Women Stem-up

WomenSTEM Up for Good Programme

Topic 3: Gender Inclusive AI

a) *Understanding Bias in AI & Data*

b) *Building Ethical & Inclusive AI*

LINKÖPING UNIVERSITY

UNIVERSITY OF THESSALY

STIMMULI for social change

NTNU Norwegian University of Science and Technology

WOMEN STEM UP

Module objectives

- ✔ Identify and explain how gender bias can emerge in AI systems, including through data collection, labeling, model design, and deployment, using real-world examples from fields such as healthcare and hiring.

- ✔ Apply inclusive design principles by evaluating the impact of diverse training data on AI model performance through hands-on activities, fostering awareness of ethical and gender-inclusive AI development.

Co-funded by the European Union

Understanding Bias in AI & Data: Useful definitions

Bias: a strong feeling in favour of or against one group of people, or one side in an argument, often not based on fair judgement (*Source: Merriam-Webster Dictionary)*

Gender Bias: prejudiced actions or thoughts based on the gender-based perception that women are not equal to men in rights and dignity (*Source: European Institute for Gender Equality)*

Artificial intelligence: the capability of computer systems or algorithms to imitate intelligent human behavior *(Source: Merriam-Webster Dictionary)*

Co-funded by the European Union

Where is AI used?

Machine learning

Machine Learning is a subset of Artificial Intelligence (AI) that focuses on building systems that can learn from data and improve over time without being explicitly programmed.

Examples:

Traditional programming: You write rules to detect spam emails.

Machine learning: You give the system thousands of examples of spam and non-spam emails, and it learns the patterns that distinguish them.

NLP stands for Natural Language Processing. It's a subfield of artificial intelligence (AI) that focuses on enabling computers to understand, interpret, generate, and respond to human language in a way that is both meaningful and useful.

Examples: chatgpt, Siri, Alexa, Google Translate

NLP combines linguistics (grammar, syntax, semantics), machine learning (training models on large datasets) and deep learning (e.g. translation , text generation)

Co-funded by the European Union

Computer Vision

Computer Vision is a field of Artificial Intelligence (AI) that enables computers to interpret and understand the visual world. Using digital images from cameras and videos, and deep learning models, machines can identify objects, classify images, and react to what they “see.”

Examples: Image Classification: Identifying what object is in an image (e.g., cat vs. dog)

Object Detection: Locating multiple objects in an image and labeling them

Facial Recognition: Identifying or verifying a person from a digital image

Image Segmentation: Dividing an image into parts for easier analysis

Action Recognition: Understanding movements in video (e.g., walking, waving)

Co-funded by the European Union

- *Who develops AI and what kind of data it is trained on, has gender implications for AI-powered solutions*

- *Can you think of what type of gender implications are there in AI-powered solutions? Think of fields such healthcare and/or hiring practices in any field.*

Co-funded by the European Union

How do AI systems become biased?

Bias in AI systems often stems from the data used to train them, the assumptions built into their models, and the environments in which they are deployed."

— Mehrabi et al., A Survey on Bias and Fairness in Machine Learning, ACM Computing Surveys, 2022.

Co-funded by the European Union

AI bias in action

- A facial recognition system trained mostly on lighter-skinned faces may perform poorly on darker-skinned individuals.
- If annotators consistently label assertive speech from women as “aggressive,” the AI may learn to associate gender with tone unfairly.
- A hiring algorithm might favor candidates from certain universities if it overemphasizes that feature.

Co-funded by the European Union

Building Ethical & Inclusive AI

Activity: Exploring AI and Data Diversity with Image Classification

- Objective: To help students understand the importance of diverse training data in AI models through a hands-on, non-coding activity.

Materials Needed:

Computers or tablets with internet access

Access to an online AI tool (e.g., Teachable Machine by Google)

A variety of images (can be collected from the internet or taken with a camera)

Paper and markers for brainstorming and reflection

Co-funded by the European Union

Building Ethical & Inclusive AI

Steps:

Introduction to AI and Bias:

1. **Begin with a discussion** on what AI is and how it is used in everyday life.

Explain the concept of bias in AI and give examples of how it can affect outcomes.

Data Collection:

2. Divide students into groups and assign each group a category (e.g., animals, vehicles, fruits).

Ask each group to collect images from diverse sources (different websites, personal photos, etc.) ensuring variety in terms of background, lighting, and angles.

Using Teachable Machine:

3. Introduce students to Teachable Machine (https://teachablemachine.withgoogle.com/), a user-friendly tool that allows anyone to train a simple AI model without coding.

Guide students through the process of uploading their collected images to Teachable Machine.

Co-funded by the European Union

Building Ethical & Inclusive AI

4.Training the Model:

Show students how to label their images and train the model using Teachable Machine.

Allow students to experiment with different sets of images to see how the diversity of data affects the model's accuracy.

5. Evaluating the Model:

Once the model is trained, test it with new images to see how well it performs.

Discuss the results and highlight the importance of having diverse training data to improve the model's accuracy and fairness.

6. Reflection and Discussion:

Have students reflect on the activity and discuss what they learned about AI and bias.

Encourage them to think about how they can apply these principles in future projects.

7. Brainstorming Session:

Use paper and markers to brainstorm ideas on how to ensure diversity in data collection for different AI applications.

Create posters or presentations to share their ideas with the class.

Co-funded by the European Union

References

1. Merriam-Webster Dictionary https://www.merriam-webster.com/
2. Glossary & Thesaurus, European Institute for Gender Equality: https://eige.europa.eu/publications-resources/thesaurus/overview
3. Ricci Lara, M.A., Echeveste, R. & Ferrante, E. Addressing fairness in artificial intelligence for medical imaging. *Nat Commun* **13**, 4581 (2022). https://doi.org/10.1038/s41467-022-32186-3
4. Heidt, A. (2025, April 7). AI for research: The ultimate guide to choosing the right tool. Nature. https://www.nature.com/articles/d41586-025-01069-0
5. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). *A Survey on Bias and Fairness in Machine Learning*. arXiv. https://doi.org/10.48550/arXiv.1908.09635

Other Resources

Global Gender Gap Report 2023
https://www.weforum.org/publications/global-gender-gap-report-2023/