

Integrating Artificial Intelligence into Undergraduate Cybersecurity Education: A Course Design for Threat Detection, Explainability, and Ethical Resilience

Vahid Heydari
Computer Science Department
Morgan State University
Baltimore, MD, USA
vahid.heydari@morgan.edu
0000-0002-6181-6826

Kofi Nyarko
Electrical and Computer Engineering Department
Morgan State University
Baltimore, MD, USA
kofi.nyarko@morgan.edu
0000-0002-7481-5080

Abstract—This paper introduces an undergraduate course, *Artificial Intelligence Applications in Cybersecurity*, designed to equip students with Artificial Intelligence (AI) and Machine Learning (ML) skills to address modern cyber threats. The curriculum integrates supervised and unsupervised learning, deep learning, explainable AI (XAI), adversarial ML, and ethical considerations. Using accessible tools (Python, Google Colab) and real-world datasets (e.g., NSL-KDD, CICIDS2017, malware corpora), students complete phased projects progressing from classical ML baselines to deep learning with interpretability (SHAP/LIME) and robustness against adversarial attacks (FGSM/PGD with mitigation). The course aligns with data science and cybersecurity workforce frameworks, emphasizing reproducibility, communication, and responsible AI practices.

Keywords—Artificial Intelligence (AI), Cybersecurity Education, Machine Learning (ML), Explainable AI (XAI), Adversarial Machine Learning, Ethics and Fairness, Undergraduate Curriculum Design, Workforce Development

I. INTRODUCTION

Artificial Intelligence (AI) is reshaping the cybersecurity landscape, offering tools for automated intrusion detection, anomaly identification, and rapid incident response. The increasing reliance on machine learning (ML) for security operations represents both an opportunity and a challenge. On one hand, AI-enabled systems can process massive data streams and detect sophisticated threats more effectively than rule-based approaches. On the other hand, AI introduces new risks, including susceptibility to adversarial manipulation, lack of transparency in decision-making, and ethical concerns about fairness, privacy, and accountability. Together, these realities underscore the urgent need for a cybersecurity workforce that is prepared not only to apply AI techniques but also to critically evaluate their trustworthiness.

The 2023 (ISC)² Cybersecurity Workforce Study reports a global shortage of approximately 3.4 million cybersecurity professionals [1], highlighting the challenge of filling critical roles in a rapidly evolving threat landscape. The NIST NICE

Cybersecurity Workforce Framework emphasizes the need for analysts with AI/ML expertise to address adversarial threats, ensure trustworthy automation, and manage decision-making ambiguity [2]. Without curricula that integrate AI techniques, explainability, and ethics, undergraduate students may enter the workforce underprepared. Historically Black Colleges and Universities (HBCUs), with their focus on equitable access, are uniquely positioned to address this gap through curricula like the proposed course, building on strategic roadmaps for AI integration [3].

Cybersecurity education has traditionally focused on networking, cryptography, and secure systems, with limited integration of modern AI/ML methods. Recent initiatives have begun introducing data science and ML concepts into curricula, yet significant gaps remain. Many programs emphasize basic classification and anomaly detection while neglecting emerging areas such as explainable AI (XAI), adversarial robustness, and ethical governance. Furthermore, most existing implementations appear in graduate programs or professional training, leaving undergraduates underexposed to the challenges and opportunities of AI in cybersecurity.

To address these shortcomings, this paper presents the design of a new undergraduate course, *Artificial Intelligence Applications in Cybersecurity*. The course integrates foundational ML techniques with security-focused applications, hands-on labs, and project-based learning. Students progress from introductory ML concepts to advanced topics such as deep learning for malware analysis, interpretability through SHAP and LIME, adversarial attacks and defenses, and ethical considerations in AI-driven security. Real-world datasets—such as NSL-KDD, CICIDS2017, and malware analysis corpora—ensure relevance and practical skill development.

This work builds upon prior research on curriculum innovation, including a strategic roadmap for AI in Cybersecurity education that emphasized experiential learning, adversarial defense, and interdisciplinary integration

[3]. By operationalizing these principles in a structured undergraduate course, the present study contributes a concrete model that can be adapted by institutions seeking to modernize their cybersecurity programs. The key contributions are:

- **Course Design:** A detailed description of objectives, student learning outcomes, and alignment with national frameworks.
- **Curriculum Implementation:** A week-by-week breakdown of modules, readings, labs, and projects.
- **Pedagogical Innovations:** Integration of explainability, adversarial robustness, and ethics into technical training.
- **Broader Impact:** A replicable framework that prepares undergraduates for AI-augmented threats while promoting fairness, transparency, and accountability.

The remainder of this paper is organized as follows: Section II reviews related work in AI-cybersecurity education. Section III outlines the course design and objectives. Section IV presents the curriculum and schedule. Section V details assessment strategies. Section VI discusses implementation insights and challenges. Section VII highlights broader impacts, and Section VIII concludes with directions for future work.

II. RELATED WORK

The integration of AI and ML into cybersecurity education has gained momentum in recent years, motivated by the growing reliance on AI-enabled systems for intrusion detection, anomaly analysis, and automated response. Several studies have introduced curricula that embed ML concepts into security contexts, often emphasizing hands-on approaches to bridge theory and practice.

Aris *et al.* [4] proposed incorporating predictive analytics and threat modeling into cybersecurity programs, though their work primarily targeted graduate-level learners. Romney *et al.* [5] developed a hands-on curriculum using R programming notebooks for novice AI practitioners, focusing on foundational ML techniques for malware and intrusion detection. At the high school level, Grover *et al.* [6] embedded AI in cybersecurity modules, demonstrating measurable gains in student understanding through pre- and post-surveys.

Professional training has also advanced in this area. For example, the SANS Institute's SEC595 course [7] emphasizes practical ML applications in intrusion detection and malware analysis, while Berkeley's MICS 207 [8] introduces ML for network analysis and anomaly detection. Recent work at Historically Black Colleges and Universities (HBCUs) by Heydari *et al.* [3] presented a strategic roadmap for AI in cybersecurity programs, stressing interdisciplinary collaboration and project-based learning. Similarly, Okpala *et al.* [9] designed AI-based cyberharassment detection projects, enabling students to apply ML methods with empirical evidence of improved engagement.

Despite these advances, gaps remain. Few undergraduate curricula integrate advanced topics such as XAI, adversarial machine learning, or ethical governance in cybersecurity contexts. Devi *et al.* [10] surveyed XAI applications in cybersecurity, highlighting tools such as SHAP and LIME for interpretability in phishing and intrusion analysis. Morris *et al.* [11] further demonstrated the value of deploying XAI tools to support analysts in high-stakes operations. For adversarial ML, the NIST National Cybersecurity Center of Excellence has outlined risks and defenses [12], and Li's CS 598 course at the University of Illinois [13] focuses on adversarial ML and game theory, but such material is rarely adapted for undergraduate education. Reports such as the Center for Security and Emerging Technology (CSET) study [14] have also stressed the vulnerabilities of AI models in cybersecurity settings, likening them to traditional software flaws.

Recent empirical studies further underscore adversarial risk and interpretability needs in security ML. Heydari and Nyarko analyze adversarial vulnerabilities in ransomware detection and propose robustness enhancements [15]. Complementary work evaluates network intrusion detection models under white-box and black-box attacks in enterprise settings [16], and introduces an adversarially trained neural network that improves NIDS robustness on UNSW-NB15 [17]. For explainability and fairness, Heydari and Nyarko demonstrate SHAP-based analyses to surface feature importance and fairness concerns in security models [18]. These results motivate early undergraduate exposure to adversarial testing, XAI, and fairness auditing.

Professional and graduate-level courses such as SANS SEC595 [7] and Berkeley's MICS 207 [8] provide valuable training in ML applications for cybersecurity, but they assume significant prior expertise and target advanced learners. In contrast, the course described in this paper is specifically designed for undergraduates, integrating foundational concepts with hands-on projects to make adversarial ML, XAI, and ethics accessible earlier in the academic pipeline. This distinction underscores the novelty of our contribution, as it offers a replicable model for institutions seeking to modernize undergraduate cybersecurity curricula.

Ethical, fairness, and privacy considerations are also underrepresented in most curricula. Marcraft [19] has suggested including ethical modules in AI-driven security courses, though practical implementations remain limited. This work builds on the roadmap for AI in cybersecurity education at HBCUs [3], contributing a concrete undergraduate course design that operationalizes these principles and links technical training with ethical responsibility.

III. COURSE DESIGN AND OBJECTIVES

The course *Artificial Intelligence Applications in Cybersecurity* is designed as an upper-division undergraduate offering (course number COSC 474) to introduce students to the intersection of AI and cybersecurity. It assumes prior

completion of programming, data structures, computer networks, and statistics, ensuring that students arrive with the foundational background necessary to engage with advanced concepts in ML and security applications.

The course is motivated by the need to prepare undergraduates for AI-augmented cybersecurity threats and defenses. While many curricula introduce ML concepts only at the graduate level, this course brings them into the undergraduate context, combining theoretical foundations with hands-on practice using real-world datasets such as NSL-KDD, CICIDS2017, and malware analysis corpora.

A. Course Objectives

The course is guided by five broad objectives that describe what students should achieve by the end of the semester:

1. Understand foundational AI and ML concepts as applied to cybersecurity.
2. Apply supervised, unsupervised, and deep learning methods to cyber data.
3. Analyze and interpret security-focused ML models using XAI techniques.
4. Evaluate the vulnerabilities of ML models to adversarial attacks and implement defenses.
5. Demonstrate practical skills and communication through a semester-long project.

B. Student Learning Outcomes (SLOs)

The Student Learning Outcomes (SLOs) are defined to be specific, measurable, and aligned with both course objectives

and the Data Science Initiative (DSI) Framework. Table I summarizes these alignments.

C. Pedagogical Approach

The course design emphasizes active learning and experiential engagement. Weekly lectures introduce theoretical material, which is reinforced by hands-on labs in Google Colab using Python, Scikit-learn, TensorFlow, and Keras. Each lab integrates short reflective questions to encourage conceptual understanding alongside technical skills. Case studies and readings supplement the labs to provide context on real-world applications and ethical dilemmas.

A phased project structure encourages cumulative learning: in the midterm project, students curate a dataset and apply classical ML models, while in the final project they extend this work with deep learning, XAI methods, and adversarial robustness. This scaffolded approach not only strengthens technical mastery but also builds confidence in communication and ethical reasoning, preparing students for workforce roles in AI-augmented cybersecurity.

IV. CURRICULUM AND SCHEDULE

The course follows a modular structure over a fifteen-week semester, with each module introducing new technical concepts and reinforcing them through hands-on labs, assigned readings, and weekly quizzes. Students progress from foundational ML concepts to advanced applications in explainability, adversarial robustness, and ethical analysis. The schedule is designed to scaffold learning such that early modules build the prerequisite skills for more advanced topics and project milestones.

TABLE I. Alignment of Course Objectives, Student Learning Outcomes (SLOs), and DSI Framework Areas

Objective	Student Learning Outcome (SLO)	DSI Framework Area
1	Define and explain core AI/ML concepts as they relate to threats, vulnerabilities, and defenses.	Mathematics & Statistics
2	Implement and compare supervised and unsupervised ML models (e.g., classification, clustering) on cybersecurity datasets.	Programming; Modeling
3	Apply SHAP and LIME to interpret model predictions and communicate findings effectively.	Modeling; Communication
4	Detect and evaluate adversarial attacks (e.g., FGSM, PGD) and propose mitigation strategies.	Ethics; Modeling
5	Design, implement, and present an AI-enabled cybersecurity project that integrates technical, analytical, and ethical considerations.	Programming; Data Curation; Communication

A. Weekly Modules

Table II provides a week-by-week outline of the curriculum, including readings, tools, and assignments. The selected textbooks [20]–[22] and datasets (e.g., NSL-KDD, CICIDS2017, malware PE images) were chosen to balance accessibility with relevance to real-world cybersecurity applications.

B. Project Phases

The course project is structured in two phases to encourage iterative development:

- **Midterm Project (Phase 1):** Students curate a dataset, apply supervised and unsupervised learning, and submit a 4–6-page technical report.
- **Final Project (Phase 2):** Students extend their work with a deep learning model, explainability analysis, and

adversarial/fairness evaluation, submitting an 8–12-page report and a demo presentation.

Figure 1 illustrates the scaffolded two-phase trajectory from classical ML baselines to deep learning with explainability, adversarial robustness, and fairness.

C. Pedagogical Rationale

This scaffolded curriculum ensures that students gain both technical proficiency and critical awareness. Weekly labs emphasize reproducibility through Google Colab, enabling students to complete assignments without the need for specialized hardware. Case studies and position papers integrate ethical reflection alongside technical training, aligning the course with CISSE’s emphasis on responsible and interdisciplinary cybersecurity education.

TABLE II. Course Modules, Readings, and Assignments

Week	Module Title	Readings / Materials	Assignments / Activities
1	Foundations of Python and ML	Géron Ch. 1–2; Colab setup guide	Lab: environment setup, implement a simple classifier
2	Introduction to AI in Cybersecurity	Chio Ch. 1; case study on threat modeling	Activity: create a basic threat model
3	Supervised Learning for Intrusion Detection	Chio Ch. 2; Géron Ch. 3–5	Lab: train Random Forest and SVM classifiers on NSL-KDD/CICIDS2017
4	Unsupervised Learning for Anomaly Detection	Géron Ch. 9	Lab: apply clustering techniques (e.g., k-means) to detect anomalies
5	Deep Learning for Malware & Network Analysis	Chio Ch. 3–4; Géron Ch. 10–11	Lab: malware detection using CNNs/RNNs on PE datasets
6	Explainable AI in Cybersecurity	Molnar Ch. 1–5, 13–18	Lab: use SHAP and LIME to interpret IDS and malware models
7	Adversarial Machine Learning: Attacks and Defenses	Chio Ch. 8; selected academic papers	Lab: simulate FGSM and PGD attacks; experiment with adversarial training
8	Ethics, Fairness, and Privacy in AI-Driven Security	Selected articles on fairness and privacy	Assignment: position paper analyzing ethical implications of AI in security
9–15	Final Project	N/A	Project deliverables: dataset curation, deep learning extension, XAI visualizations, adversarial/fairness evaluation, report, and demo

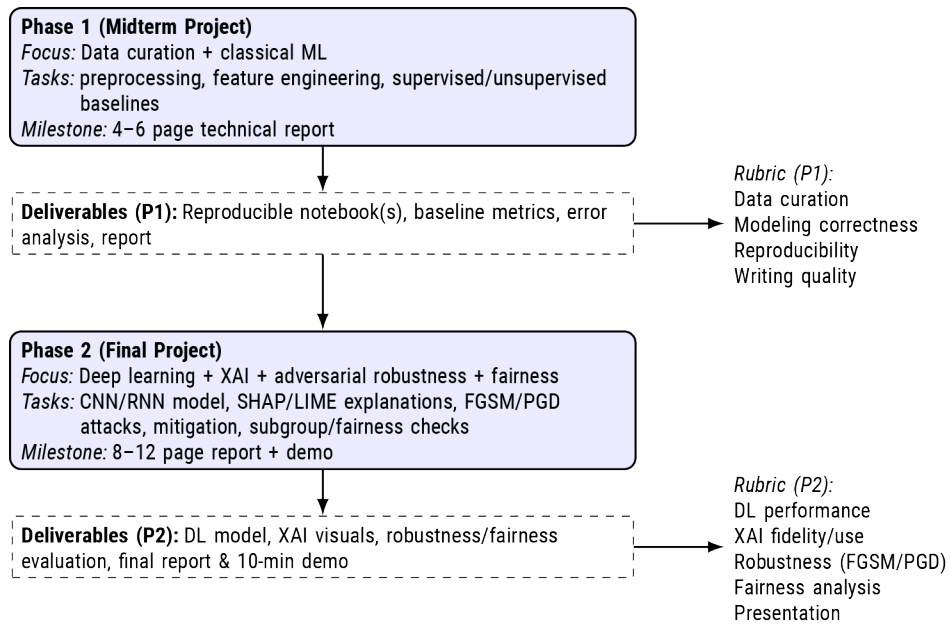


Fig. 1. Scaffolded project flow aligned with SLOs 2–5: Phase 1 establishes baselines; Phase 2 extends to deep learning, XAI, robustness, and fairness, culminating in a demo.

V. ASSESSMENT AND EVALUATION

Assessment in *Artificial Intelligence Applications in Cybersecurity* is designed to measure both technical proficiency and higher-order outcomes such as interpretability, ethical reasoning, and communication. The evaluation framework balances individual assignments, team-based projects, and in-class engagement to ensure students develop a broad set of skills.

A. Grading Breakdown

The grading scheme (Table III) balances summative projects with formative checkpoints. In addition to weekly labs and quizzes that reinforce core concepts, five low-stakes project milestones and a revision cycle distribute feedback across the term (see Subsection V-B). The phased project structure (Phase 1 → Phase 2) emphasizes growth: grades reflect progress on baselines, explainability, and robustness rather than a single high-stakes submission.

TABLE III. Assessment Categories and Weights

Assessment Category	Description	Weight
Weekly Labs	Hands-on Colab notebooks with embedded questions and brief writeups	25%
Weekly Quizzes	Short quizzes (\approx 30 MCQs) reinforcing weekly concepts	10%
Project Milestones & Peer Review	Five low-stakes checkpoints with structured feedback (proposal; dataset & ethics brief; baseline models; XAI plan; robustness protocol) + two guided peer-reviews	10%
Midterm Project (Phase 1)	Dataset curation; supervised/unsupervised models; 4–6 page report	20%
Final Project (Phase 2)	Deep learning with XAI and adversarial robustness; 8–12 page report + demo	25%
Revision & Reflection Memos	Post-feedback revisions (1 cycle) for Phase 1 and 2 + 1-page reflection on changes	5%
Attendance / Participation	In-class engagement, studio sessions, code walkthroughs	5%
Total		100%

B. Iterative Feedback and Mastery Policy

To reduce end-loaded grading and promote learning through feedback, the course incorporates structured milestones and revision cycles:

- **Milestones (10% total):** Five checkpoints (Weeks 2, 4, 6, 7, 9) with rubric-aligned feedback:
 1. *Project Proposal* (problem, scope, risks) – 2%
 2. *Dataset & Ethics Brief* (sources, curation, privacy/fairness risks) – 2%
 3. *Baselines Checkpoint* (classical models, metrics, error analysis) – 2%
 4. *XAI Plan* (intended SHAP/LIME use, target questions, fidelity checks) – 2%
 5. *Robustness Protocol* (FGSM/PGD parameters, defenses to test, reporting plan) – 2%
- **Peer Review (within Milestones):** Two guided peer-reviews (rubric prompts) to surface issues early and improve communication.
- **Revision Cycle (5% total):** One optional resubmission for Phase 1 and Phase 2 deliverables after feedback. Revisions can recover up to 50% of points lost on the original submission; each revision is accompanied by a 1-page reflection memo documenting changes and rationale.

- **Studio Time:** Dedicated in-class studio blocks for formative feedback on code, XAI artifacts, and robustness analyses; instructor/TA feedback is returned within 72 hours.

This structure distributes assessment across the semester, incentivizes early progress, and increases transparency of expectations while preserving the integrative value of capstone-style projects.

C. Alignment with Learning Outcomes

Each assessment is aligned with the SLOs defined in Section III. Table IV summarizes how different assessment components map to SLOs and course objectives.

D. Evaluation Approach

The evaluation strategy emphasizes:

- **Technical competence:** measured through labs and project implementation.
- **Critical thinking:** assessed via quizzes, lab reflections, and ethical analysis in Module 8.
- **Communication:** reinforced through project reports, position papers, and final presentations.
- **Collaboration:** embedded in team-based project work and classroom participation.

This multi-dimensional framework ensures students not only acquire AI/ML technical skills but also develop the interpretive, ethical, and communication competencies necessary for professional practice in cybersecurity.

TABLE IV. Alignment of Assessments with Student Learning Outcomes (SLOs)

Assessment	Aligned SLOs	Skills Assessed
Weekly Labs	SLO 2 (implement ML models), SLO 3 (apply XAI), SLO 4 (adversarial defense)	Programming, modeling, data analysis, reproducibility
Weekly Quizzes	SLO 1 (define and explain AI/ML concepts), SLO 2 (compare models)	Conceptual mastery, terminology, critical recall
Midterm Project (Phase 1)	SLO 2 (supervised/unsupervised ML), SLO 5 (project execution)	Data curation, technical writing, teamwork, analysis
Final Project (Phase 2)	SLO 3 (XAI), SLO 4 (adversarial defense), SLO 5 (project delivery)	Deep learning, explainability, robustness, ethics, communication
Attendance	All SLOs (indirectly)	Engagement, collaboration, professional readiness

VI. IMPLEMENTATION INSIGHTS AND CHALLENGES

Although the course has not yet been piloted, several insights and anticipated challenges were identified during the design process. These considerations are important for ensuring successful implementation and for guiding institutions that may wish to adapt the curriculum.

A. Anticipated Student Engagement

The course emphasizes project-based and hands-on learning, which is expected to promote strong engagement. Weekly labs in Google Colab lower technical barriers by providing a uniform environment that does not require local installations or specialized hardware. Students are anticipated to be most engaged during the phased project, where they progress from building classical ML models to implementing deep learning, XAI, and adversarial defenses on real-world cybersecurity datasets.

B. Accessibility and Resource Considerations

A core design principle is accessibility. By relying on open-source software (Python, Scikit-learn, TensorFlow) and cloud-based infrastructure (Google Colab), the course minimizes dependence on high-performance computing resources. This makes it suitable for institutions with limited infrastructure, including minority-serving and resource-constrained universities. However, anticipated challenges include varying levels of student preparation in programming and statistics. To mitigate this, the course incorporates optional review materials and early diagnostic assessments.

C. Faculty Preparation

Delivering a course at the intersection of AI and cybersecurity requires interdisciplinary expertise. Institutions may face challenges related to faculty readiness, particularly where instructors have strong backgrounds in one area but not both. Potential solutions include faculty development workshops, AI/cybersecurity boot camps, and co-teaching models that leverage expertise across departments.

D. Equity and Inclusion

As highlighted in prior work [3], Historically Black Colleges and Universities (HBCUs) and other minority-serving institutions have a unique opportunity to lead in AI-cybersecurity education. However, they also face structural challenges such as resource limitations and diverse student preparation levels. The use of cloud-based tools and open educational resources (OERs) in this course is intended to reduce barriers and ensure equitable access, while the explicit inclusion of fairness and ethics reinforces the importance of responsible AI in diverse learning contexts.

E. Planned Evaluation Approach

To strengthen credibility and ensure continuous improvement, a structured evaluation plan will be implemented when the course is offered. Planned methods include:

- **Pre- and Post-Course Surveys:** To measure changes in student confidence and self-assessed knowledge of AI/ML, cybersecurity applications, and ethical reasoning.
- **Rubric-Based Project Scoring:** Midterm and final projects will be evaluated using standardized rubrics that assess technical accuracy, reproducibility, interpretability, and ethical analysis.
- **Formative Assessments:** Weekly quizzes will provide data on conceptual understanding of topics such as supervised learning, anomaly detection, XAI, and adversarial ML.
- **External Review:** Selected final projects may be reviewed by faculty peers or industry partners to benchmark outcomes against workforce expectations.

These evaluation mechanisms will provide both quantitative and qualitative evidence of student learning outcomes. The results will inform iterative refinement of the curriculum and contribute to the broader research community on cybersecurity education.

F. Sample Evaluation Instruments (Illustrative)

To make the evaluation plan more concrete, we provide examples of survey items and a rubric excerpt that will be adapted for the pilot offering. Table V summarizes the Phase 2 project rubric dimensions used to score deep learning performance, explanation fidelity (SHAP/LIME), robustness testing (FGSM/PGD), and ethical/fairness analysis.

Sample Pre/Post Likert Items (1=Not Confident to 5=Very Confident):

- Rate your confidence in implementing **SHAP** to interpret a trained classifier in a cybersecurity task (e.g., IDS).
- Rate your confidence in generating **FGSM/PGD** adversarial examples and evaluating their impact on model performance.
- Rate your confidence in explaining differences between **ROC-AUC** and **Precision-Recall AUC** for imbalanced security datasets.
- Rate your confidence in articulating at least two **fairness risks** in AI-driven security pipelines and proposing mitigations.

Open-Ended Items:

- Describe one way **XAI** can influence a security analyst's decision during incident response (2–3 sentences).
- Briefly identify one **ethical concern** that may arise when deploying ML for malware or intrusion detection in production.

TABLE V. Excerpt from Project Rubric (Phase 2: DL + XAI + Robustness)

Criterion	Points	Anchors / Evidence
Technical Accuracy (DL)	0–10	Model trains correctly; appropriate architecture/regularization; justified hyperparameters; sensible baselines
Reproducibility	0–5	Clean repo/notebooks; fixed seeds; clear README; data processing documented
XAI Fidelity (SHAP/LIME)	0–10	0–3: superficial visuals, no linkage to decisions; 4–7: correct plots with partial explanation; 8–10: clear, faithful explanations tied to specific alerts/cases, limitations discussed
Robustness Evaluation (FGSM/PGD)	0–10	Attacks correctly implemented; effect on metrics reported; defense/mitigation attempted and analyzed
Ethical/Fairness Analysis	0–5	Identifies risks; reports subgroup metrics; proposes realistic mitigation or monitoring
Communication & Presentation	0–10	Clear writing/figures; concise demo; evidence-based Q&A

G. Measuring Ethical Reasoning

Ethical reasoning is assessed with a scenario-based case study (Module 8), an embedded fairness audit in the final project, and a brief oral “responsible AI” defense during the demo. Evidence is collected via (i) a position paper (2–3 pages) and (ii) rubric-scored artifacts in the Phase 2 deliverables (fairness metrics, XAI justifications, and mitigation plans).

Case Prompt (Module 8): Your team is deploying an AI-based phishing detector for an enterprise with multilingual staff. A pilot shows higher false-positive rates for messages from non-native English writers. Decide whether and how to deploy the system. Identify stakeholders and harms, analyze fairness/privacy trade-offs, use model explanations (e.g., SHAP) to justify decisions, propose subgroup metrics to monitor, and outline a mitigation/appeals process.

Ethical Reasoning Rubric (Position Paper & Project Artifacts): Table VI details the scoring dimensions. Two graders will calibrate on sample submissions; disagreements > 2 points trigger discussion and rescoreing.

Embedded Fairness Audit (Phase 2): Teams must compute at least one subgroup fairness analysis (e.g., per-group precision/recall or equalized odds gap), describe observed disparities, and attempt a mitigation (e.g., threshold adjustment, sample reweighting, or post-processing). Results

are scored under the rubric’s *Fairness Metrics & Mitigation* and *Evidence & XAI Use* criteria.

Oral Responsible-AI Defense (3 minutes): During the final demo, teams answer one ethics/fairness prompt drawn from their project (e.g., “What harm could arise from false negatives in your IDS setting, and how would you monitor it post-deployment?”). Clarity and evidence-based responses contribute to the *Trade-offs & Communication* score.

Sample Pre/Post Survey Items (Likert 1–5):

- I can **identify stakeholders and harms** for an AI-based security system in a given context.
- I can use **SHAP/LIME** to justify or revise a deployment decision for a security model.
- I can **compute and interpret subgroup metrics** (e.g., precision/recall by group) and propose a mitigation.
- I can **explain privacy vs. security trade-offs** when deploying automated detection in sensitive settings.

Weighting in the Course Grade: The Ethical Reasoning Rubric contributes **5%** of the overall course grade without altering the category weights in Table III: (i) the Module 8 position paper counts for **3%** (within *Weekly Labs*); and (ii) Phase 2 project artifacts (fairness audit and XAI justification) account for **2%** (within the *Final Project*).

TABLE VI. Ethical Reasoning Rubric (10 Points Total)

Criterion (0–2 each)	Performance Anchors
Stakeholders & Harms Identified	0: superficial; 1: partial list; 2: comprehensive with concrete harms
Principles & Policies Applied	0: none; 1: mentions privacy/bias/accountability; 2: applies them to decisions with justification
Evidence & XAI Use	0: no evidence; 1: SHAP/LIME used but not tied to action; 2: explanations support/alter a decision
Fairness Metrics & Mitigation	0: none; 1: reports subgroup metrics; 2: metrics plus realistic mitigation/monitoring plan
Trade-offs & Communication	0: vague; 1: acknowledges trade-offs; 2: clear argument, limits acknowledged

VII. BROADER IMPACTS

The proposed course has broader impacts that extend beyond its immediate implementation. By combining technical AI/ML training with cybersecurity applications, explainability, adversarial defense, and ethics, it contributes to national priorities in workforce development, educational innovation, and equitable participation in STEM.

A. Cybersecurity Workforce Development

The curriculum is explicitly aligned with the NICE Cybersecurity Workforce Framework, preparing students for roles such as *Cyber Defense Analyst*, *Data Analyst*, and *AI Security Specialist*. By emphasizing real-world datasets, adversarial robustness, and model interpretability, the course ensures that graduates are not only technically proficient but also capable of deploying AI responsibly in high-stakes environments. This workforce orientation is particularly critical in the age of AI and automation, where ambiguous and evolving threats demand adaptable professionals.

B. Open Educational Resources and Dissemination

A major impact of this course is its potential for dissemination as an OER. All labs, lecture slides, and assessments are designed for release under a permissive license, enabling adoption by faculty at other institutions. Materials will be published on open platforms such as GitHub for direct access to code and Jupyter/Colab notebooks, and Clark Creative Education for structured curricula and instructor training resources. Institutional repositories and the CISSE community portal will further support distribution. This ensures broad scalability and encourages collaboration across institutions.

C. Scalability and Institutional Adoption

Because the course relies on cloud-based infrastructure (e.g., Google Colab) and open-source tools, it is scalable to institutions with varying levels of resources. This makes it an attractive option for universities seeking to modernize their cybersecurity programs without investing in expensive hardware. The modular structure also allows instructors to adopt individual components—such as the XAI or adversarial ML modules—even if they cannot offer the full course.

D. Equity, Diversity, and Inclusion

This course design advances diversity and inclusion by creating an AI in cybersecurity pathway accessible to undergraduates at HBCUs and other minority-serving institutions. Integrating fairness and ethics as technical topics reinforces the message that responsible AI is inseparable from cybersecurity education. This approach equips students not only with technical expertise but also with the ethical grounding needed to lead in shaping trustworthy AI systems.

E. Contribution to Educational Research

The course offers a replicable case study for the integration of AI/ML into cybersecurity curricula at the undergraduate level. Planned evaluation methods, including surveys and project-based assessments, will generate empirical insights into student learning outcomes. These results can inform future curriculum design, contribute to the research literature on cybersecurity education, and support cross-institutional collaboration on AI-driven cybersecurity training.

F. Summary of Broader Impacts

In summary, the broader impacts of this work include:

- Strengthening the cybersecurity workforce by integrating AI/ML skills with ethical and explainability training.
- Disseminating OER course materials through GitHub, Clark Creative Education, and institutional repositories.
- Enhancing scalability through cloud-based, open-source infrastructure.
- Promoting diversity and inclusion by designing content accessible to HBCUs and resource-limited institutions.
- Contributing to educational research through empirical evaluation and publication.

VIII. CONCLUSION AND FUTURE WORK

This paper presented the design of an undergraduate course, *Artificial Intelligence Applications in Cybersecurity*, that integrates foundational AI/ML methods with advanced topics such as explainable AI, adversarial robustness, and ethics. The course operationalizes experiential learning, interdisciplinary integration, and responsible AI practices within a structured, semester-long curriculum. Through weekly labs, quizzes, and phased projects, students build practical competence in applying ML to security datasets while developing the critical awareness needed to evaluate trustworthiness, transparency, and operational risk in AI-enabled security systems.

The course addresses gaps in current cybersecurity education, where undergraduate exposure to AI often remains limited to basic modeling concepts without sufficient attention to explainability, adversarial threat models, or ethical governance. By emphasizing real-world datasets, cloud-based tooling for accessibility, and explicit treatment of fairness and accountability, the curriculum prepares students to navigate cybersecurity practice in the age of AI and automation. The modular design also supports adoption in diverse institutional contexts, from resource-limited programs to research-intensive universities.

To address long-term success and institutional expectations, the course is designed to produce measurable evidence that aligns with common program outcomes in undergraduate computing and cybersecurity programs, including problem analysis and solution design, secure and ethical professional practice, effective communication, and teamwork. Evaluation will combine direct and indirect measures across offerings. Direct evidence will include rubric-scored artifacts from the phased projects (modeling correctness, reproducibility, SHAP/LIME-based justification quality, adversarial robustness protocols and results, and fairness/ethics audits), along with pre/post measures of conceptual mastery on core topics such as imbalanced evaluation metrics and robustness testing. Indirect evidence will include student self-efficacy surveys, reflection and revision memos, peer-review quality indicators, and

practitioner or faculty peer review of selected final projects. Results will be summarized each year and mapped to both course SLOs and program outcomes to support continuous improvement and program-level assessment.

Future work will focus on piloting the curriculum, collecting empirical evidence of student learning outcomes, and refining modules based on assessment results and stakeholder feedback. Planned enhancements include deeper integration of fairness-aware training, development of faculty support materials and training workshops, and expansion into interdisciplinary collaborations with policy and law programs. Longer-term efforts will disseminate the course as an open educational resource (OER) to support adoption by HBCUs and other institutions seeking to modernize their cybersecurity curricula.

In summary, this course represents a concrete step toward advancing undergraduate cybersecurity education in the age of AI. By combining technical depth with interpretability, robustness, and ethics, it equips the next generation of cybersecurity professionals to both harness and critically evaluate AI in practice, contributing to more resilient and trustworthy digital systems.

ACKNOWLEDGEMENT

This work is supported in part by the Center for Equitable Artificial Intelligence and Machine Learning Systems (CEAMLS) at Morgan State University. Generative AI and automated tools were used to assist in the production of this paper, but no AI tools were credited as authors, in accordance with the submission guidelines.

REFERENCES

- [1] ISC², "2023 cybersecurity workforce study," (ISC)², Tech. Rep., 2023, accessed: 2025-08-28. [Online]. Available: <https://www.isc2.org/Research/Workforce-Study>
- [2] NIST National Initiative for Cybersecurity Education (NICE), "Nice cybersecurity workforce framework," <https://www.nist.gov/itl/applied-cybersecurity/nice/nice-framework-resource-center>, 2020, accessed: 2025-08-28.
- [3] V. Heydari and K. Nyarko, "Empowering the next generation: A strategic roadmap for ai in cybersecurity education," in *Journal of The Colloquium for Information Systems Security Education*, vol. 12, no. 1, 2025, pp. 8–8. [Online]. Available: <https://doi.org/10.53735/cisse.v12i1.202>
- [4] D. O. M. R. M. F. AHMET ARIS, Luis Puche Rondon and A. Uluagac, "Integrating artificial intelligence into cybersecurity curriculum: New perspectives," in *2022 ASEE Annual Conference & Exposition*, no. 10.18260/1-2-41761. Minneapolis, MN: ASEE Conferences, August 2022, <https://peer.asee.org/41761>.
- [5] G. W. Romney, J. Guymon, M. D. Romney, and D. A. Carlson, "Curriculum for hands-on artificial intelligence cybersecurity," in *2019 18th International Conference on Information Technology Based Higher Education and Training (ITHET)*, 2019, pp. 1–8.
- [6] S. Grover, B. Broll, and D. Babb, "Cybersecurity education in the age of ai: Integrating ai learning into cybersecurity high school curricula," in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, ser. SIGCSE 2023. New York, NY, USA: Association for Computing Machinery, 2023, p. 980–986. [Online]. Available: <https://doi.org/10.1145/3545945.3569750>

- [7] SANS Institute, "Sec595: Applied data science and machine learning for cybersecurity," <https://www.sans.org/cyber-security-courses/applied-data-science-machine-learning/>, accessed: 2025-08-28.
- [8] UC Berkeley School of Information, "Mics 207: Artificial intelligence and machine learning in cybersecurity," <https://www.ischool.berkeley.edu/courses/cyber/207>, accessed: 2025-08-28.
- [9] E. Okpala, N. Vishwamitra, K. Guo, S. Liao, L. Cheng, H. Hu, Y. Wu, X. Yuan, J. Wade, and S. Khorsandroo, "Ai-cybersecurity education through designing ai-based cyberharassment detection lab," *arXiv preprint arXiv:2405.08125*, 2024.
- [10] S. N. Devi, E. and S. C. Babu, "A comprehensive survey on explainable ai in cybersecurity domain," https://setsindia.in/wp-content/uploads/2024/06/XAI_Cybersecurity.pdf, Society for Electronic Transactions and Security (SETS), Tech. Rep., 2024, accessed: 2025-08-28.
- [11] E. Morris, M. Nyre-Yu, M. Smith, B. Moss, and C. Smutz, "Explainable ai in cybersecurity operations: Lessons learned from xai tool deployment." Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), Tech. Rep., 2022.
- [12] NIST NCCoE, "Artificial intelligence: Adversarial machine learning," <https://www.nccoe.nist.gov/ai/adversarial-machine-learning>, accessed: 2025-08-28.
- [13] B. Li, "Cs 598: Special topics on adversarial machine learning," <https://aisecure.github.io/TEACHING/CS598/CS562.html>, accessed: 2025-08-28.
- [14] M. Musser, A. Lohn, J. X. Dempsey, J. Spring, R. S. S. Kumar, B. Leong, C. Liaghati, C. Martinez, C. D. Grant, D. Rohrer *et al.*, "Adversarial machine learning and cybersecurity: Risks, challenges, and legal implications," *arXiv preprint arXiv:2305.14553*, 2023.
- [15] V. Heydari and K. Nyarko, "Adversarial vulnerabilities in ransomware detection: Enhancing machine learning model robustness," in *2025 IEEE 22nd Consumer Communications & Networking Conference (CCNC)*, 2025, pp. 1–4.
- [16] —, "Evaluating network intrusion detection models for enterprise security: Adversarial vulnerability and robustness analysis," in *Proceedings of the 27th International Conference on Enterprise Information Systems - Volume 1: ICEIS, INSTICC*, SciTePress, 2025, pp. 699–708.
- [17] —, "Enhancing adversarial robustness in network intrusion detection: A novel adversarially trained neural network approach," *Electronics*, vol. 14, no. 16, 2025. [Online]. Available: <https://www.mdpi.com/2079-9292/14/16/3249>
- [18] —, "Fairness in machine learning for cybersecurity: Enhancing trust through feature importance and shap analysis," in *2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2024, pp. 1–6.
- [19] Marcraft, "4 ways to include ai in cybersecurity education," <https://marcraft.com/4-ways-to-include-ai-in-cybersecurity-education/>, 2024, accessed: 2025-08-28.
- [20] C. Chio and D. Freeman, *Machine learning and security: Protecting systems with data and algorithms.* O'Reilly Media, Inc., 2018.
- [21] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow.* O'Reilly Media, Inc., 2022.
- [22] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.* Leanpub, 2023, accessed: 2025-08-28. [Online]. Available: <https://christophm.github.io/interpretable-ml-book/>