ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE

Paper ID #36570

Promoting AI Trustworthiness through Experiential Learning

Alvis Fong

Dr. Fong is with the Department of Computer Science at Western Michigan University. His research interests revolve around data-driven knowledge discovery and aspects of machine intelligence, such as learning for classification and knowledge representation and reasoning. His scientific contributions include two books, fourteen book sections/chapters, two international patents, and 215 papers in reputable journals and conference proceedings. Leading journals that carry his work include IEEE T-KDE, IEEE T-ITBiomed, IEEE T-MM, IEEE T-Evolutionary Computing, IEEE T-Affective Computing, IEEE T-II, and a few other IEEE Transactions titles. He has served on several journal editorial boards and numerous conference committees. Dr. Fong holds four degrees in EE and CS. He is a registered Chartered Engineer and European Engineer.

Steven Carr (Professor & Chair)

Ajay Gupta (Professor)

Shameek Bhattacharjee (Assistant Professor)

Promoting AI Trustworthiness through Experiential Learning (WIP)

The authors are with Western Michigan University

Abstract – Despite highly publicized advances in artificial intelligence (AI) systems and technologies, there have been numerous reports that indicate AI can sometimes exhibit undesirable behaviors. When AI algorithms run on high-performance cyberinfrastructure (CI), such misbehaviors can multiply to obscure the root causes. Secure, safe, and reliable (SSR) computing principles can mitigate these problems. This project aims to inform curriculum development by creating and evaluating experiential learning materials to educate students from the outset. Three levels of preparedness cater to a wide range of learners. Specifically, members of the Transformative Interdisciplinary Human + AI Research Group at Western Michigan University, together with public and private partners, aim to address a critical shortage in STEM workforce that understands the anticipated changes brought about by AI. Taking a convergent approach, the project primarily aims to integrate core literacy and advanced skills at the intersection of SSR Computing, High Performance Computing (HPC), and AI into the educational curriculum across multiple STEM disciplines. The work focuses on institutions that have comparatively lower levels of advanced CI adoption. The project's secondary aim is to lay the groundwork for future broadening adoption of advanced CI training resources that have the potential to influence wide segments of CI communities. The project is currently in progress and this paper presents findings to date.

1. Introduction

Artificial intelligence (AI) systems can sometimes perform undesirably or be manipulated to exhibit biases and abusive behaviors. Machine learning (ML), which originated from an AI's ability to learn, has led the way in recent advances in AI. However, contemporary ML including the well-publicized deep learning (DL) neural networks are statistical in nature. ML algorithms can be trained to provide exemplary performances in a statistical sense, but individual results can be unreliable. For example, DL-based image recognition engines can routinely outperform humans with near perfect accuracies achievable in a range of tasks, e.g., [1] – [3]. Unfortunately, individual results can sometimes be surprisingly poor. When such algorithms fail, the failure mode is often unclear [4]. Furthermore, because ML/DL algorithms rely heavily on data for training and validation, the quality and quantity of data available can be a critical issue [5]. Apart from biases inherent in data, even "well-intentioned" researchers and developers will replicate societal biases and distort results. This is often due to the lack of diverse perspectives and contributions to the AI systems created. It is important to understand that AI is not an objective tool. In addition, bad actors can potentially inject malicious content into data to negatively influence the training process.

When AI/ML algorithms are parallelized on high-performance cyberinfrastructure (CI), such misbehaviors and uncertainty can multiply to obscure the root causes. Secure, safe, and reliable (SSR) computing techniques, which are pillars that support AI trustworthiness, can mitigate these problems. This project aims to inform curriculum development by creating and evaluating experiential learning materials to educate computer science (CS) and other STEM students who use AI from the outset. There is an emphasis on experiential learning because it has been found effective in a wide range of use cases across multiple disciplines involving both theory (concepts) and practice (labs), e.g. [6]-[8].

Using the developed educational materials, learners first become aware of the issues and then they are guided towards developing a range of practical skills toward mitigating those issues. Intensive, multi-faceted, modular, experiential learning units are designed to rapidly upgrade the skills of current and future CI users, so they become equipped with new skills to apply to their tasks. The loosely coupled modules can be taken as standalone self-directed units to suit CI professionals. The modules can also be integrated into existing classes, starting with CS 1 and CS 2, which are routinely taken by many non-CS STEM students. Three levels of preparedness (foundation, intermediate, and advanced), which roughly correspond to lower undergraduate, upper undergraduate, and graduate, cater to a wide range of learners. In a sandbox environment, learners of each module take measured risks when guided on a challenging yet fun journey of discovery and knowledge acquisition.

Supported by an NSF grant (no. 2017289), members of the Transformative Interdisciplinary Human + AI Research Group at Western Michigan University (WMU), together with public and private partners at various US locations and beyond, aim to address a critical shortage in STEM workforce that understands the anticipated immense technical and societal changes brought about by AI. The need for a strong AI-informed workforce is exemplified by the American AI Initiative (https://www.whitehouse.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/). Taking a convergent approach, the project primarily aims to integrate core literacy and advanced skills at the intersection of SSR Computing, High Performance Computing (HPC), and AI into the educational curriculum across multiple STEM disciplines. This will prepare faculty, undergraduate, and graduate students for large-scale data handling and analytics. The work focuses on institutions that have comparatively lower levels of advanced CI adoption, such as second-tiered institutions (Carnegie Classification R2), historically black colleges and universities (HBCU), and community colleges.

The project's secondary aim is to lay the groundwork for future broadening adoption of advanced CI training resources that have the potential to influence wide segments of CI communities. This is achieved through identification of best practices derived from the project, modular curricula, and experiential hands-on learning materials. The course is further advanced with carefully designed outreach activities to establish and maintain a pipeline of talents from pre-college to 2-year and 4-year colleges, and graduate programs. Although there is an emphasis on CS curriculum, non-CS STEM students and practitioners who frequently apply AI to their tasks are also intended users of the educational materials.

The project, which began in August 2020, is currently underway. The purpose of this paper is to share the project team's exciting endeavor broadly. Specific information to be disseminated at the conference include:

- Design of a series of reproducible, customizable, modular, experiential educational units that can be integrated within existing courses and/or taken as standalone self-directed learning activities.
- Results (to date) of actual use of the educational units in multiple settings across STEM disciplines (CS, branches of engineering, statistics, business analytics).
- Proposed outreach activities.
- Highlights of on-going and future work.

2. Experiential Learning Modules

This project aims to design and develop experiential modular learning materials that complement – not compete with – available resources from educational and commercial entities, e.g., MIT OpenCouseWare (https://ocw.mit.edu/index.htm), Edx (https://www.edx.org/), and Coursera (https://www.coursera.org/). In addition, the newly developed materials will be freely available on the research group's and project websites (https://fong.cs.wmich.edu/ and https://wmich.edu/cs/cybertraining).

2.1 Characteristics of the learning materials

Table I summarizes the key characteristics of the developed learning materials.

Table I. Key characteristics of developed materials.

Characteristic	Rationale / Purpose
Active, experiential learning	Enrich learning experience and outcome
Modular	Manageable chunks of learning materials
Loosely coupled	No strict order to complete the modules; multiple entry points
In class or online	Flexible delivery modes for students and other learners
Sandbox environment	Encouragement to take calculated risks
Multi-level	Three difficulty levels to cater to a range of backgrounds
Multidisciplinary experts	Both to inform the project and test the modules
Relevant examples	Customizable in consultation with domain instructors

A centerpiece of the developed modules is a strong emphasis on active experiential learning. In an immersive environment, learners gain first-hand experience in various aspects of SSR AI that can be readily applied to their problems across different quantitative disciplines. Active learning helps learners internalize the newly gained knowledge and skills. The modular design of learning units, which are loosely coupled, allow learners to mix and match modules that best meet their needs. There is no strict order or specific pathways that dictate how learners use the various modules.

The current pandemic has had a strong influence on learning activities. Also known as a hi-flex design, which stands for high flexibility, the learning modules have been developed to be compatible in both a traditional classroom environment and self-directed online learning. In fact, a mixed mode approach has been taken in the evaluation phase of the study. The learn-by-doing ethos associated with active learning is also manifested in a sandbox environment in which learners are encouraged to think “outside the box” for possible solutions that mitigate some of the vulnerabilities of using AI/ML. Learners feel safe to try and take calculated risks without serious repercussions that might affect their grades or degree of satisfaction.

Another important characteristic of the modular learning materials is that they are designed to approximately map to three levels of readiness equivalent to lower, upper undergraduate, and graduate. These do not represent any strict delineation; an especially well-prepared lower classman can take a module designed for a higher-level audience whenever they feel comfortable. Finally, a panel of experts drawn from multiple quantitative disciplines, who come from a 2- or 4-year academic institution, private and government research organizations, and other relevant stakeholders are on board. These experts inform the development of the learning modules, help disseminate the work among their circles, and some have additionally offered to test some of the modules in their respective environments. Through a “train-the-trainers” scheme, they help propagate the work. Furthermore, individual instructors who are interested in adopting any of the available modules are encouraged to work with the research team to customize the learning modules with domain-relevant examples and use cases to better cater to their students’ needs.

2.2 Outline of initial learning modules

The following twelve initial modules, which cover the depth and breadth of SSR AI, have been developed with a launch date of late August 2021:

1. Math Toolkit for SSRAI running on HPC CI. (foundational) This module can be integrated with foundational math taken by CS students.

2. Algorithmic Exploration and Exploitation of an Intelligent System’s weakness. (foundational) This module can be integrated with CS 1110 Computer Science I.

3. Modular and Structured Software Development for Robust Intelligent Systems that run on HPC CI. (foundational) This module can be integrated with CS 1120 Computer Science II.

4. Data Structures for SSRAI running on HPC CI. (intermediate). This module can be integrated with a Data and File Structures course.

5. Deep learning with HPC. (intermediate) This module can be integrated with a Data and File Structures course.

6. SSRAI Software Development for HPC CI deployment. (advanced, undergraduates and graduates) This module can be integrated with a Software Development / Engineering course.

7. Vulnerabilities of Machine Learning. (advanced, undergraduates and graduates) This module can be integrated with a Machine Learning course.

8. Beyond current generation AI and Toward Artificial General Intelligence. (advanced, undergraduates and graduates) This module can be integrated with an Artificial Intelligence or Cyber-Physical Systems course.

9. Adversarial Machine Learning and Robust Trust Scoring Models. (advanced, undergraduates and graduates) This module can be integrated with courses related to ML, Cybersecurity, Networks, or Computer Vision.

10. Societal Impact of AI. (advanced, undergraduates and graduates) This module can be integrated with an AI or Advanced Software Development course.

11. Pitfalls of applying AI to Information Retrieval tasks. (advanced, graduates)

This module can be integrated with an Information Retrieval (IR) course.

12. Real-Time SSRAI with HPC CI. (advanced, undergraduates and graduates) This module can be formulated as a capstone project.

Note that the above suggested course-module integration pairs apply equally to similar courses.

3. Evaluation and Results to Date

3.1 Evaluation instruments

Evaluation of the effectiveness of the developed learning materials is primarily conducted in the form of a pair of pre- and post-intervention questionnaires. These questionnaires have been prepared by an independent evaluation specialist, who also performs analysis after the questionnaires have been completed voluntarily and anonymously. The questionnaires consist of a mix of free-text questions, such as “How would you define Safe, Secure, and Reliable AI?”, and questions that ask participants to rate something like “How strongly do you agree or disagree with this statement: It is worth learning about safe, secure, and reliable AI.” on a 5-point Likert scale. The pre-intervention questionnaire comprises a total of five questions. The post-intervention questionnaire, which comprises 20 questions, is significantly more involving than the pre-intervention questionnaire.

3.2 Mixed-mode pilot

Designed and developed with flexibility in mind, the learning modules can be fully integrated into existing or future courses. For example, complete or partial modules can be adopted by individual instructors as classroom activities. Assessment can be formative or summative as determined by the instructors. An important consideration is that the complete or partial adoption of any of the learning modules should not increase the time to complete their respective courses.

Alternatively, the flexible learning modules can be attempted in a self-directed online mode. This mode is specifically useful for non-traditional learners, e.g., practitioners interested in upgrading their knowledge and skills while working full time. Indeed, anyone who might be interested in finding out more is welcome to use the freely available learning materials. To demonstrate flexibility, a mixed-mode delivery approach was employed in the initial pilot.

After the initial learning modules were reviewed and became available in late August 2021, they were ready for a trial run in the fall 2021 semester that ended in mid-December. A mixed-mode approach was adopted to launch the modules in affected classes. For example, Module 11 Pitfalls of applying AI to Information Retrieval tasks was introduced to level 6000 Information

Retrieval students during regular class time as an optional extra credit component. Students who chose to attempt the module were given one week to complete it in their own time, outside the classroom environment. All students in the class elected to attempt the module.

During the introduction session in class, students were made aware that the voluntary extra credit exercise made up of two separate but interrelated components: education and research. The former meant attempting to complete the module to earn extra credit. The latter meant additionally completing the anonymous pre-/post-intervention questionnaires developed by the independent evaluator. It was made perfectly clear that whether students completed the research component had no bearing on their grade; it was an extra task they could do to help the research team gain a better understanding. In this instance, the response rate was 75%.

4. On-going and Future Work

Analysis of the raw data (completed questionnaires) is currently underway and preliminary results are encouraging. Raw data are currently being gathered by several faculty across multiple disciplines (Computer Science, Mechanical Engineering, Civil Engineering, Statistics, Business Analytics, etc.) in the current (spring 2022) semester. A clearer picture will emerge with respect to the efficacy of the intervention sometime in summer 2022. Finetuning of existing modules and future developments will be dependent on the findings.

In addition, the project has an integrated outreach component that takes a multiprong approach. Specific tasks include dissemination at good conferences (such as this), train-the-trainers program and publicity activities involving the panel of experts, and outreach to area high schools. A short non-residential workshop that lasts five half days is in the pipeline to engage up to 20 high school students each summer. Known as “Can AI be trusted?”, the workshop exposes high school students to the hype and reality surrounding trustworthy AI systems and technologies. The research team is watching closely at the pandemic situation because an in-person experiential workshop is the preferred approach to fully immerse participants in this exciting venture.

5. Conclusion

This work-in-progress (WIP) paper has presented an NSF-funded project aimed at improving AI literacy among diverse disciplinary learners who likely / actually encounter AI in their work. The work emphasizes on safety, security, and reliability as three key pillars that support trustworthiness in AI systems, algorithms, and technologies. The developed experiential learning modules serve as a vehicle of conveying the key concepts to a broad range of learners. Preliminary results from an independent evaluator show early promise in a pilot run conducted in fall in 2021. A further and broader trial run is currently underway across multiple disciplines and institutions in spring 2022. Key findings that will emerge in summer 2022 will inform further development in this line of research. The aim of this WIP paper is to disseminate the project broadly at a top education conference so as to share this exciting venture with likeminded researchers and educators. Up to date findings will be presented at the conference. All artifacts of the project (learning modules, suggested use cases, summarized/anonymized findings, etc.) are freely available on the project website. The project team welcomes feedback from the broader community beyond the contributing panel of experts on any aspect of the project.

6. References

[1] J. Liu, "Survey of the Image Recognition Based on Deep Learning Network for Autonomous Driving Car," 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT), 2020, pp. 1-6, doi: 10.1109/ISCTT51595.2020.00007.

[2] T. Treebupachatsakul and S. Poomrittigul, "Microorganism Image Recognition based on Deep Learning Application," 2020 International Conference on Electronics, Information, and Communication (ICEIC), 2020, pp. 1-5, doi: 10.1109/ICEIC49074.2020.9051009.

[3] S. Liu and B. Liu, "Application Analysis of Image Enhancement Method in Deep Learning Image Recognition Scene," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), 2021, pp. 1949-1952, doi: 10.1109/ICESC51422.2021.9532597.

[4] R. Zhang, W. Xiao, H. Zhang, Y. Liu, H. Lin and M. Yang, "An Empirical Study on Program Failures of Deep Learning Jobs," 2020 IEEE/ACM 42nd International Conference on Software Engineering (ICSE), 2020, pp. 1159-1170, doi: 10.1145/3377811.3380362.

[5] Ethem Alpaydin, "2 MACHINE LEARNING, STATISTICS, AND DATA ANALYTICS," in Machine Learning , MIT Press, 2021, pp.35-69.

[6] J. Cecil and A. Gupta, "An Experiential Approach to Support Learning of Cyber Physical Systems Concepts involving Mixed Reality Platforms," 2021 IEEE Frontiers in Education Conference (FIE), 2021, pp. 1-6, doi: 10.1109/FIE49875.2021.9637125.

[7] W. -S. Soh, "Experiential Learning Through Remote Electrical Engineering Labs During the COVID-19 Pandemic," 2021 IEEE International Conference on Engineering, Technology & Education (TALE), 2021, pp. 01-05, doi: 10.1109/TALE52509.2021.9678756.

[8] S. Vorapojpisut, "Applying Experiential Learning Cycle for Electrical Measurement Laboratory," 2021 6th International STEM Education Conference (iSTEM-Ed), 2021, pp. 1-4, doi: 10.1109/iSTEM-Ed52129.2021.9625094.