2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
Paper ID #53889

Teaching the Intuition Behind Artificial Intelligence: Understanding Before Automation

Dr. Tracy Anne Hammond, Texas A&M University

Dr. Tracy Anne Hammond is a Professor of Computer Science and Engineering at Texas A&M University and Director of the Sketch Recognition Lab. She earned her Ph.D. in Electrical Engineering and Computer Science from the MIT Computer Science and Artificial Intelligence Laboratory at Massachusetts Institute of Technology, along with additional degrees from Columbia University in computer science, mathematics, applied physics, and anthropology.

Her research focuses on artificial intelligence, sketch recognition, intelligent user interfaces, haptics, and engineering education, with applications in education, health, accessibility, and related domains. She contributes to research, teaching, and professional service in AI and engineering education, with an emphasis on responsible and inclusive design.

Jobin Varughese, Texas A&M University

Jobin Varughese is a doctoral student in the College of Engineering at Texas A&M University, where he also serves as Senior Educational Data Analyst at the Center for Teaching Excellence. He holds an M.S. in Management Information Systems with an emphasis in Data Science from Oklahoma State University and a B.S. in Electronics and Communication Engineering from Christ University. His research draws on learning analytics, AI in education, and human-computer interaction to observe, build, and study systems that learn how people learn. In this work, he partners closely with faculty, instructional staff, and students — believing that the most meaningful scholarship emerges not from research alone, but from collaborations with the people closest to the learning experience.

Teaching the Intuition Behind Artificial Intelligence: Understanding Before Automation

Abstract

Before generative AI reshaped the educational landscape, a course was designed around a simple premise: understand AI before you use it. This paper examines that intuition-first approach through *Foundations of Data Science and Machine Learning*, originally developed and taught in 2020 and offered again in 2025 using the same conceptual curriculum. Students engaged with foundational AI concepts by hand, without code or libraries. Using a concept-focused written assessment and final project reports from the 2025 offering, we analyze how students reason about model assumptions, evaluation tradeoffs, and failure modes. Results show consistent evidence of conceptual reasoning, including the ability to articulate why models fail and what probabilistic systems cannot know. These forms of reasoning align with the critical thinking that generative AI literacy requires, despite generative AI systems never being explicitly taught. This case also illustrates how institutional constraints on AI coursework can narrow interdisciplinary access while conceptual depth remains intact. The findings suggest that intuition-first instruction provides a durable foundation for AI literacy that remains relevant as tools evolve.

Introduction

Artificial intelligence and machine learning have become central topics across engineering and non-engineering disciplines alike, prompting rapid growth in AI-related coursework. In many cases, these offerings emphasize coding proficiency, software frameworks, or the use of specific tools [1–3]. While such approaches support technical fluency, they can leave students without the conceptual understanding needed to explain model behavior, evaluate assumptions, or reason about failure. This limitation has become increasingly visible as generative AI systems such as ChatGPT are integrated into educational and professional contexts. These tools produce fluent and convincing output, which can obscure underlying limitations related to training data, probabilistic inference, and generalization [4, 5]. Students who lack foundational intuition may interpret such systems as authoritative or intelligent, rather than as statistical models constrained by the data and assumptions that produced them. This renews a longstanding question in engineering education: what does it mean to be AI-literate, and how should that literacy be taught?

This paper argues that intuition-first instruction offers a durable foundation for AI literacy. Rather than beginning with code or tools, intuition-first instruction emphasizes reasoning about data, probability, model structure, and failure modes. Students learn to explain why a model works before learning how to implement it. This approach is particularly important for interdisciplinary audiences, including students who may never become machine learning developers but who will increasingly interact with AI systems in professional contexts. To explore this approach, we examine a vertically integrated course titled *Foundations of Data Science and Machine Learning*, designed for both undergraduate and graduate students from engineering and non-engineering

backgrounds. The course was taught twice, first in 2020 and again in 2025, using the same conceptual curriculum. Students solved core machine learning problems by hand—including Bayesian inference, confusion matrix analysis, and Hidden Markov Models—with the goal of developing intuition about uncertainty, generalization, and model limitations [3, 6].

The two offerings differed markedly in institutional context. The earlier course enrolled a large and interdisciplinary cohort under open enrollment conditions, while the later offering was significantly constrained by departmental policies introduced following the rise in popularity of AI coursework. This contrast provides insight not only into pedagogical outcomes, but also into how institutional responses to AI shape access to meaningful learning experiences. Using a concept-focused written assessment and student-generated artifacts, this paper examines how intuition-first instruction supports reasoning about machine learning concepts and generative model behavior without relying on code or tool use. In doing so, it contributes to ongoing discussions in engineering education about AI literacy, interdisciplinary access, and the risks of reducing AI education to tool-centric instruction [7–9]. Prior work has emphasized the importance of interdisciplinary approaches to AI education to ensure that conceptual understanding is not restricted to a narrow subset of students [10, 11]; this study contributes empirical evidence to that discussion by documenting how institutional constraints shaped access while preserving depth of engagement.

Related Work

Teaching Artificial Intelligence and Machine Learning

Instruction in artificial intelligence and machine learning has traditionally emphasized algorithm implementation, coding proficiency, and performance optimization. Many introductory and intermediate courses focus on training models using established libraries and frameworks, often assuming substantial prior programming experience [1–3]. While this approach supports technical fluency, it can obscure the probabilistic assumptions and conceptual foundations that govern model behavior, particularly for students outside of computer science. Prior work has documented how implementation-focused instruction can mask conceptual misunderstandings, with students often learning surface procedures without developing deep mental models of underlying processes [12–14]. Recent calls within engineering education have emphasized the importance of broader AI literacy that extends beyond technical skills to include conceptual understanding, interpretability, and critical evaluation of model behavior [7, 8, 15]. This work aligns with those calls by positioning intuition-first instruction as a means of expanding access to AI concepts across disciplines and levels of preparation. Assessing this form of understanding requires tasks that surface reasoning rather than implementation; explanation-based assessments and written reasoning tasks have been proposed as more appropriate measures than programming assignments or benchmark performance alone [16, 17].

Conceptual and Intuition-First Pedagogies

Research in learning sciences consistently demonstrates that conceptual understanding improves transfer, retention, and problem-solving flexibility compared to procedural or tool-driven instruction [18, 19]. In statistics and data science education, scholars have argued that reasoning

about uncertainty, data-generating processes, and model assumptions is more important for long-term competence than computational mechanics alone [20, 21]. Within machine learning education, several authors have advocated for teaching models through probabilistic reasoning and analogy-based explanations to reduce cognitive barriers for learners without extensive mathematical backgrounds [22, 23]. Recent work in data science education similarly emphasizes framing modeling as an interpretive activity shaped by data quality, assumptions, and context [24, 25].

Understanding and Explaining Generative AI Failures

The widespread adoption of generative AI systems such as ChatGPT has intensified interest in helping students understand why these models fail, hallucinate, or behave unpredictably. Prior work has documented how probabilistic sequence models are fundamentally constrained by their training data and assumptions [3, 6]. More recent critiques of large language models emphasize that fluent output can obscure these limitations, leading users to overestimate model reliability [4, 5]. Educational responses to generative AI have largely focused on tool use, prompt engineering, and academic integrity [9], with comparatively less attention to foundational conceptual understanding. This study contributes to emerging discussions by showing how intuition-first instruction grounded in probabilistic reasoning can help students interpret and critique generative model behavior without requiring direct instruction on specific tools.

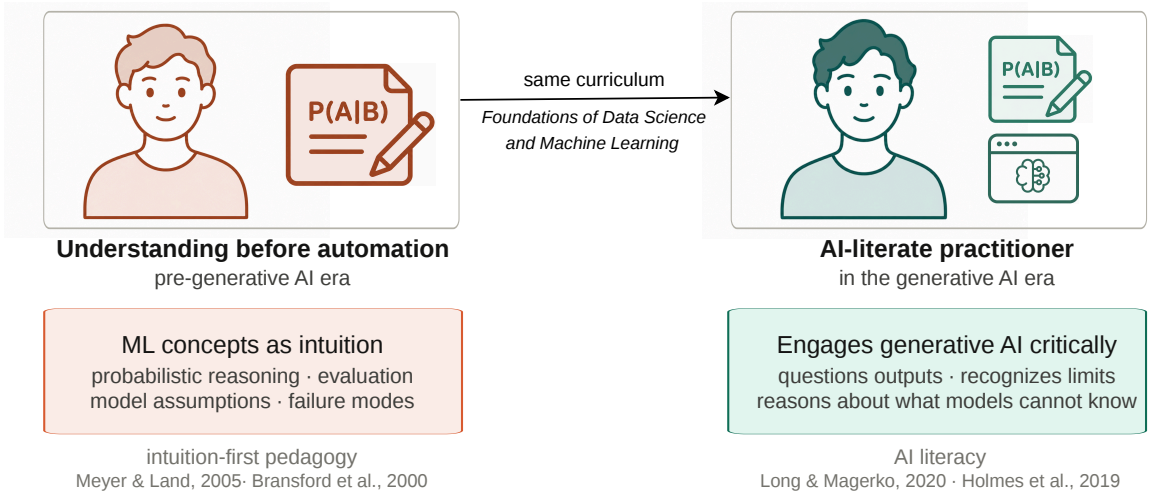

FIGURE 1. Two states of the same learner. A student equipped with machine learning concepts as intuition before the generative AI moment develops the foundational reasoning that enables critical engagement with generative AI systems they were never explicitly taught. The curriculum predates ChatGPT. The critical reasoning developed does not expire when the tools change. Theoretical grounding: intuition-first pedagogy [18, 26]; AI literacy [7, 8].

Theoretical Framework

This study is grounded in two complementary frameworks that together explain both the pedagogical design of the course and the reasoning outcomes observed in student work.

Threshold Concepts

Meyer and Land [26] introduced the notion of threshold concepts to describe a particular class of ideas that, once understood, fundamentally transform how a learner sees a subject. Threshold concepts are *transformative*—they change the learner's perspective in ways that extend beyond the concept itself—and *irreversible*—once grasped, they are difficult to unlearn. They are also *integrative*, in that they reveal connections across ideas that were previously obscure, and often *troublesome*, in that they resist straightforward transmission and require genuine engagement to cross.

The course examined in this study is organized around threshold concepts in machine learning. Bayesian inference is one such concept [26]: a student who genuinely understands that probability is a measure of belief updated by evidence—not a fixed property of the world—sees all probabilistic models differently. The distinction between accuracy and recall in imbalanced classification is another [26]: once a student understands that a classifier can be simultaneously correct 99% of the time and completely useless for its intended purpose, no metric can be read uncritically again. The structure of a Hidden Markov Model [6]—that a system has hidden states it cannot directly observe, and can only emit outputs consistent with what it has seen—is perhaps the most transferable threshold concept in the curriculum [26]. A student who has drawn and traced an HMM by hand has internalized something durable about what any probabilistic sequence model can and cannot do.

The course does not teach these as a checklist. It teaches them as a disposition: that understanding a model means understanding its assumptions, its failure modes, and the boundaries of what it can know. That disposition, once developed, does not expire when the tools change.

AI Literacy

Long and Magerko [8] define AI literacy as a set of competencies that allow people to critically evaluate AI technologies, communicate and collaborate effectively with AI, and use AI as a tool while understanding its societal implications. This is distinct from AI fluency—the ability to use AI tools effectively—in that it requires conceptual understanding of how AI systems work, not just operational familiarity with their interfaces.

The distinction matters for this paper. A student who has been taught to prompt a language model effectively is AI-fluent. A student who understands why a model produces confident but incorrect outputs—because it is a probabilistic system constrained by its training data, optimizing for plausibility rather than truth—is AI-literate. Intuition-first instruction in machine learning, as examined here, is designed to produce the second kind of student. The threshold concepts described above are precisely the conceptual tools that AI literacy requires.

Together, these frameworks explain the figure at the core of this paper (Figure 1). The left state represents a learner building threshold concepts through direct engagement with machine learning ideas, before the widespread adoption of generative AI. The right state represents the same learner entering a world of generative AI with those concepts intact—equipped not just to use these systems, but to reason about them critically. The arrow between the two states is not a claim about causal transfer. It is a claim about what conceptual preparation makes possible.

Data Sources and Evidence

The primary sources of evidence for this study are a concept-focused written assessment and a set of final project reports, both drawn from the 2025 offering of *Foundations of Data Science and Machine Learning* (CSCE 689) at a large public research university. The course was a compressed 4.5-week summer offering enrolling eleven graduate students ($n = 11$) from engineering and non-engineering backgrounds, including computer science, electrical engineering, and interdisciplinary programs. Enrollment was restricted by departmental policies introduced following the rise in popularity of AI coursework—institutional constraints that narrowed the interdisciplinary access the course had previously attracted while preserving the depth of conceptual engagement.

This course was originally developed and taught in 2020, before the widespread adoption of generative AI. The same conceptual curriculum was used again in 2025. This paper focuses exclusively on the 2025 offering, where student artifacts were available for analysis.

Assessment Design

To assess conceptual understanding independent of programming ability or software tooling, students completed the concept-focused assessment, a one-hour, ten-question written assessment. The full instrument is reproduced in Appendix A. Questions required students to reason through foundational machine learning problems by hand, without access to code, libraries, or computational tools. Tasks included explaining differences between baseline and ensemble classifiers, applying Bayes' Theorem step by step, constructing confusion matrices from narrative descriptions, computing and interpreting evaluation metrics, reasoning about overfitting from training and test performance, and constructing a Hidden Markov Model including hidden states, emissions, and transition probabilities [6].

The assessment was designed so that correct responses required genuine conceptual understanding rather than procedural recall. Seven students submitted typed responses; four submitted handwritten work. Both formats were treated as equivalent evidence of conceptual reasoning, as the assessment design foregrounded reasoning process over computational execution.

Evidence of Intuition-First Learning

The assessment instrument provides direct evidence of intuition-first learning because it removes all implementation scaffolding. There is no code to run, no library to call, and no tool to consult. What remains is the student's own understanding. A correct response to the Hidden Markov Model construction task, for example, requires a student to hold the structure of a probabilistic state model in mind, assign meaningful probabilities, and reason through what the model can and cannot infer from an observed sequence [6]. That is not a task that can be completed by recalling syntax. It requires internalized conceptual understanding—which is precisely what intuition-first instruction is designed to produce.

Use of Evidence in This Study

This study uses the assessment submissions and final project reports as qualitative evidence of conceptual engagement, not as measures of learning gain or comparative performance. Direct student quotes are used where student language precisely captures the reasoning pattern being described. All quotes are anonymized by participant code and question number. This study was determined to be exempt from full IRB review. Student artifacts are used with instructor authorization.

Methods and Limitations

Methods

This study adopts a qualitative, artifact-based case study methodology. Rather than measuring learning gains through pre/post-testing or statistical comparison, the analysis focuses on student-generated work that directly reflects reasoning processes. The primary artifact analyzed is the concept-focused assessment, a one-hour, ten-question written assessment designed to evaluate conceptual understanding of classification, evaluation metrics, Bayesian reasoning, and Hidden Markov Models. Final project reports from the same cohort serve as a secondary source to triangulate patterns observed in the assessment data.

Student submissions were reviewed using inductive thematic analysis. Responses were examined for patterns in how students expressed reasoning about model selection, probabilistic inference, evaluation tradeoffs, generalization, and model limitations. Themes were identified based on recurring patterns in student language across all eleven submissions rather than frequency counts or correctness scores. Coding was conducted independently by two reviewers. Themes were confirmed through discussion, with agreement reached on five of seven themes in the first pass and consensus established on all themes through discussion.

Limitations

Several limitations should be considered when interpreting these findings.

First, the sample is small ($n = 11$) and drawn from a single course offering at one institution. The findings are illustrative rather than generalizable and should be interpreted as a detailed instructional case rather than a population-level study. This small enrollment was itself a consequence of institutional constraints that restricted access to the course following the rise in popularity of AI coursework—a limitation that is also a finding, in that it documents how departmental policy can narrow access to foundational AI education.

Second, the study does not include a comparison condition. No claim is made that intuition-first instruction outperforms alternative approaches. The analysis documents what students were able to demonstrate, not how their outcomes compare to students taught differently.

Third, the instructor also served as the researcher. To mitigate interpretive bias, the analysis focuses on concrete student language and problem-solving steps rather than evaluative judgments, and claims are bounded to what is directly observable in student work. A second independent coder reviewed all submissions and themes were confirmed through discussion.

Fourth, this study does not include an instrument that directly measures transfer of ML reasoning to generative AI evaluation. The connection between the threshold concepts developed in the course and the critical reasoning capabilities described in the discussion is argued structurally, not demonstrated empirically. That empirical demonstration remains future work.

Results

This section reports findings from qualitative analysis of student responses to assessment and final project reports from the 2025 course offering. Seven themes were identified through inductive thematic analysis and confirmed through independent coding by two reviewers. Table I provides a summary of themes and representative student language.

TABLE I. Reasoning patterns observed across student responses, with representative anonymized quotes.

Reasoning pattern	**What it looks like**	**Student language**
Reasoning from a floor	Students ask whether a model is doing better than chance before accepting its performance as meaningful.	*"If the model can't beat ZeroR, something is probably wrong."*—P3, Q1
Treating inference as a process	Students work through each step of probabilistic reasoning, naming what each term represents before computing.	*"P(cube | red) × P(red) / P(cube)"* with every prior defined before substitution.—P5, Q2
Questioning what a metric is actually measuring	Students explain why a high score can still mean total failure, and choose metrics based on what the problem needs.	*"High accuracy does not reflect the model's effectiveness in detecting the rare but important class."* —P6, Q8
Distinguishing fitting from learning	Students identify when a model has memorized rather than generalized, and explain what that means for new data.	*"The model achieved 100% accuracy by memorizing exact values, but fails to generalize to new examples."* —P5, Q10
Reasoning about what a model cannot know	Students identify the boundaries of what a model can infer, including when a problem lacks the information needed to reach a conclusion.	*"The transition probabilities are not given — it is not possible to figure out a most likely sequence of states."*—P8, Q7
Treating simplicity as a deliberate choice	Students explain when a simpler model is the better option, rather than assuming complexity is preferable.	*"Decision tree tries to overfit in that case and can give results that are not expected."*—P11, Q6
Reading failure as information	Students use model failure as a starting point for reasoning about data, assumptions, and design.	*"It is useless since it will never actually classify the task we are interested in."*—P8, Q8

Conceptual Reasoning in Written Assessments

Analysis of student responses revealed consistent evidence of reasoning about core machine learning concepts across multiple domains. Students demonstrated the ability to explain differences between baseline and learning models, articulate probabilistic reasoning using Bayes' Theorem, construct confusion matrices from narrative classification scenarios, and compute and interpret evaluation metrics.

The most detailed reasoning emerged in the imbalanced classification question. All eleven students computed 99.72% accuracy and immediately identified it as misleading. Explanations went beyond naming an alternative metric to articulating the mechanism. P6 wrote: *"high accuracy does not reflect the model's effectiveness in detecting the rare but important class."* P1 computed Recall and Precision explicitly, finding both equal to zero, demonstrating that the classifier has no utility for its stated purpose despite near-perfect accuracy.

Reasoning About Generalization and Overfitting

Student responses to the training and testing question demonstrated consistent understanding of the distinction between fitting training data and learning generalizable structure. All eleven students correctly identified that the classifier achieved 100% training accuracy by memorizing specific values rather than learning an underlying pattern, and correctly predicted degraded performance on unseen data. P5 wrote: *"the model achieved 100% accuracy on the training data by memorizing exact values, but it fails to generalize to new examples. A good model should capture underlying patterns, not just memorize training inputs."* This reasoning indicates that students could connect abstract concepts of overfitting and generalization to concrete examples without relying on implementation.

Probabilistic State Modeling

Responses to the Hidden Markov Model construction task showed that students could distinguish between hidden states and observable emissions and reason about transitions and priors in a probabilistic system [6]. Most students correctly drew the full HMM structure with labeled probabilities. Two responses demonstrated deeper engagement with the structural constraints of the model. P4 extended the task beyond what was required, reasoning through the most probable state sequence given the observed emissions and noting: *"Predicted location for first three days: Lab $\rightarrow$ Field $\rightarrow$ Lab."* P8 identified an underspecification in the problem itself: *"the transition probabilities are not given, so it is not possible to really figure out a most likely sequence of states."* This response is notable because it demonstrates reasoning about what a model requires to make inferences—not just what it produces when fully specified, but what it cannot determine when information is missing.

Patterns in Project Reasoning

Final project artifacts from the 2025 course offering further supported patterns observed in the written assessment. Across projects spanning infrastructure systems modeling, environmental data analysis, health-related classification, and text classification, students demonstrated intentional model selection grounded in data characteristics and task constraints. Reports consistently included discussion of data quality, representational limits, and evaluation tradeoffs.

Rather than emphasizing optimized performance, project reports focused on interpretation of results and explanation of failure modes. When models performed poorly or inconsistently, students linked outcomes to data limitations, imbalance, or violations of assumptions. One project applied Hidden Markov Model reasoning to real infrastructure maintenance data, explicitly noting that the model could not predict deterioration patterns outside its training history—a direct application of the structural reasoning demonstrated in the assessment.

Discussion: Intuition-First Learning in a Post-ChatGPT World

The reasoning patterns documented in this study extend beyond quiz performance. Students who worked through machine learning concepts by hand, without tools or libraries, developed ways of thinking about AI systems that do not expire when the tools change. This suggests that durability may be a key strength of intuition-first instruction.

Reasoning Transfers When the Foundation Is Conceptual

The threshold concepts students internalized in this course are the same ones needed to engage critically with generative AI [26]. Once a student understands that a majority classifier can achieve 99% accuracy while detecting nothing, that insight restructures how they evaluate any model's performance claims. This is consistent with learning science evidence that conceptual understanding improves transfer and retention more reliably than procedural instruction [18, 19], and with calls in computing education for assessment designs that surface reasoning rather than implementation [16].

A student equipped to ask "what is this metric actually measuring?" is better positioned to evaluate generative AI output than one who has only learned to prompt effectively. This is not a claim of direct causal transfer. The study does not include an instrument measuring whether students applied these concepts to evaluate a generative AI output. The claim is structural: the conceptual tools this course builds are precisely those that AI literacy requires [8]. Intuition-first instruction does not produce students who know more about ChatGPT. It produces students who know how to reason about any probabilistic system, including ones that did not exist when they were taught.

The HMM Insight Is Particularly Transferable

Among the threshold concepts in this curriculum, Hidden Markov Model reasoning deserves particular attention. Students who drew and traced HMMs by hand internalized a specific and durable insight: a probabilistic sequence model can only generate or infer states consistent with its training data. It has no access to what it has not seen.

This is structurally analogous to how large language models fail. Hallucination is not random noise. It is the predictable consequence of a model generating plausible continuations outside its learned distribution. Large language models differ architecturally from HMMs in important ways, and this analogy does not substitute for understanding attention mechanisms or token prediction [4, 5]. But the probabilistic intuition transfers at the level that matters most for critical evaluation: the boundary between what a model can know and what it cannot.

One student's response captures this precisely. Noting that transition probabilities were not

provided, the student wrote that it was *"not possible to really figure out a most likely sequence of states."* That is a student reasoning about the informational requirements of a model, which is exactly the disposition that AI literacy demands [7, 8].

Institutional Constraints Narrowed Access While Depth Held

The 2025 offering enrolled eleven graduate students under departmental restrictions introduced following the rapid rise of AI coursework. The earlier offering attracted a larger, more interdisciplinary cohort under open enrollment. That contrast reflects a pattern playing out across institutions: as generative AI becomes ubiquitous, curricular responses have often prioritized tool-based instruction and prompt engineering over foundational conceptual preparation [9, 10].

The evidence from this study provides a counterpoint to that trend. Conceptual depth was preserved under constrained conditions. The institutional pressure that narrowed access to this course also increases the need for this type of course. There is a real risk that engineering education responds to the generative AI moment by producing AI-fluent graduates who are not AI-literate [8, 11].

Implications for Curriculum Design

Three implications follow for educators designing AI-related coursework. Conceptual instruction should precede tool use. Students who understand why a model works are better equipped to evaluate what a tool produces. Assessment design shapes what gets learned: written, tool-free tasks that require students to explain why a model succeeds or fails make conceptual understanding visible and reinforce it. And access matters. Restricting foundational AI coursework to narrow departmental boundaries limits who develops the capacity to reason critically about these systems [7, 11].

Conclusion

This paper presented an intuition-first approach to teaching machine learning that emphasizes conceptual understanding over programming or tool use. Through a course originally developed before the widespread adoption of generative AI and taught again in 2025 using the same curriculum, students demonstrated the ability to reason about probabilistic models, evaluation metrics, and generalization using foundational principles rather than implementation details. Analysis of a concept-focused written assessment showed that students could articulate why models succeed or fail, including recognizing limitations related to data imbalance, overfitting, and the structural boundaries of what a probabilistic model can know. These reasoning patterns were consistent across all eleven submissions and were further supported by final project artifacts from the same cohort.

The findings suggest that intuition-first instruction produces understanding that remains relevant as AI technologies evolve [7, 8]. The threshold concepts students internalized in this course are the same ones that AI literacy requires [26]. A student who has reasoned through why a Hidden Markov Model cannot infer states absent from its training is equipped to ask the right questions about any probabilistic system, including ones that did not exist when they were taught.

The institutional contrast documented here carries its own lesson. Departmental restrictions

introduced following the rise of generative AI narrowed enrollment and reduced interdisciplinary access while the depth of conceptual engagement remained intact. Teaching students to use AI systems they do not understand is not an adequate response to this moment. By prioritizing understanding before automation, educators can prepare students not only to use AI, but to evaluate, critique, and adapt to it responsibly as the field continues to change [7–9].

References

[1] E. Alpaydin, *Introduction to Machine Learning*, 4th ed. Cambridge, MA: MIT Press, 2020.

[2] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. Sebastopol, CA: O'Reilly Media, 2019.

[3] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press, 2012.

[4] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?" in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. New York, NY: Association for Computing Machinery, 2021, pp. 610–623, doi: 10.1145/3442188.3445922.

[5] G. Marcus and E. Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York, NY: Pantheon Books, 2019.

[6] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989, doi: 10.1109/5.18626.

[7] W. Holmes, M. Bialik, and C. Fadel, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Paris, France: UNESCO and Center for Curriculum Redesign, 2019.

[8] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. New York, NY: Association for Computing Machinery, 2020, pp. 1–16, doi: 10.1145/3313831.3376727.

[9] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer *et al.*, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023, doi: 10.1016/j.lindif.2023.102274.

[10] R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, *Intelligence Unleashed: An Argument for AI in Education*. London, UK: Pearson, 2016.

[11] National Academies of Sciences, Engineering, and Medicine, *Data Science for Undergraduates: Opportunities and Options*. Washington, DC: National Academies Press, 2018, doi: 10.17226/25104.

[12] S. Fincher and M. Petre, *Computer Science Education Research*. London, UK: RoutledgeFalmer, 2004.

[13] L. Porter, C. B. Lee, and B. Simon, "Halving fail rates using peer instruction: A study of four computer science courses," in *Proceedings of the 44th ACM Technical Symposium on Computer Science Education (SIGCSE '13)*. New York, NY: Association for Computing Machinery, 2013, pp. 177–182, doi: 10.1145/2445196.2445250.

[14] A. Luxton-Reilly, Simon, I. Albluwi, B. A. Becker, M. Giannakos, A. N. Kumar, L. Ott, J. Paterson, M. J. Scott, J. Sheard, and C. Szabo, "Introductory programming: A systematic literature review," in *ITiCSE 2018 Companion: Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*. New York, NY: Association for Computing Machinery, 2018, pp. 55–106, doi: 10.1145/3293881.3295779.

[15] M. Kandlhofer, G. Steinbauer, S. Hirschmugl-Gaisch, and P. Huber, "Artificial intelligence and computer science in education: From kindergarten to university," in *2016 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2016, pp. 1–9, doi: 10.1109/FIE.2016.7757570.

[16] J. M. Wing, "Computational thinking," *Communications of the ACM*, vol. 49, no. 3, pp. 33–35, 2006, doi: 10.1145/1118178.1118215.

[17] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, and D. Dennison, "Hidden technical debt in machine learning systems," in *Advances in Neural Information Processing Systems 28 (NIPS 2015)*. Curran Associates, 2015, pp. 2503–2511, available: https://proceedings.neurips.cc/paper/2015/hash/86df7dcfd896fcaf2674f757a2463eba-Abstract.html.

[18] J. D. Bransford, A. L. Brown, and R. R. Cocking, *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academies Press, 2000, doi: 10.17226/9853.

[19] M. T. H. Chi, "Active-constructive-interactive: A conceptual framework for differentiating learning activities," *Topics in Cognitive Science*, vol. 1, no. 1, pp. 73–105, 2009, doi: 10.1111/j.1756-8765.2008.01005.x.

[20] J. Garfield and D. Ben-Zvi, *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*. Dordrecht, Netherlands: Springer, 2008, doi: 10.1007/978-1-4020-8383-9.

[21] G. W. Cobb, "Mere renovation is too little too late: We need to rethink our undergraduate curriculum from the ground up," *The American Statistician*, vol. 69, no. 4, pp. 266–282, 2015, doi: 10.1080/00031305.2015.1093029.

[22] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012, doi: 10.1145/2347736.2347755.

[23] M. I. Jordan, "On statistics, computation and scalability," *Bernoulli*, vol. 19, no. 4, pp. 1378–1390, Sep. 2013, doi: 10.3150/12-BEJSP17.

[24] J. S. Saltz and I. Shamshurin, "Data science education: Theory and practice," *Journal of Computing Sciences in Colleges*, vol. 37, no. 6, pp. 69–78, 2022.

[25] U. Fayyad and H. Hamutcu, "Toward foundations for data science and analytics: A knowledge framework for professional standards," *Harvard Data Science Review*, vol. 2, no. 2, 2020, doi: 10.1162/99608f92.1a99e67a.

[26] J. H. F. Meyer and R. Land, “Threshold concepts and troublesome knowledge (2): Epistemological considerations and a conceptual framework for teaching and learning,” *Higher Education*, vol. 49, no. 3, pp. 373–388, 2005, doi: 10.1007/s10734-004-6779-5.

Appendix A: Concept-Focused Assessment Instrument

Quiz 6: Classification and Evaluation was administered as a one-hour, ten-question written assessment. No code, libraries, or computational tools were permitted. Students were required to show reasoning and intermediate steps where applicable.

Q	Question
Q1	Explain the key differences between ZeroR and Random Forest in terms of model complexity, learning capability, and application scenarios.
Q2	You have a box of objects: 40% are red and 60% are blue; 20% are balls and 80% are cubes; half of the red objects are cubes. If you randomly select a cube, what is the probability that it is red? Show all steps using Bayes' Theorem.
Q3	Given the following classification outcomes for a three-class classifier—20 Cats: 18 correctly classified, 2 as Pigs; 20 Dogs: 16 correctly classified, 3 as Pigs, 1 as Cat; 15 Pigs: 6 correctly classified, 5 as Cats, 4 as Dogs—construct the confusion matrix.
Q4	Using the confusion matrix from Q3, compute overall accuracy, TPR for Pigs, and Precision for Dogs. Show all steps.
Q5	Given TP = 16, FP = 6, FN = 4 for Class A, calculate Precision, Recall, and F1-score. Show each formula and intermediate step.
Q6	What is a decision stump, and how does it differ from a J48 decision tree? In what situations might a decision stump outperform deeper trees?
Q7	Given: In-lab observations: Stylus (70%), Touch (20%); In-field observations: Stylus (20%), Touch (50%); Prior probabilities: Lab (70%), Field (30%); Sequence: Day 1 = Stylus, Day 2 = Touch, Day 3 = Mouse. Draw an HMM with hidden states (Lab, Field), observed emissions (Stylus, Touch, Mouse), transition probabilities, and emission probabilities.
Q8	A system detects "brushing teeth" for 4 minutes out of 1440. A majority classifier labels everything as "not brushing." Compute the accuracy. Explain why accuracy is misleading here. Suggest two alternative evaluation metrics.
Q9	Explain how a Support Vector Machine improves upon a basic linear classifier, both geometrically and in terms of generalization.
Q10	A classifier was trained: `if weight == 14 or weight == 16: return Dog; else: return Cat`. Training data: Cats [9,10,11]; Dogs [14,16,16]; 100% training accuracy. Test on: Cats [10,12,14]; Dogs [16,22,24]. How many predictions will be correct? What does this illustrate about training vs. testing on the same dataset?