Cybersecurity Education in the Age of Artificial Intelligence: A Novel Proactive and Collaborative Learning Paradigm

Jin Wei-Kocsis
Purdue University
West Lafayette, USA
kocsis0@purdue.edu

Moein Sabounchi
Purdue University
West Lafayette, USA
msabounc@purdue.edu

Baijian Yang
Purdue University
West Lafayette, USA
byang@purdue.edu

Tonglin Zhang
Purdue University
West Lafayette, USA
tlzhang@purdue.edu

2022 IEEE Frontiers in Education Conference (FIE) | 978-1-6654-6244-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/FIE56618.2022.9962643

***Abstract*—This Innovative Practice Work-in-Progress paper presents a virtual, proactive, and collaborative learning paradigm that can engage learners with different backgrounds and enable effective retention and transfer of the multidisciplinary AI-cybersecurity knowledge. While progress has been made to better understand the trustworthiness and security of artificial intelligence (AI) techniques, little has been done to translate this knowledge to education and training. There is a critical need to foster a qualified cybersecurity workforce that understands the usefulness, limitations, and best practices of AI technologies in the cybersecurity domain. To address this import issue, in our proposed learning paradigm, we leverage multidisciplinary expertise in cybersecurity, AI, and statistics to systematically investigate two cohesive research and education goals. First, we develop an immersive learning environment that motivates the students to explore AI/machine learning (ML) development in the context of real-world cybersecurity scenarios by constructing learning models with tangible objects. Second, we design a proactive education paradigm with the use of hackathon activities based on game-based learning, lifelong learning, and social constructivism. The proposed paradigm will benefit a wide range of learners, especially underrepresented students. It will also help the general public understand the security implications of AI. In this paper, we describe our proposed learning paradigm and present our current progress of this ongoing research work. In the current stage, we focus on the first research and education goal and have been leveraging cost-effective Minecraft platform to develop an immersive learning environment where the learners are able to investigate the insights of the emerging AI/ML concepts by constructing related learning modules via interacting with tangible AI/ML building blocks.**

***Index Terms*—Artificial intelligence, machine learning, cybersecurity, education**

I. Introduction

The phenomenal growth of AI techniques, especially ML, impacts every aspect of human life, including autonomous and semi-autonomous security systems that are demonstrating impressive promises for increasing awareness, reacting in real time, and improving the overall effectiveness [1]–[3]. According to VynZ Research, the global AI in cybersecurity market reached USD 12 billion in 2020 and will grow to USD 30.5 billion in 2025 [4]. However, increasing evidence shows that AI techniques can be manipulated, evaded, and misled, which results in new and profound security implications [5], [6]. While prominent research progress has been made in understanding the trust and security of AI techniques [7], there is an education and training gap to foster the qualified cyber-workforce that understands the usefulness, limitations, and best practices of AI technologies in cybersecurity domain. Recent reports also indicate that this education gap will throttle aspirations in the advance of AI and intensify the shortage problem in cybersecurity workforce [8]–[10].

Efforts have been made to incorporate a comprehensive curriculum to meet the demand. However, there still remain essential challenges for effectively educating students on the interaction of AI and cybersecurity including: (1) due to the emerging and growing features of AI technologies and zero-day exploits, the integration of AI and cybersecurity technologies are rapidly and dynamically evolving; (2) students can have very diverse knowledge background, ranging from conventional information technology to data science, and thus may have varied needs for inspiring skill and interaction engagement; and (3) while significant studies have been developed in understanding AI/ML-specific threats, most of the existing research focuses on computer vision domain and very limited efforts have been made in the cybersecurity domain that is complex and rife with adversaries. To address these challenges, in this research, we aim to educate and train a qualified cyber-workforce in this new era where security breaches, privacy violations, and artificial intelligence have become commonplace.

The rest of the paper is organized as follows. Section II introduces background and related work. The proposed learning paradigm and the current progress of our research work are elaborated in details in Section III. The paper is concluded in Section IV with future work highlighted.

II. Background and Related Work

Statistical ML algorithms have been extensively used in the field of cybersecurity, such as spam detection [11], malware detection [12], and network intrusion detection [13]. The rise of deep learning (DL) approaches offers a promising direction

This work is funded by the U.S. National Science Foundation Award #2114974

in discovering sophisticated and unseen attack patterns [14]. While the benefits of DL is immense, the black-box nature of DL casts doubts in the decision making process. More importantly, when this new intelligent component is applied to an existing cybersecurity system, it increases the attack surface and is subjected to additional attacks. A typical machine learning workflow starts from data collection, followed by data pre-processing to clean up the noises, normalize the scales and manage the missing data. Feature selection and feature engineering are often needed before feeding the training data to various machine learning algorithms. Since there is no single model fits all, inference models are evaluated on the validation data and the best model will be selected to predict unseen new data. When the performance deteriorates, model update will be triggered to retrain the model. Every step of the machine learning process is subjected to attacks, as detailed in the next paragraph.

Recently, adversarial attacks on machine learning techniques, especially the deep learning systems, received a lot of attentions from both the AI community and the security community [5], [15], [16]. To attack data acquisition and data prepossessing, various attacks were proposed to mislead the classifier using data poisoning attacks [17], [18]. Likewise, feature selections could be reduced to impair the accuracy of machine learning algorithms [19]. A significant portion of the work in adversarial attacks was targeting the machine learning algorithms themselves. Adversarial perturbations were generated in the direction of the gradient to attack the models, e.g. FGSM [20] and JSMA [21]. Attacks could also be launched to degrade the confidence scores of the classifiers, such as the ZOO attack presented in [22]. Decisions could also be confused if the attackers feed the system with adversarial examples crafted with Generative Adversarial Network (GAN) techniques [23]. The above descried attacks are well discussed in the context of computer vision and natural language processing. More studies are needed to understand how these attacks impact the security of a system. Given that cybersecurity problems are often manifested in the form of binary classifiers, attacks on AI/ML driven security system will be even more dangerous because attacks on one model can be easily transferred to another model in the case of binary classifications [24].

From education perspective, new technologies are increasingly adopted to innovate teaching and learning. Recent developments in visualization and virtual reality (VR) possess great potentials in education and training [25], [26]. Various studies have been developed to apply VR technologies to develop immersive learning environment in different fields, including medicine, engineering, and construction. All of them demonstrated valuable practices of applying VR in education [27], [28]. However, the established work mainly focused on teaching procedural, practical knowledge, and declarative knowledge that can be benefited from realistic surrounding function of VR. In contrast, the AI-cybersecurity conjecture has many abstract concepts and theories. Additionally, existing work in VR often overlooked collaborative learning and social connectivism, which are essential for staying current in rapidly evolving information ecology [29].

In summary, the current research trend on AI and cybersecurity usually focuses on two different themes: the trustworthiness of the AI systems and the application of AI in the cybersecurity domain. While it is wise for the researchers to have their own concentrations, it will be ill-advised to treat AI and cybersecurity as two distinct subjects to educate the next generation scholars. In addition, existing curriculum on integrating AI and cybersecurity are often reactive in nature: lectures and labs are created in response to known attacks. A transformative approach is needed to educate the learners to become proactive thinkers and practitioners. Although innovative technologies, have been adopted in the education domain, it is not yet clear how complex, dynamic and abstract disciplines can benefit from this innovative pedagogy.

III. Proposed Proactive and Collaborative Learning Paradigm and Current Progress

The overview of our proposed virtual, proactive and collaborative learning paradigm, which enables the innovative integrated cybersecurity and AI/ML curriculum, is illustrated in Fig. 1. As shown in Fig. 1, our proposed paradigm

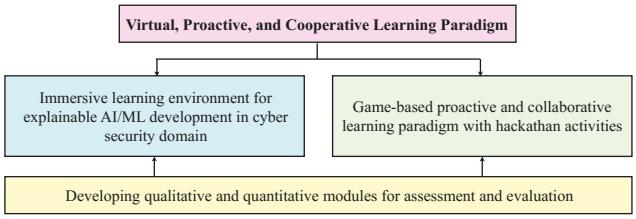

Fig. 1. Overview of our proposed virtual, proactive and collaborative learning paradigm.

mainly consists of two main components: (1) a cost-effective immersive learning environment that motivates the students to explore AI/ML development in the context of real-world cybersecurity scenarios by constructing learning models with tangible objects; and (2) a game-based proactive education paradigm with the use of hackathon activities that engages students with diverse background to collaboratively formulate AI/ML-specific threats and develop trustworthy and robust AI/ML solutions in cybersecurity domain.

In the current stage of research, we have been leveraging cost-effective Minecraft platform to develop an immersive learning environment where the learners are able to investigate the insights of the emerging AI/ML concepts by constructing related learning modules via interacting with tangible AI/ML building blocks. The building blocks are developed to represent primitive units with different degrees of granularity. The interaction between the building blocks throughout the AI/ML workflow is also virtualized to the learners for their individual or collaborative investigation. In our current version of the immersive learning environment, the individual and interactive investigation is supported and the learning modules include

logistic regression, fully-connected neural network, convolutional neural network, recurrent neural network, autoencoder, and generative adversarial networks.

1) Logistic Regression: As illustrated in Fig. 2, in our immersive learning environment that is developed by leveraging MCPI API [30] and PythonTool Mod [31], the learners are able to select a dataset from the multiple available datasets and develop a logistic regression model with tangible building blocks by customizing the hyperparameters of the building blocks and the settings of the learning procedure via the interactive text-based user interface. The selected dataset can be visualized in our proposed learning environment. Additionally, the critical parameters of the logistic regression model during training and testing procedures, such as weights and bias, are also visualized via color scale. Further, to enhance the transparency of the training and testing procedure, the decision boundary during the training and testing can also be visualized in real time. By using the decision boundary visualization, the learners are able to achieve more insights about the impact of different parameters and hyperparameters on the performance of the realized learning model.

Fig. 2. Screenshot of immersive learning environment for developing logistic regression model and explanations.

2) Fully-Connected Neural network (FCNN): As illustrated in Fig. 3, while integrating the development of FCNN model in our learning environment, we also explore another type of user interface where the learners are able to customize building blocks and ML procedure by setting the switches on a control panel instead of typing any text. While enabling the

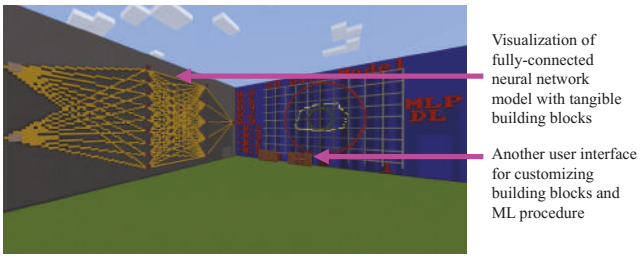

Fig. 3. Screenshot of immersive learning environment for developing FCNN model and explanations.

development of FCNN, we noticed that the immersive learning platform has limitations on supporting multiple learners for collaborative investigation and on supporting the development of ML modules with high complexity and scalability. To address the limitations, we explore alternative strategy to realize immersive learning environment by leveraging flask server, Bukkit server, and Spigot API. The two screenshots of the immersive learning environment using this new strategy with different angles are shown in Fig. 4. In this environment,

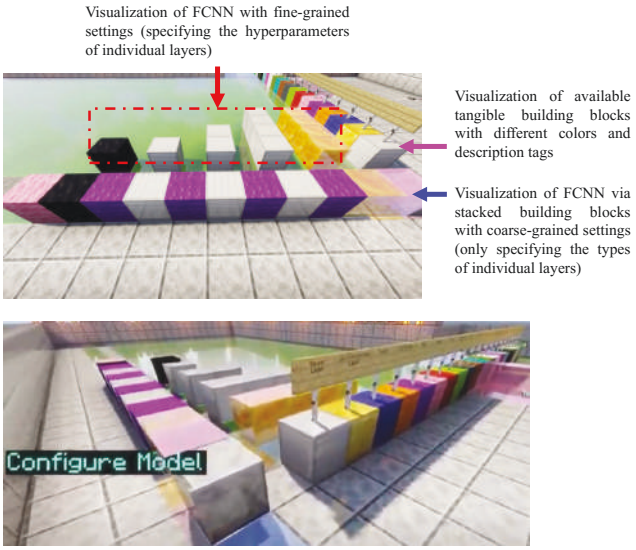

Fig. 4. Screenshots of immersive learning environment with new strategy for developing FCNN model via two different angles.

there are tangible building blocks whose concepts are provided via the associated colors and tags. For example, a building block can be used for realizing a dense layer with sigmoid or softmax activation functions. In this environment, the learners are able to: (1) select the tangible building blocks based on the colors and the tags associated with the available building blocks; (2) interact with the selected tangible building blocks with different granularities; and (3) develop the ML models by stacking the building blocks.

3) Convolutional Neural Network (CNN): As shown in Fig. 5, the representation of immersive learning environment for developing CNN model is very similar to the screenshot in Fig. 4. The main difference is that we integrate the ad-

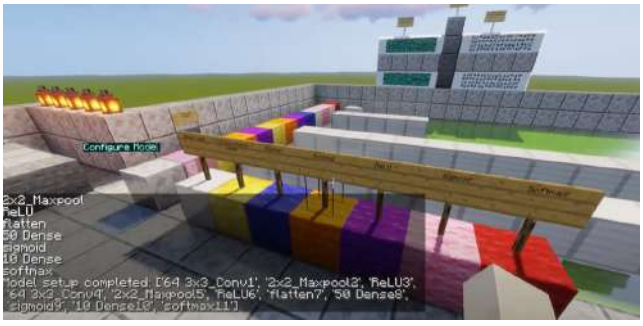

Fig. 5. Screenshot of immersive learning environment for developing CNN model with the outputs of visualization mechanism on the top.

ditional visualization mechanism for enhancing the learners' experience on CNN development. The additional visualization mechanism includes visualizing the kernel weights and feature maps associated with the convolutional layers. Figure 6 shows

a screenshot by zooming in the outputs of the visualization mechanism in Fig. 5.

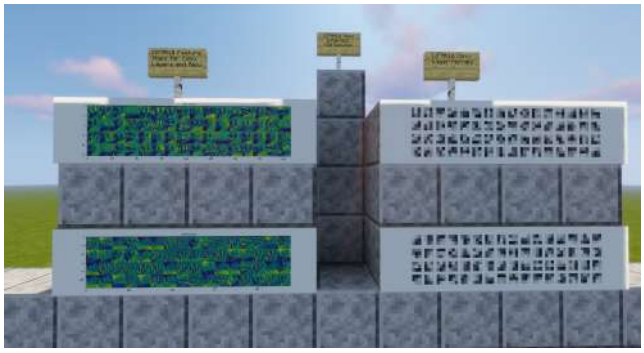

Fig. 6. Screenshot of the outputs of the visualization mechanism of developing CNN for a classification task: Left: visualization of feature maps after given epochs, and Right: visualization of kernel weights after given epochs.

4) Autoencoder: The screenshot of the immersive learning environment for developing autoencoder model is shown in Fig. 7. The outputs of the visualization mechanism, as shown in Fig. 7, include the reconstructed data with different hyperparameters and parameters.

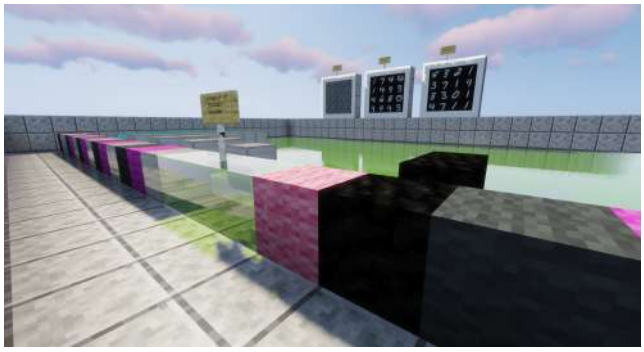

Fig. 7. Screenshot of immersive learning environment for developing an autoencoder model with the outputs of visualization mechanism on the top.

5) Generative Adversarial Network (GAN): The screenshot of the immersive learning environment for developing GAN model is shown in Fig. 8. As shown in Fig. 8, the GAN

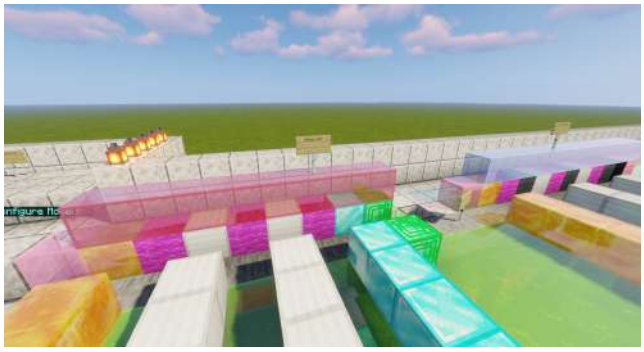

Fig. 8. Screenshot of immersive learning environment for developing GAN model.

model consists of generator associated with the pink glass block and discriminator associated with the blue glass block. The training of GAN is realized by triggering generator and discriminator blocks interactively. As shown in Fig. 9, the outputs of the visualization mechanism include the generated data with different hyperparameters and parameters.

Fig. 9. Screenshot of immersive learning environment for developing GAN model with another angle, where the outputs of the visualization mechanism are shown on the top.

IV. Conclusions and Future Work

The overarching goal of our research is to address a critical need to foster a qualified cybersecurity workforce that understands the usefulness, limitations, and best practices of AI technologies in the cybersecurity domain. To achieve this goal, we leverage multidisciplinary expertise in cybersecurity, AI, and statistics to design and implement a virtual, proactive, and collaborative learning paradigm that can engage learners with different backgrounds and enable effective retention and transfer of the multidisciplinary AI-cybersecurity knowledge. At the current research stage, we have been leveraging Minecraft to develop an immersive learning environment where the learners are able to investigate the insights of the emerging AI/ML concepts by constructing related learning modules via interacting with tangible AI/ML building blocks.

For our future work, we will continue to work on completing the cost-effective immersive learning environment that motivates the students to explore AI/ML development in the context of real-world cybersecurity scenarios by constructing learning models with tangible objects. In addition to add more deep learning models in the learning environment, data representing cyberattacks will be loaded to the environment for learners to investigate and explore. We will also develop a game-based proactive education paradigm with the use of hackathon activities to engage students with diverse background to collaboratively formulate AI/ML-specific threats and develop trustworthy and robust AI/ML solutions in cybersecurity domain.

Acknowledgment

This work was funded by the U.S. National Science Foundation (NSF) collaborative research grant 2114974. We thank Gihan J. Mendis who provided help with the immersive learning environment development.

REFERENCES

[1] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505–2516, 2017.

[2] E. Tsukerman, *Machine Learning for Cybersecurity Cookbook: Over 80 recipes on how to implement machine learning algorithms for building security systems using Python*. Packt Publishing; 1st edition, 2019.

[3] D. Berman, A. Buczak, J. Chavis, and C. Corbett, "A survey of deep learning methods for cyber security," *Information*, vol. 10, no. 4, pp. 1–35, 2019.

[4] V. Research. (2019) Global artificial intelligence (ai) in cyber security market – analysis and forecast (2019-2025). [Online]. Available: https://www.vynzresearch.com/ict-media/artificial-intelligence-in-cyber-security-market

[5] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial machine learning at scale," in *Proceedings of International Conference on Learning Representations (ICLR)*, Toulon, France, April 2017.

[6] "Artificial intelligence and cybersecurity: Opportunities and challenges technical workshop summary report," *A Report by Networking & Information Technology Research and Development Subcommittee and the Machine Learning & Artificial Intelligence Subcommittee of the National Science & Technology Council*, March 2020.

[7] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision: A survey," *IEEE Access*, vol. 6, pp. 14 410–14 430, 2018.

[8] "A report to the president on supporting the growth and sustainment of the nation's cybersecurity workforce: Building the foundation for a more secure american future," *Transmitted by The Secretary of Commerce and The Secretary of Homeland Security*, May 2017.

[9] "Federal cybersecurity research and development strategic plan," *Prepared by the Cyber Security and Information Assurance Interagency Working Group Subcommittee on Networking & Information Technology Research & Development Committee on Science & Technology Enterprise of the National Science & Technology Council*, December 2019.

[10] "U.S. Department of Homeland Security Artificial Intelligence Strategy," December 2020.

[11] K. Tretyakov, "Machine learning techniques in spam filtering," in *Data Mining Problem-oriented Seminar, MTAT*, vol. 3, no. 177. Citeseer, 2004, pp. 60–79.

[12] D. Gavriluţ, M. Cimpoeşu, D. Anton, and L. Ciortuz, "Malware detection using machine learning," in *2009 International Multiconference on Computer Science and Information Technology*. IEEE, 2009, pp. 735–741.

[13] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *2010 IEEE symposium on security and privacy*. IEEE, 2010, pp. 305–316.

[21] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami, "The limitations of deep learning in adversarial settings," in *2016 IEEE European symposium on security and privacy (EuroS&P)*. IEEE, 2016, pp. 372–387.

[14] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, and C. Wang, "Machine learning and deep learning methods for cybersecurity," *IEEE Access*, vol. 6, pp. 35 365–35 381, 2018.

[15] L. Huang, A. D. Joseph, B. Nelson, B. I. Rubinstein, and J. D. Tygar, "Adversarial machine learning," in *Proceedings of the 4th ACM workshop on Security and artificial intelligence*, 2011, pp. 43–58.

[16] W. Brendel, J. Rauber, and M. Bethge, "Decision-based adversarial attacks: Reliable attacks against black-box machine learning models," *arXiv preprint arXiv:1712.04248*, 2017.

[17] S. Alfeld, X. Zhu, and P. Barford, "Data poisoning attacks against autoregressive models," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, 2016.

[18] J. Steinhardt, P. W. W. Koh, and P. S. Liang, "Certified defenses for data poisoning attacks," in *Advances in neural information processing systems*, 2017, pp. 3517–3529.

[19] F. Zhang, P. P. Chan, B. Biggio, D. S. Yeung, and F. Roli, "Adversarial feature selection against evasion attacks," *IEEE transactions on cybernetics*, vol. 46, no. 3, pp. 766–777, 2015.

[20] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[22] P.-Y. Chen, H. Zhang, Y. Sharma, J. Yi, and C.-J. Hsieh, "Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models," in *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*, 2017, pp. 15–26.

[23] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, pp. 2672–2680, 2014.

[24] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," *arXiv preprint arXiv:1312.6199*, 2013.

[25] A. Natale, C. Repetto, G. Riva, and D. Villani, "Immersive virtual reality in k-12 and higher education: A 10-year systematic review of empirical research," *British Journal of Educational Technology*, vol. 51, no. 6, pp. 2006 – 2033, 2020.

[26] J. Radianti, T. Majchrzak, J. Fromm, and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda," *Computer & Education*, vol. 147, pp. 1– 29, 2020.

[27] J. Delgado, L. Oyedele, P. Demian, and T. Beach, "A research agenda for augmented and virtual reality in architecture, engineering and construction," *Advanced Engineering Informatics*, vol. 45, pp. 1 – 21, 2020.

[28] J. Pattle, "Virtual reality and the transformation of medical education," *Future Healthcare Journal*, vol. 6, no. 3, pp. 181 – 185, 2019.

[29] G. Siemens, "Connectivism: A learning theory for the digital age," *International Journal of Instructional Technology & Distance Learning*, vol. 2, pp. 1 – 7, 2004.

[30] "Minecraft: Pi edition API Python Library," *[Available Online:] https://github.com/martinohanlon/mcpi*.

[31] "PythonTool Mod," *[Available Online:] https://ngcm.github.io/PythonTool-Mod*.