering & Computer Science
University of California, Berkeley
Berkeley, USA
norouzi@berkeley.edu

***Abstract*—This Research Full paper focuses on the challenges in artificial intelligence, machine learning, and data science education—referred to as "artificial intelligence" courses hereafter—often characterized by extensive prerequisites that limit student access. We analyze the course structures and prerequisites of these courses in computing departments at 50 Research-1 institutions in the United States, recognized for their "Very High Research Activity." Our methodology involves analyzing course syllabi to examine the structure and prerequisites of these courses, using open coding to develop a unified codebook to identify prerequisites and determine the earliest exposure levels for students. A clustering analysis was also conducted to identify common and differing curriculum approaches among institutions. Results show that data science courses require less initial exposure, while artificial intelligence and machine learning courses require more prerequisites. Standard requirements for artificial intelligence courses include basic data structure (Computer Science 2) and algorithms, with machine learning courses requiring more mathematics preparation. Moreover, public institutions offer advanced courses with more prerequisites compared to private institutions. Overall, this study recognizes considerable diversity in curricular frameworks across Research-1 institutions and encourages institutions to revise curricula to broaden access to artificial intelligence education and increase participation in research.**

***Index Terms*—Curriculum Design, Artificial Intelligence, Machine Learning, Data Science, Prerequisites, Academic Retention**

I. Introduction

There has been increased interest in areas related to Artificial Intelligence (AI), Machine Learning (ML), and Data Science (DS) (referred to as *AI* courses here on) as well as applications of these areas in various fields [1]–[4]. Integrating such disciplines in the educational framework is highly demanded in any industry today [5] as it prepares students to drive innovation and address complex challenges in a variety of sectors [6], [7]. However, AI fields face challenges in ensuring diversity [8], [9]. Studying and teaching *AI* courses have been reported by both students and instructors as challenging [6], [10], [11] and Computer Science (CS) educators often hesitate to teach *AI* subjects to non-major students due to their complexity [12]. Learning these topics often requires a grasp of a range of mathematical prerequisite skills like calculus, linear algebra, and probability, as well as programming skills [13]–[16]. Such requirements can potentially delay students from engaging in these fields at early stages.

Our study analyzes R1 (Research-1): Doctoral Universities in the United States (U.S.) with "Very High Research Activity" from the Carnegie classification[1]. We specifically focus on R1 institution since we want to investigate the process and the associated timeline that prepares students for AI research after taking relevant courses. Our study aims to compare and contrast the various curriculum designs across institutions to evaluate student accessibility to *AI* courses and understand the effects of a light prerequisite structure on students' readiness for AI research.

We seek to answer the following three research questions:

- **RQ1**. What is the earliest exposure of students to the *AI* curriculum in R1 computing departments in the U.S.?
- **RQ2**. What common approaches are institutions using to structure prerequisites for *AI* courses in R1 computing

*These students contributed equally to this work.
† These faculty contributed equally to this work as co-advisors.

[1] https://carnegieclassifications.acenet.edu/institutions/

departments in the U.S.?

- **RQ3**. What different approaches are institutions using to structure prerequisites for *AI* courses in R1 computing departments in the U.S.?

This study intends to guide educational institutions in creating learning pathways that are rigorous yet flexible and align with the evolving demands of *AI* courses to balance the need for foundational knowledge with accessible research entry points into these fields.

II. Related Work

Previous studies showed the importance and impact of *AI* courses on a diverse group of students. Ng et al. [17] emphasized the importance of AI in education, noting a shift from traditional university-level CS to inclusive methods for K-12 and non-technical learners. In addition, the AI curriculum can be taught to non-technical majors as well. Menkhoff and Lydia Teo [18] conducted a case study with its chatbot workshop to non-technical undergraduate students to teach basic skills in the first-year AI course. Moreover, de Freitas and Weingart [12] demonstrated that AI concepts can be effectively taught to non-technical students with a curriculum specifically designed for first-year students. These findings confirm the versatility and applicability of AI education across diverse technical levels of students.

The CS community is interested in examining the role of prerequisites in education. Krause-Levy et al . [19] examined instructors' views on computing education prerequisites, revealing the complexity of prerequisite course implementation and the challenges instructors face when aligning course content with student preparation and curriculum requirements. Another study by Krause-Levy et al. [20] analyzed demographic factors affecting students' readiness for an advanced DS course. The study revealed disparities in prerequisite proficiency among different student groups and emphasized the importance of addressing diverse educational backgrounds in the field. Krause-Levy et al. [21] found significant correlations between students' success in prerequisite courses and their performance in advanced courses, indicating that doing well in prerequisite courses strongly predicts success in subsequent advanced computing courses.

There are diverse approaches in which prerequisites are integrated into *AI* courses. Li and Liu [22] emphasize the importance of core theoretical courses like "Matrix Computation" and "Optimization" as prerequisites in AI. Another study found that incorporating just-in-time prerequisite reviews, consisting of targeted questions and instructional videos before each lecture, effectively addressed knowledge gaps in ML courses [13].

Earlier works demonstrate that *AI* can be taught without extensive technical prerequisites. Barretto et al. [8] recommends enhancing AI and ML participation by adding courses on their societal and cultural impacts, targeting underrepresented students interested in these broader topics over technical aspects. Moreover, in a month-long course teaching ML and Natural Language Processing (NLP) to high school students who were taught programming in the course itself, students enhanced their understanding of AI. They underscored the importance of foundational programming skills in AI education [23]. Allen et al . [11] advocate combining theoretical and practical teaching methods, tailored support to address students' mathematics anxiety and confidence issues, and adaptive teaching strategies for complex AI concepts to broaden participation.

With the recent technological advancements, interest in conducting AI research has risen dramatically, particularly with undergraduates [24]. However, AI research is often not available to students without related experience in coursework. Access to *AI* courses at an early stage empowers students to engage in undergraduate research earlier. This has important implications for student retention, as Bhattacharyya et al . [25] found that engaging in undergraduate research increased student retention and graduation rates. Moreover, the Early Research Scholars Program (ERSP) at UC San Diego demonstrated that participants had higher GPAs, improved retention rates, and a stronger sense of belonging and confidence compared to control groups, which consisted of students who did not participate in the program and followed the standard curriculum [26]. These findings suggest that early research experiences help students develop essential academic skills and foster a positive academic identity, which is vital for success. Similarly, the Students Tackling Advanced Research (STAR) Scholars Program at Drexel University found that both STEM and non-STEM students experienced significant learning gains, particularly in understanding research work, developing independent working skills, and improving communication abilities [27]. These results highlight the benefits of early research opportunities, supporting diverse student populations and fostering broader interest in pursuing further research and advanced studies. Additionally, a review by Linn et al. [28] emphasized the importance of effective mentoring in undergraduate research experiences. They found that while students highly rate undergraduate research, more rigorous research designs, and validated assessments are needed to understand and fully optimize their benefits. Effective mentoring is critical in helping students integrate scientific practices and concepts, ensuring they gain a coherent and comprehensive understanding of their research field.

Our framework is grounded in educational accessibility and curriculum design. Prior research has primarily focused on demographic analyses [29], instructor perspectives [23], [30], and the relationship between prerequisite courses and student performance in specific computing courses [31]. These studies emphasize how variations in student backgrounds and instructional strategies impact academic outcomes. Building on these foundations, our study systematically examines curriculum structures, particularly *AI* courses across R1 institutions, analyzing how prerequisites influence course accessibility and AI research preparation eventually.

III. METHODOLOGY

A. Data Collection

We randomly sampled 50 R1 universities. Our sample included 37 public universities and 13 private universities. To be included in our sample, the university must have a computing department advertised on its website. Universities must also have at least one AI, ML, or DS course offered by their computing department. We excluded and resampled two universities that were originally in our sample but did not have a computing department and did not offer any undergraduate courses. The geographic distribution of the sampled universities across the U.S. is illustrated in Figure 1.

We then collected, for each university, a list of undergraduate *AI* courses offered by their computing department. First, we identified relevant courses from each university's academic calendar and classified them as AI, ML, or DS based on the 2023 offering guidelines by examining the course syllabi. This process ensured that the selected courses were representative of their respective subjects. While we considered potential edge cases in categorization, such instances were not prevalent. To maintain relevancy and accuracy, we excluded special topic courses and those that were not recently offered.

We collected information for each relevant undergraduate course, including course type (AI, ML, or DS), course name, level, immediate prerequisites, and offering frequency. Course levels were defined as "Introductory," "Intermediate," "Advanced," or "Cross-listed" (open to both undergraduate and graduate students) according to each university's course numbering scheme reflecting the complexity of the course content. Offering frequency was categorized into "More than once a year," "Once a year," or "Less than once a year," based on the universities' academic schedules.

Out of our sampled universities, three universities had less than 2 *AI* courses, 36 offered between 2 and 3 *AI* courses, and 11 universities offered more than 3 *AI* courses, as seen in Figure 2.

Further, as shown in Table I, our sampled courses included 55 AI, 54 ML, and 40 DS courses. The course level from "Introductory" to "Cross-listed" is coded as levels 1 to 4. AI and DS had a minimum course level of introductory, whereas ML had a minimum intermediate. All three subjects were most often offered at the advanced level. Further, AI and ML courses were most often offered once a year, whereas DS courses were most often offered more than once a year.

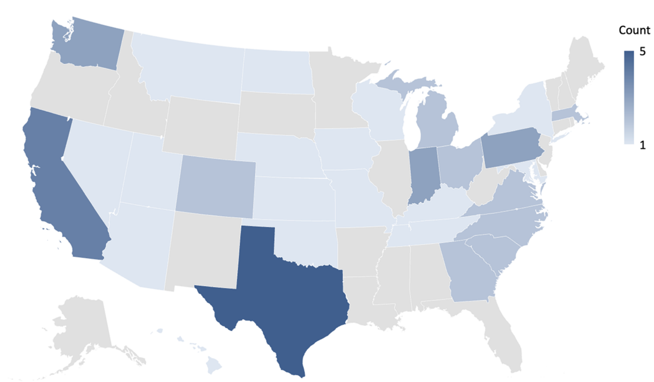

Fig. 1. Sampled university distribution across the U.S.

B. Coding of Prerequisites

To structure the diverse set of prerequisites provided by each university, we first gathered information on the immediate prerequisites of each course. Three researchers, also authors of this study, independently conducted open coding of each course, resulting in three sets of codes. The open coding involved identifying recurring themes and patterns in the prerequisite descriptions, allowing us to generate initial categories. Next, the three sets were merged into a single codebook. During this phase, we held multiple review sessions to discuss and reconcile differences in the codes, ensuring consistency and accuracy. We removed duplicates and aligned our codes with courses provided in ACM's CS Curricula 2023 (Version Gamma)[2]. The course names (codes) in our final codebook is shown in Table II.

To enhance reliability, two researchers then coded the prerequisites for each course using the final codebook. There was a 10% overlap in the coding process to check reliability. After this phase, the coders discussed with a third mediator to iterate upon the overlapping codes. Any discrepancies were discussed during this part, and additional modifications were made to the coding, particularly in instances where courses spanned multiple topics. This iterative process achieved high inter-rater reliability, as evidenced by a Cohen's kappa score of 0.95, high substantial agreement [32]. This iterative process ensured that our coding scheme was robust and could be applied uniformly across different institutions.

C. First Exposure Level

To understand when students can first access *AI* courses (RQ1), we focused on identifying the initial entry point for engagement with these courses. Thus, we analyzed how the *the first exposure level* (the depth of each course within the prerequisite hierarchy) differed between course types. For example, a course with several layers of prerequisites would indicate a higher first exposure level compared to a course that is directly connected to the initial course. We detailed the differences in the distribution of first exposure levels between *AI* courses by plotting a histogram for each course type.

D. Common Prerequisite Approach

To understand the common prerequisite approaches (RQ2), we developed a Sankey graph plotting each course's prerequisite chain. We chose this graph to visualize which prerequisites were most often required to determine the most foundational prerequisites for each of the *AI* courses. For each of the 50 institutions in our sample, we construct a prerequisite graph: a directed graph where the nodes are courses needed to take one of the *AI* courses, and the edges denote direct prerequisite

[2]https://csed.acm.org/wp-content/uploads/2023/09/Version-Gamma.pdf

TABLE I
DISTRIBUTION OF COURSE LEVELS FOR *AI* COURSES.

Course Type	Number of Courses	Average Course Level	Min Course Level	Max Course Level	Mode Course Level	Mode Frequency
AI	55	2.96	1	4	3	Once a year
ML	54	3.02	2	4	3	Once a year
DS	40	2.33	1	4	3	More than once a year

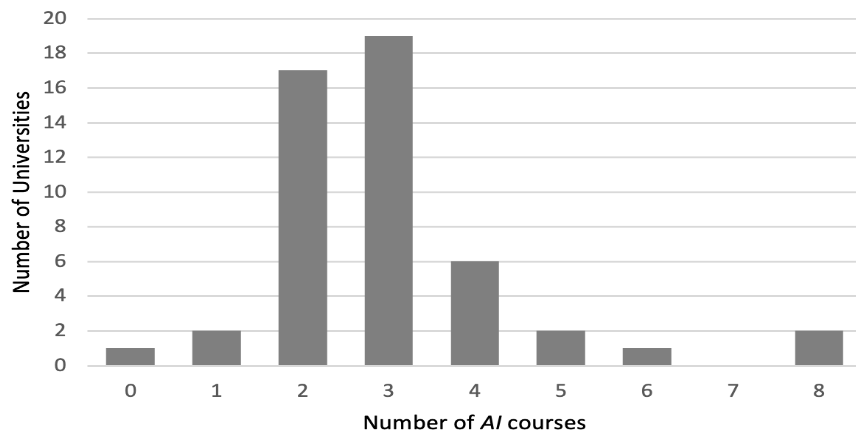

Fig. 2. Histogram of the number of *AI* courses per university.

TABLE II
CATEGORIZATION OF CODE NAMES.

Category	Code Names
Mathematics	Discrete Mathematics, Linear Algebra, Multi-variable Calculus, Probability, Statistics, Single-variable Calculus
Computing	Algorithms, Architecture and Organization, Artificial Intelligence, CS1, CS2, Data Management, Data Science, Machine Learning, Foundations of Programming Languages, Software Engineering
Others	"Society, Ethics and Professionalism", Signal Processing

relation. This approach allowed us to trace each course to its direct and indirect prerequisites.

E. Clustering Analysis

We used a clustering method to reveal the characteristics of different institutions and their corresponding courses (RQ3). Specifically, we analyzed the data using k-means clustering to create course and university clusters. Our features for the course level clustering use five groups of features: Type of course, course level, course frequency, prerequisite statistics, and prerequisite courses using the codebook above. These groups of features were necessary to define how *AI* is currently being offered in R1 institutions; taken together, they capture statistics related to access to courses. By considering course levels and prerequisites, we could determine the first term where students can enroll in *AI* courses. Further, the course frequency metric offered insights into barriers to access to classes even after prerequisites are completed. Through our analysis, we aimed to expose the roadblocks students may face to access these courses and, therefore, potential roadblocks in their pursuit of research opportunities. Our university-level features included: course level, frequency, immediate prerequisite courses, total prerequisites, and first exposure level, as calculated by the averages over the relevant *AI* courses at the university.

For k-means clustering, we used one-hot encoding to represent categorical features. The number of clusters (k) was chosen based on the highest silhouette score between 2 and $\sqrt{\frac{n}{2}}$, a common heuristic to find an optimal k, where n is the number of data points. This resulted in $k = 8$ clusters for the course level and $k = 4$ clusters for the university level clustering.

IV. RESULTS

A. RQ1: Distribution of Exposure Level:

We compared the first exposure levels of different course types, as shown in Figure 3. For AI courses, the majority have a first exposure level of 2, indicating they are typically encountered after completing foundational prerequisites. Only a small number are available at the first exposure level of 1, and none at level 0, indicating a higher entry barrier. ML courses show an even higher entry barrier, with most at the first exposure level 2. Some ML courses reach a first exposure level of 5, suggesting they require several semesters of prerequisite coursework. DS courses tend to be more introductory, with a significant number having a first exposure level of 0, indicating they are available to students in their first semester or year without prerequisites. This early accessibility contrasts sharply with AI and ML courses. These results suggest that DS courses are generally more accessible early in students' academic careers, while AI and ML courses have delayed initial exposure.

B. RQ2: Relationship Between Prerequisites

As shown in Figure 4, all prerequisite chains were based on one of three course types: CS1 for more programming-related courses and Linear Algebra or Single-variable calculus for more mathematics-related courses. DS courses generally require minimal prerequisites, with only a few universities requiring additional courses like CS2 or Algorithms. In other words, DS courses have fewer and more direct prerequisites, making them more accessible with minimal entry barriers. AI courses typically require a strong programming background, starting from CS1. AI courses also frequently list CS2 and Algorithms as prerequisites, indicating a need for advanced programming and algorithmic knowledge. ML courses require more extensive preparation, often involving long prerequisite chains on both the programming and mathematics sides, starting from CS1 and Single-variable Calculus. ML courses exhibit more extensive prerequisite chains, usually beginning with Single-variable Calculus and extending through Linear Algebra, Probability, and Statistics, highlighting the substantial mathematical foundation needed for ML concepts. CS1 is a foundational prerequisite for many AI and ML courses, reflecting the importance of introductory programming skills.

C. RQ3: Course Level Clustering

To understand the different approaches institutions use to structure *AI* prerequisites, we studied the clustering of different courses offered by the sampled institutions. Figure 5 is a heatmap of the course clusters. On the x-axis, we have our selected features as described previously, and on the y-axis, we have 8 clusters. The color gradient indicates how important each feature was to the cluster. A lighter shade, such as yellow, indicates that the feature's average value is higher in that cluster than in others. In other words, clusters with lighter shades have a higher importance of that feature compared to clusters with darker shades. For example, the first cluster has a yellow color for AI, which means that its cluster is defined by having AI courses. A darker shade, such as navy blue, indicates that the feature's average value is lower in that cluster. Notably, the dark blue in Cluster 6 for the "Once a Year" feature suggests that the cluster mainly contains courses not offered once a year. Lastly, the medium-tone colors, such as teal, indicate that a feature was not as relevant to that cluster as to other clusters.

Using these definitions, we find that features including course type, course level, and some prerequisite features help define our clusters, as seen by the light yellows and dark blues in those regions. However, features such as exposure level and number of prerequisites do not define our clusters as much, except for cluster 4, where relatively low prerequisite levels are relevant. The eight clusters have the following characteristics:

1) AI courses, advanced level, offered once a year, common prerequisites include CS1 and CS2 ($N = 45$).
2) ML courses, advanced level, offered once a year, low number of immediate prerequisites, and common prerequisites include CS1 and Algorithms ($N = 31$).
3) ML Courses, advanced level, offered once a year, high level of total and immediate prerequisite courses, and common prerequisites include CS1, Algorithms, Linear Algebra, Probability, Statistics, and Single Variable Calculus ($N = 23$).
4) DS courses at an introductory level ($N = 17$).
5) Intermediate level, offered more than once a year, low first exposure levels, low total and immediate numbers of prerequisites ($N = 15$).
6) Advanced level, offered more than once a year, high total number of prerequisites, and common prerequisites include CS2, Statistics, Foundations of Programming Languages, and Single Variable Calculus ($N = 7$).
7) ML courses, advanced level, high first exposure level, low number of immediate prerequisites, and common prerequisites include Linear Algebra and Machine Learning ($N = 6$).
8) DS courses, advanced level, offered once a year, and common prerequisites include CS1, CS2 Algorithms, and Data Management ($N = 5$).

D. University Level Clustering: RQ3

Table III summarizes the comparative statistics between public and private universities, highlighting differences in average course levels, number of prerequisites, and first exposure to *AI* courses. The table shows that public institutions offer more advanced courses with a greater number of prerequisites and greater exposure levels than private institutions on average. For the clustering, we looked at the institution level, and the clusters were based on the following features: public vs. private institutions, course levels, exposure levels, prerequisite numbers, and frequency. The four centroids informed our clustering analysis. Figure 6 is a heatmap of the university clusters with the same coloring scheme as the course clusters. The four clusters have the following characteristics:

1) Public institutions, advanced course levels, high first exposures ($N = 25$).
2) Public institutions frequently offer courses and a lower number of prerequisites ($N = 11$).
3) Private institutions, a high number of average prerequisites ($N = 10$).
4) Public and private institutions, introductory course levels, infrequent offerings, minimal total prerequisites, low first exposure ($N = 4$).

V. Discussion

We analyzed the impact of our findings on the varied prerequisite structures for *AI* courses across R-1 institutions. Our findings showed an apparent disparity in how DS is introduced

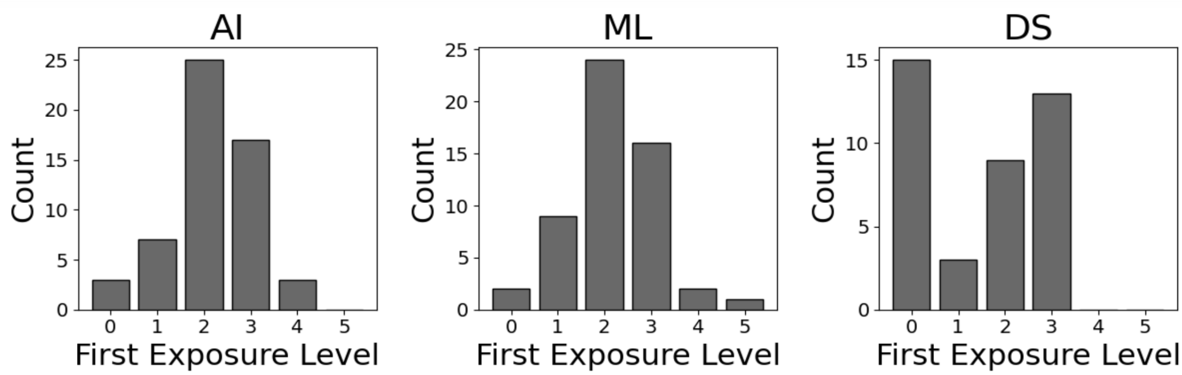

Fig. 3. Histogram of the exposure levels of *AI* courses.

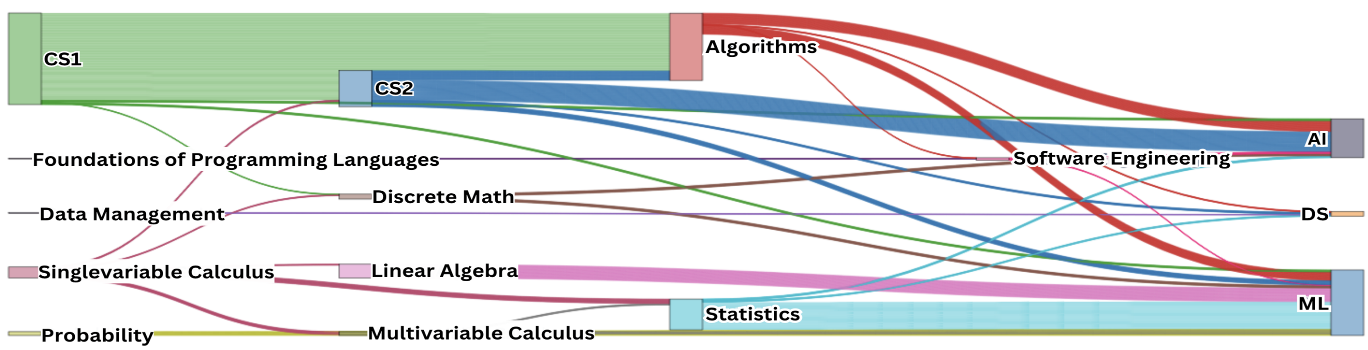

Fig. 4. Sankey diagram representing prerequisite chains of *AI* courses. The nodes represent prerequisite courses, and the lines connecting each node represent the relationships. For example, if one university's AI course has its Algorithms course as a prerequisite, which has CS1 as a prerequisite, two lines would be plotted connecting CS1 to Algorithms and Algorithms to AI (from left to right). The width of a connection between two nodes corresponds with the frequency of that particular prerequisite relationship. Connections that occur less than three times are not plotted for graph clarity. DS courses have relatively few prerequisites, as seen from their sparse and thick connections.

TABLE III
COMPARATIVE STATISTICS OF *AI* COURSES BY UNIVERSITY TYPE

University Type	Count	Average Course Level	Average Number of Prerequisites	Average First Exposure Level
Public	37	1.90	3.53	2.10
Private	13	1.66	3.27	1.90

as a fundamental, accessible subject from the first term, unlike the more advanced AI and ML courses that require solid foundational knowledge built over several semesters. This discrepancy highlights the potential for educational strategies that include early, integrated exposure to AI concepts within DS curricula, thereby fostering a smoother transition into more complex areas and enhancing research involvement from the onset of academic pursuits. The following analysis delves into the varied approaches institutions can adopt and proposes suggestions to ease student access to AI education.

A. *RQ1. What is the earliest exposure of students to the AI curriculum R1 computing departments in the U.S.?*

Nearly a third of universities offer DS courses with no prerequisites, allowing students to engage with DS from their first term. This approach contrasts with AI and ML courses, which typically require two to three semesters of prerequisites. The results suggest that DS is considered a fundamental subject in early university education, while AI and ML are positioned as advanced subjects, accessible after building a foundational knowledge base in the first year.

Thus, we believe that offering DS courses that introduce some AI content can be a helpful strategy that computing departments can use to develop undergraduate researchers capable of contributing to applied AI research. For example, some institutions can integrate introductory AI concepts into early DS courses, using class projects that apply DS tools to solve real-world problems with AI components. This approach could spark interest in AI fields early on, encouraging a more diverse range of students to pursue advanced AI studies. Institutions could also develop modules within DS courses that cover basic AI principles or ML algorithms, making these concepts accessible without needing advanced mathematical or programming prerequisites.

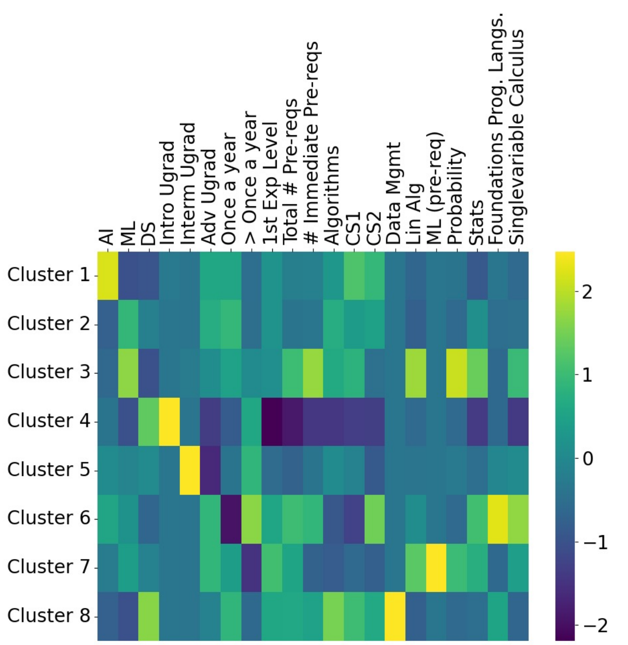

Fig. 5. Heatmap of course clusters. Each column represents a feature (course type, level, frequency, number of prerequisites, prerequisite courses).

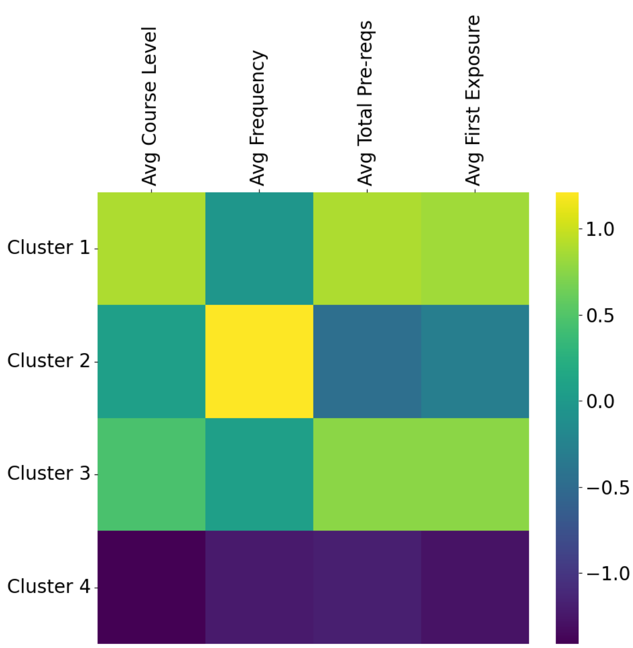

Fig. 6. Heatmap of university clusters.

B. *RQ2. What common approaches are institutions using to structure prerequisites for AI courses in R1 computing departments in the U.S.?*

AI courses typically require a strong programming background, often starting with CS1 courses as prerequisites. This indicates a focus on programming skills as essential for AI courses. ML courses, on the other hand, demand extensive preparation in both programming and mathematics. Their long prerequisite chains usually begin with CS1 and Single-variable Calculus, reflecting the need for a deeper understanding of both fields.

Compared to AI and ML, DS courses are more accessible, generally requiring fewer prerequisites, and thus, are positioned as more introductory subjects in the curriculum. However, Figure 4 shows that a keen student should be able to satisfy the prerequisites to take an ML course in the latter half of year 2 of study if the university's curriculum allows it. Therefore, scheduling courses so interested and capable students can take an ML course in the latter half of year 2 is another way that computing departments can support undergraduate research. An example of effective sequencing can be seen in institutions that align their calculus and programming courses in the first year, allowing students to take foundational ML courses earlier. This strategy can accelerate students' learning curve. Institutions might also consider restructuring their curriculums to reduce overlaps and redundancies in prerequisites for ML courses, perhaps by integrating applied mathematics directly within foundational programming courses.

C. *RQ3. What different approaches are institutions using to structure prerequisites for AI courses in R1 computing departments in the U.S.?*

Introductory AI courses typically require fewer prerequisites than ML courses, indicating easier access for beginners. Some ML courses require DS as a prerequisite, suggesting an institutional focus on integrating DS skills in ML education. Private universities offer these courses more frequently and with fewer prerequisites than public universities, facilitating quicker access to AI and ML studies. In contrast, public universities have higher-level introductory courses, potentially providing a more thorough foundational education in these fields but with delayed student exposure.

Learning from the differences between the two types of institutions, private universities can ensure foundational depth by offering optional preparatory courses alongside entry-level AI and ML courses to provide additional background knowledge, enabling a more inclusive yet thorough educational experience. On the other hand, public institutions can streamline prerequisites for introductory AI and ML courses where possible, allowing more students quicker access while maintaining the rigor of higher-level courses for those seeking a deeper specialization. For instance, a public university might offer an "AI Fundamentals" course with minimal prerequisites to attract a broader array of students, including those from non-technical backgrounds. In contrast, private institutions might provide a comprehensive series of courses relevant to a deeper understanding of AI for students seeking to gain more thorough knowledge. Institutions could enhance the efficiency of AI education by implementing courses designed to rapidly cover necessary prerequisites, thereby accelerating students' readiness for advanced AI topics without compromising the educational depth.

The findings suggest various approaches to *AI* curricula. Many institutions offer introductory DS courses, showing that such courses provide students with early exposure to AI-related content. Such exposure is crucial because it helps demystify complex concepts and makes students feel more comfortable engaging with more advanced AI and ML topics later on. Structuring AI and ML courses to gradually build on the foundational knowledge provided by early DS courses can be another strategy to structure a progressive learning journey for students in these fields. This gradual progression allows for an incremental increase in difficulty and depth, ensuring that students develop the requisite skills and confidence to tackle advanced concepts. Such an approach can help foster a stronger interest in research and practical applications of AI and ML, as students would have a solid base of understanding before moving into specialized areas. It may also mitigate potential disparities in student preparedness and help instructors tailor their teaching to different learning levels.

VI. Limitations

Our data relies on publicly accessible course calendar information on the institutions' websites. However, this data may be incomplete or contain outdated information. The dynamic nature of course offerings in rapidly evolving fields such as *AI* means that some relevant courses might have been overlooked in their latest iteration. Prerequisite links were also collected from the publicly accessible academic calendar and may not capture prerequisite and course requirements that are not explicitly listed. As for data collection, there is an uneven geographical distribution of the sampled universities, as depicted in Figure 1. The concentration of institutions within certain regions may influence the generalizability of our findings.

Moreover, we only considered courses offered by the computing department, which may exclude courses from other departments that also contribute to education in AI. These departments may include mathematics, statistics, engineering, or even the social sciences, where interdisciplinary *AI* courses are offered. Failing to account for such courses means that the data may not fully capture the breadth of AI education provided across institutions. Furthermore, excluding these interdisciplinary courses could overlook innovative curricular approaches or collaborative research opportunities that bridge traditional departmental boundaries. Thus, future studies could benefit from a more comprehensive evaluation of *AI* courses across multiple disciplines.

VII. Future Work

In future work, we plan to explore the impact of initial course entry points on students' subsequent engagement in AI research. Given R1 institutions' heavy emphasis on research, this investigation will involve a longitudinal study that tracks students' exposure to *AI* courses to their active participation in AI research. We will analyze the timing of research engagement relative to coursework and retention rates within the AI field. We are interested in determining how early exposure to *AI* courses influences students' involvement and success in research activities, thereby providing insights into optimizing educational pathways in AI.

Further investigation is also suggested into whether certain prerequisite structures act as barriers to entry, This exploration could involve examining the impact of different prerequisite configurations on student accessibility and success.

In addition, incorporating qualitative data supplement to our current analysis through collaboration with professors and industry professionals is another direction for future research to make *AI* accessible at an earlier exposure. This collaboration would involve a qualitative analysis, such as interviews, to determine the most effective course structures for achieving various goals, such as conducting research in specific areas. Such insights can provide a more nuanced understanding of how educational frameworks can be tailored to meet diverse career and research objectives [33], [34].

VIII. Conclusion

Our study showed variations in *AI* education in R1 institutions in the U.S. DS courses are generally more accessible, often with little to no prerequisites, allowing early student engagement. In contrast, AI and ML courses typically require extensive prerequisites. Moreover, AI courses emphasize programming skills, starting with CS1, while ML courses demand comprehensive preparation in both programming and mathematics. Private universities offer *AI* courses with fewer prerequisites than public universities, facilitating quicker access. In contrast, public universities have higher average course levels with more prerequisites, potentially providing a more thorough foundational education but with delayed student first exposure.

Our analysis emphasizes the need for educational strategies that balance foundational knowledge with accessible entry points into *AI* courses. Institutions should consider modular prerequisite structures that let students gradually develop necessary skills while providing early opportunities to learn about AI and ML applications. Flexible curricula could help students from diverse backgrounds engage in these fields. As AI technologies advance and integrate into various sectors, educational frameworks must be flexible and accessible. Early access to *AI* courses can significantly impact students' ability to participate in research early in their academic careers.

Our study lays the groundwork for future research and curriculum development, focusing on adjusting prerequisite structures to better align with diverse student needs. By addressing the barriers identified, institutions can create flexible curricula that encourage early research engagement in AI.

References

[1] S. Reddy, S. Allan, S. Coghlan, and P. Cooper, "A governance model for the application of ai in health care," *Journal of the American Medical Informatics Association*, vol. 27, no. 3, pp. 491–497, 2020.

[2] L. Cao, “Ai in finance: challenges, techniques, and opportunities,” *ACM Computing Surveys (CSUR)*, vol. 55, no. 3, pp. 1–38, 2022.

[3] K. Rajan and A. Saffiotti, “Towards a science of integrated ai and robotics,” pp. 1–9, 2017.

[4] E. Horvitz, “Ai, people, and society,” 2017.

[5] A. Urtasun Alonso, “Empowering undergraduates through machine learning,” *Industry and Higher Education, 36 (3), 443-447*, 2022.

[6] P. Langley, “An integrative framework for artificial intelligence education,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 9670–9677.

[7] I. M. Cockburn, R. Henderson, and S. Stern, “The impact of artificial intelligence on innovation: An exploratory analysis,” in *The economics of artificial intelligence: An agenda*. University of Chicago Press, 2018, pp. 115–146.

[8] D. Barretto, J. LaChance, E. Burton, and S. N. Liao, “Exploring why underrepresented students are less likely to study machine learning and artificial intelligence,” in *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1*, 2021, pp. 457–463.

[9] S. Wongvibulsin, “Educational strategies to foster diversity and inclusion in machine intelligence,” *Nature machine intelligence*, vol. 1, no. 2, pp. 70–71, 2019.

[10] O. Hazzan and K. Mike, “The pedagogical challenge of machine learning education,” in *Guide to Teaching Data Science: An Interdisciplinary Approach*. Springer, 2023, pp. 199–208.

[11] B. Allen, A. S. McGough, and M. Devlin, “Toward a framework for teaching artificial intelligence to a higher education audience,” *ACM Transactions on Computing Education (TOCE)*, vol. 22, no. 2, pp. 1–29, 2021.

[12] A. A. de Freitas and T. B. Weingart, “I’m going to learn what?!? teaching artificial intelligence to freshmen in an introductory computer science course,” in *Proceedings of the 52nd ACM technical symposium on computer science education*, 2021, pp. 198–204.

[13] L. Zhang and S. Allin, “Just-in-time prerequisite review for a machine learning course,” in *Proceedings of the 25th Western Canadian Conference on Computing Education*, 2023, pp. 1–2.

[14] M. Kunda, “The ai triplet: Computational, conceptual, and mathematical knowledge in ai education,” *arXiv preprint arXiv:2110.09290*, 2021.

[15] M. Neumann, “Ai education matters: a first introduction to modeling and learning using the data science workflow,” *AI Matters*, vol. 5, no. 3, pp. 21–24, 2019.

[16] J. S. Hardin and N. J. Horton, “Ensuring that mathematics is relevant in a world of data science,” *Notices of the AMS*, vol. 64, no. 9, pp. 986–990, 2017.

[17] D. T. K. Ng, M. Lee, R. J. Y. Tan, X. Hu, J. S. Downie, and S. K. W. Chu, “A review of ai teaching and learning from 2000 to 2020,” *Education and Information Technologies*, vol. 28, no. 7, pp. 8445–8501, 2023.

[18] T. Menkhoff and Y. Q. Lydia Teo, “Engaging undergraduate students in an introductory ai course through a knowledge-based chatbot workshop,” in *Proceedings of the 6th International Conference on Information System and Data Mining*, 2022, pp. 119–125.

[19] S. Krause-Levy, A. Salguero, R. S. Lim, H. McTavish, J. Trajkovic, L. Porter, and W. G. Griswold, “Instructor perspectives on prerequisite courses in computing,” in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, 2023, pp. 277–283.

[20] S. Krause-Levy, S. Valstar, L. Porter, and W. G. Griswold, “A demographic analysis on prerequisite preparation in an advanced data structures course,” in *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education-Volume 1*, 2022, pp. 661–667.

[21] ——, “Exploring the link between prerequisites and performance in advanced data structures,” in *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 2020, pp. 386–392.

[22] M. Li and B. Liu, “A brief discussion on the reform of mathematics teaching in artificial intelligence majors-taking matrix computation and optimization as examples,” in *National Conference of Theoretical Computer Science*. Springer, 2022, pp. 132–141.

[23] N. Norouzi, S. Chaturvedi, and M. Rutledge, “Lessons learned from teaching machine learning and natural language processing to high school students,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 09, 2020, pp. 13 397–13 403.

[24] H. Crompton and D. Burke, “Artificial intelligence in higher education: the state of the field,” *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, p. 22, 2023.

[25] P. Bhattacharyya and C. W. Chan, “Can undergraduate research participation reduce the equity gap?” *Journal of the Scholarship of Teaching and Learning*, vol. 21, no. 1, 2021.

[26] C. Alvarado, S. Villazon, and B. Tamer, “Evaluating a scalable program for undergraduate cs research,” in *Proceedings of the 2019 ACM Conference on International Computing Education Research*, 2019, pp. 269–277.

[27] J. S. Stanford, S. E. Rocheleau, K. P. Smith, and J. Mohan, “Early undergraduate research experiences lead to similar learning gains for stem and non-stem undergraduates,” *Studies in Higher Education*, vol. 42, no. 1, pp. 115–129, 2017.

[28] M. C. Linn, E. Palmer, A. Baranger, E. Gerard, and E. Stone, “Undergraduate research experiences: Impacts and opportunities,” *Science*, vol. 347, no. 6222, p. 1261757, 2015.

[29] N. Norouzi, H. Habibi, C. Robinson, and A. Sher, “An equity-minded multi-dimensional framework for exploring the dynamics of sense of belonging in an introductory cs course,” in *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, 2023, pp. 131–137.

[30] B. Akram, J. Leinonen, N. Norouzi, J. Prather, and L. Zhang, “Ai in computing education from research to practice,” in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2*, 2024, pp. 1521–1522.

[31] N. Norouzi and R. Hausen, “Quantitative evaluation of student engagement in a large-scale introduction to programming course using a cloud-based automatic grading system,” in *2018 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2018, pp. 1–5.

[32] M. L. McHugh, “Interrater reliability: the kappa statistic,” *Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[33] S. Kross and P. J. Guo, “Practitioners teaching data science in industry and academia: Expectations, workflows, and challenges,” in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–14.

[34] E. Sulmont, E. Patitsas, and J. R. Cooperstock, “What is hard about teaching machine learning to non-majors? insights from classifying instructors’ learning goals,” *ACM Transactions on Computing Education (TOCE)*, vol. 19, no. 4, pp. 1–16, 2019.