Module 4. Machine Learning & AI – Data Science Education for Ethical Use & Active Citizenship

Introduction

Module 4 is designed to equip participants with essential knowledge, critical perspectives, and practical strategies for introducing data science concepts within the context of Artificial Intelligence (AI) and Machine Learning (ML), with a strong emphasis on civic engagement and ethical responsibility. In response to the growing influence of data and AI across different sectors of society—affecting decisions in healthcare, media, education, public policy, and beyond—the module underscores the importance of data science literacy as foundational to informed and engaged citizenship. Participants will explore key ML processes, including data collection/generation, data preparation, AI model training, and result assessment, while also addressing critical issues such as bias identification in data and AI models, and the broader ethical implications of ML on society. They will also be introduced to instructional tools and strategies for integrating these topics into STEAM education, with a focus on enhancing students' data science literacy and promoting civic engagement and social justice.

Learning objectives

It is expected that upon completion of this module participants will be able to:

- Recognize the significance of data science in shaping modern society, particularly within the context of AI.
- Provide concrete examples for illustrating the role of data in ML processes and its influence on AI-driven decision-making processes.
- Evaluate the ethical and social implications of data science and AI, emphasizing issues such as algorithmic bias, privacy, and equity.
- Design STEAM activities that foster critical engagement with AI-driven data practices, promoting data science literacy and ethical awareness among students.

Expected/estimated time for completion: 5 hours

Outline of course structure and activities

Name of activity	Type of activity	Duration
1. Essential reading 1.1 Jiang, S., Nocera, A., Tatar, C., Yoder, M. M., Chao, J., Wiedemann, K., Finzer, W., & Rosé, C. P. (2022). An Empirical Analysis of High School Students' Practices of Modelling with Unstructured Data. *British Journal of Educational Technology, 53*, 1114–1133. https://doi.org/10.1111/bjet.13253 1.2 Olari, V., & Romeike, R. (2021). Addressing AI and data literacy in teacher education: A review of existing educational frameworks. In WiPSCE '21: *The 16th Workshop in Primary and Secondary Computing Education*, *17*. (pp. 1–2).	Pre-requisite reading	1.5 hr
2. Preparatory activity: Before joining class	Hands-on activity on machine learning in preparation for the synchronous session, and reflection on online forum	0.5 hrs
3. Introduction to data-centered STEAM education 3.1. Presentation of key concepts 3.2. Group activity 1: ChatGPT in Action 3.3. Group activity 2: Exploring AI's Role in Education 3.4. Group activity 3: Exploring Code.org Learning Resources	Synchronous session (face-to-face or online)	2.5 hrs
4. Post a short reflection message on online forum	Reflection on online forum	0.5 hrs

Description of the activities

1. Essential reading about STEAM education

1.1. Read 'An Empirical Analysis of High School Students' Practices of Modelling with Unstructured Data' (Jiang et al., 2022)

This article reports on a study which examined student learning of modelling real-world text data. By analyzing the learning process of a group of high school students, the study shows how students used social and cultural knowledge to create predictive features, reasoned about patterns in training data, and managed errors within their models.

1.2. Read 'Addressing AI and data literacy in teacher education: A review of existing educational frameworks.' *(Olari, V., & Romeike, 2021).*

This article argues for an integrated approach to teaching AI and data literacy in K-12, noting that while AI use is increasingly relevant across subjects, current frameworks often overlook data literacy. The authors propose a preliminary model that uses the data lifecycle to highlight essential data literacy competencies for understanding AI.

2. Preparatory activity: before joining class

This hands-on preparatory activity introduces you to machine learning by guiding you in how to build an image classifier using Google's web-based tool Teachable Machine. It helps you understand how data is used to train models and how these models make predictions. It also prompts critical thinking about the use of AI in educational contexts.

Step 1: Exploring Teachable Machine

1. Open the Teachable Machine website on your browser.
2. Once on the homepage, you will see options to create different types of models:
    - Image Project (image classifier)
    - Audio Project (sound classifier)
    - Pose Project (pose classifier)

    For today's activity**, select the Image Project** option, as we will focus on building an image classifier model.

Step 2: Collecting Training Data

1. Choose **two or three categories** that you want your model to classify. These could be:
    - Objects (e.g., "book" vs. "pen")
    - Gestures (e.g., "thumbs up" vs. "thumbs down")

2. In the Teachable Machine app:
   - **Create Your Classes**:
     - Click "Get Started" and create **2-3 classes** (e.g., "Thumbs Up," "Thumbs Down," or "Book," "Pen").
   - **Capture Training Data**:
     - Using your webcam, take **50-100 images for each class.**
     - Make sure to vary your images by changing angles, lighting, and distance to improve the model's ability to generalize.
   - **Train Your Model**:
     - After collecting the images, click "Train Model" to let Teachable Machine process the data.

Step 3: Testing the Model

1. Once your model has been trained, test it by showing it **new examples** that you didn't include during training.
   - For example, if you trained it on hand gestures, try doing the gestures at different angles or lighting.
2. Observe how accurately your model classifies the new examples.
3. Reflect on the results:
   - Did the model perform as expected?
   - Were there any misclassifications?
   - What might have caused these errors (e.g., insufficient variety in the training data)?

3. Machine Learning & AI – Data Science Education for Ethical Use & Active Citizenship (Synchronous Meeting)

The synchronous session, supported by a PowerPoint presentation (PPT), will introduce key machine learning (ML) processes, including data collection and generation, data preparation, AI model training, and result assessment. It will also address critical issues such as bias in data and AI models, along with the broader ethical implications of ML in society. Additionally, participants will explore instructional tools and strategies for integrating these topics into STEAM education, aiming to equip students with essential data science skills for critically engaging with AI/ML and fostering a more informed, participatory, and equitable society.

The following activities will take place during the synchronous (face-to-face or online) meeting of the module.

3.1. Group activity 1: ChatGPT in Action (PPT slide 8)

In groups of 2-3, participants will brainstorm and write a creative or challenging prompt for ChatGPT related to data science or AI. They will use ChatGPT to generate a response and analyze it, noting any surprising elements, inaccuracies, or unclear aspects. Each group will share their prompt and ChatGPT's response with the class. Finally, participants will connect

their observations to concepts from the video, discussing how probability and mathematics influence ChatGPT's responses and identifying factors that may cause inaccuracies or creative leaps.

3.2. Group activity 2: Exploring AI's Role in Education (PPT slide 35)

Participants will split into two teams with opposing viewpoints: Group A will argue that AI enhances equity and fairness in education, while Group B will argue that AI increases inequality and unfairness. Teams will spend 10 minutes collaborating to discuss their assigned perspective and develop clear, well-structured arguments. Each group will then present their arguments to the class, followed by a brief debate to compare perspectives and gain deeper insights into the role of AI in education.

3.3. Group activity 3: Exploring Code.org Learning Resources (PPT slide 42-43)

Participants will select the lesson most relevant to their teaching level—*AI for Oceans* or *How AI Makes Decisions* for primary educators, and *AI for Oceans* or *Bias in the Machine* for secondary educators. In small groups, they will spend 30 minutes exploring the chosen lesson and reflecting on ways to adapt it for their classrooms. The session concludes with a whole-class discussion, where participants share key takeaways, including what they found useful or inspiring and how they might apply similar approaches in their teaching practice.

4. Post a short reflection message on online forum

Upon completing the module activities, take a moment to reflect on the following questions.

(i) Machine learning process: How does the model "learn" from the data, and how does it use training data to make predictions?
(ii) Data diversity: Why is diverse data important, and how does including varied examples in the dataset (e.g., different lighting, angles) improve the model's accuracy?
(iii) Bias in Data: What happens if the dataset is not diverse? How can this impact the model's predictions?
(iv) Privacy: What ethical questions arise when using machine learning in classrooms, especially regarding student data?
(v) Instructional Implications: How would you incorporate data science concepts related to AI/ML into STEAM education, in ways that promote civic engagement and ethical use of AI/ML?

Share your insights by writing a brief post (around 100 words) in the module's online forum, highlighting your key thoughts

Additional resources for further study

The following videos introduce the AI/ML process and key issues related to ML/AI bias, exploring their underlying mechanisms and broader societal impact:

Generative AI: Output & Probabilities
3 types of bias in AI
How AI works: Training Data & Bias
AI Impact on Society

Teachers interested in teaching their students more about AI can consider engaging in the self-paced professional learning module, **Preparing to Teach AI and Machine Learning.**

This module provides teachers with the tools, skills, and knowledge needed to teach the **AI and Machine Learning** unit that is freely available on the Code.org website.

It includes familiarization with AI Lab, a user-friendly programming tool developed by Code.org that lets users create ML Models to solve problems, use patterns in data to make decisions, or test predictions.

Machine Learning for Kids: https://machinelearningforkids.co.uk

Machine Learning for Kids is an interactive tool designed to introduce young learners to the concepts of machine learning in a fun and accessible way. It provides a user-friendly platform for building simple machine learning models, allowing students to train models using their own data and see how AI works in practical applications. By integrating with popular coding environments like Scratch, Python, and App Inventor, it enables hands-on experimentation, fostering an understanding of how machines can recognize patterns and make decisions. This tool is ideal for cultivating curiosity and foundational knowledge in AI among beginners.

Kofteros, A. *Artificial Intelligence: Teaching & Learning with AI*. Published by the Author. Nicosia, 2024. ISBN 978-9925-8055-1-8. Available Online: https://tinyurl.com/aiebook2024

A Creative Commons ebook on Artificial Intelligence, developed to introduce teachers to related technologies and tools, through a series of chapters on its development and its applications in the past few years. A series of tools are introduced that enable the generation of text, sound, presentations, video, along with chapters on Machine Learning.

Zimmermann-Niefield, A., Turner, M., Murphy, B., Kane, S.K., & Shapiro, R. B. (2019). Youth Learning machine learning through building models of athletic moves. In *Proceedings of the*

18th ACM International Conference on Interaction Design and Children (pp. 121-132). Association for Computing Machinery. https://doi.org/10.1145/3311927.3323139

This article investigates how young people without programming experience can use ML in an athletic context to create and explore models of their own physical movements. Using AlpacaML, an iOS app that connects to wearable sensors, middle- and high-school students participated in a workshop where they learned to collect data, build and test ML classifiers, and iteratively refine their models based on their athletic activities.

Srikant, S., & Aggarwal, V. (2017). Introducing data science to school kids. In *Proceedings of the 2017 ACM SIGCSE technical symposium on computer science education* (pp. 561–566). https://doi.org/10.1145/3017680.3017717

The article discusses a hands-on data science tutorial designed for students aged 10-15 to introduce them to the full cycle of supervised machine learning, including data collection, visualization, and model testing. Tailored to minimal prerequisites like basic arithmetic and computer skills, the program emphasized engagement through carefully designed pedagogy and datasets. The pioneering initiative aimed to inspire global educators to incorporate data science into school curricula, leveraging its accessible, interactive framework to build comprehensive courses for future workforce readiness.

Pangrazio, L. & Selwyn, N. (2021) Towards a school-based 'critical data education, *Pedagogy, Culture & Society*, *29*(3), 431-448, https://doi.org/10.1080/14681366.2020.1747527

The article explores the challenges of implementing "critical data education" in schools to help young people understand the role of digital data in their lives, particularly within the data economy and surveillance capitalism. Based on three research projects, it highlights objectives like materializing data processes, understanding their implications, and teaching strategies for data protection. The findings emphasize the need to align educational approaches with students' media use expectations and suggest rethinking traditional cybersafety education to better address technological, ethical, and social dimensions of data literacy.

Fergusson, A., & Pfannkuch, M. (2022). Introducing high school statistics teachers to predictive modelling and APIs using code-driven tools. Statistics Education Research Journal, 21(2), 8–8. https://doi.org/10.52041/serj.v21i2.49

The article outlines the development of a web-based task aimed at teaching high school teachers predictive modeling and APIs through a design-based research approach. During a professional development workshop, six teachers explored predictive modeling via an

informal strategy that incorporated data visualizations, a relevant dataset (movie ratings), and a browser-based platform for executing R code. The study concluded that this approach successfully introduced foundational machine learning concepts, presenting an effective framework for embedding statistical and computational methods in educational settings.

Druga, S., & Ko, A. J. (2021). How do children's perceptions of machine intelligence change when training & coding smart programs? In Interaction Design and Children. June 24–30, 2021, https://doi.org/10.1145/3459990.3460712

The study examines how children perceive machine intelligence after creating and training custom machine learning models. Conducted over four weeks with 52 children aged 7–12, it found that children used scientific methods in training, coding, and testing their smart programs. The experience led to increased skepticism about smart devices, with children attributing agency to programmers rather than devices. Peer debates further influenced this shift. The findings suggest strategies to enhance children's understanding and sense of agency in interacting with AI technologies.

Adisa, I.O., Herro, D., Abimbade, O., & Irgens, G. A. (2024). *Information and Learning Sciences*, *5*(7/8), 513-544. https://doi.org/10.1108/ILS-06-2023-0062

This study, part of a participatory design project, developed pedagogical tools to integrate computational thinking and data science into elementary education. Using a phenomenological case study approach, it analyzed classroom observations, student focus groups, and artifacts to explore how learners engaged with data practices. Findings showed that student engagement improved when data problems were connected to their experiences, though challenges arose in reconciling data patterns with real-world experiences. The research highlights the importance of developing reasoning, visualization, and decision-making skills in data science education at the elementary level.