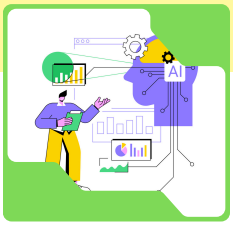




# BIAS: Genre matters

**AI exploration phase**

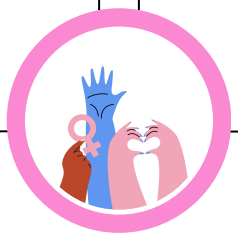
P2 - Building Hypothesis on AI

**AI and Exploration Topics.** What are we exploring?

The **BIAS gap**: why large language models make assumptions about genre
**SDG 5**: avoid discrimination in prompting phase
**Practical approach**: understand by doing, testing prompting tecniques

**Materials Needed and to Create.** What must be prepared before the activity?

Analisys sheets (paper or digital) – *"The BIAS Gap Analisys Sheet"*
Pens / tablets / smartphones for note-taking
Whiteboard or projector for collective analysis



**Exploration Steps and Description.** What do participants actually do?

**Step 1 – The Power of Parity: SDG 5 (Classroom – 30 minutes)**
**Key Question: Why is gender equality a fundamental human right and a necessity for a sustainable world?**

**Suggested Facilitation Approach**: Open with a discussion on "**Hidden Barriers**." Ask students if they believe everyone has the same starting line in life.
**Introduce SDG 5**: Gender Equality as a global "to-do list" to remove those barriers.

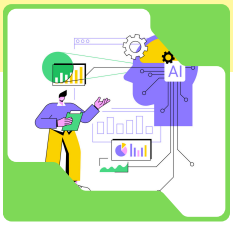
**Focus Areas**:

- **Target 5.3 (Safety & Dignity)**: Discuss the elimination of harmful practices like early/forced marriage and FGM. How do these practices strip away a person's future?
- **Target 5.a (Economic Power)**: Explore why owning land, inheriting property, and accessing bank accounts are the keys to true independence.

**Activity: The Rights Audit (20 min)**: Students work in pairs to research a specific country's laws regarding property and marriage.

- **Goal**: Identify one legal "gap" where women do not have the same rights as men.
- **Fundamental Clarification**: Empowerment isn't just a feeling; it's backed by legal and economic infrastructure.

Co-funded by the European Union






# BIAS: genre matters

**Exploration Steps and Description.** What do participants actually do?

**Step 2 – The Mirror of Bias: Training LLMs (Classroom – 60 minutes)**
**Key Question**: If an AI learns from the internet, does it learn our progress or our prejudices?

**Non-Technical Explanation**: **Large Language Models (LLMs)** are trained on massive datasets. If those datasets contain historical gender gaps or stereotypes, the AI "assumes" those patterns are universal truths.

**Activity**: **The BIAS Gap Analysis (40 min)**: Distribute the "BIAS Gap Analysis Sheet." Students input specific prompts into an AI and analyze the output:

1. The Prompt: "**Write a story about a doctor and a nurse**."
   - Check: Did the AI assign genders automatically?
2. The Prompt: "**List 10 qualities of a strong leader**."
   - Check: Are the traits coded as "masculine" or "feminine"?
3. The Hunt: Students find **three more examples** where the AI defaults to a gender stereotype.

**Goal**: Understand that an LLM is a reflection of the "**Data Gap**"—it doesn't know what is fair, only what is frequent.

**Step 3 – Binary Classification: The NN Experiment (Lab – 45 minutes)**
**Key Question: Can a machine accurately define "Man" or "Woman" based only on two numbers?**

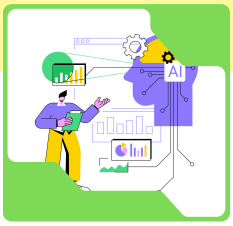
The Technical Framework: Introduce a **simple Neural Network (NN)** designed for **binary classification**. The goal is to predict gender based on Height and Weight.

The Program Logic:

- **Inputs**: Height and Weight.
- **The Math**: Uses a **perceptron** multilayer neural network to learn by data.
- **The Conflict**: Discuss the limitations. What happens to outliers? What about **athletes**?
- **Activity**: Students run a pre-written script (Python/Colab) and observe the "**Decision Boundary**."

**Discuss**: If we only use physical data, what "**human**" elements of gender are we ignoring? Is a binary classification enough to represent SDG 5's inclusive goals?

Co-funded by the European Union






# BIAS: genre matters

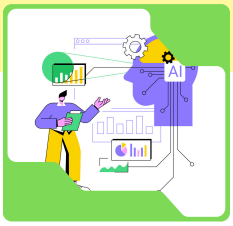
**Exploration Steps and Description.** What do participants actually do?

**Step 4 – Recap & Reflection: The Sticky Note Game (15 minutes)**
**Key Question: What did we learn about the intersection of human rights and machine logic?**

**The Activity:** Every student receives three sticky notes to post on the classroom exit board:

- **Plus** (+): What worked well today? What was your "Aha!" moment regarding AI or SDG 5?
- **Delta**: What should we change? Was the math too hard? Was the bias discussion too short?
- **Dot**: Your satisfaction rating. Place a dot on a 1–4 scale (1 = Confused, 4 = Empowered).

Co-funded by the European Union





# BIAS: genre matters

**Expected Output.** What do participants create, produce or gain from the exploration?

- Completed "**BIAS Gap Analysis Sheet**" documenting real-world examples of gender-coded language and data skews in LLMs.
- A **trained Binary Classification Model** (Neural Network) demonstrating the mathematical logic behind **gender prediction** and its inherent **physical limitation**s.
- Ability to explain the **legal and economic barriers** identified in SDG 5.3 and 5.a (e.g., land rights vs. harmful practices).
- Trace how a "**Data Gap**" in a training set translates into automated **social prejudice**.
- Practical insight into the role of an **ethical creator**: moving beyond "**default**" settings to build more inclusive technology.

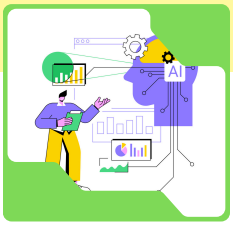
**The Hook and the Playfulness.** What makes this fun/exciting for young people?

- **Detective Mode (The Bias Hunt)**: Students become "Digital Detectives," using the **BIAS Gap Analysis Sheet** to catch world-class AI making embarrassing or unfair assumptions.
- **The Gender Decoder**: Turning a complex Neural Network into a game of "**Height vs. Weight**" to see if a machine can actually "guess" a human identity based only on a few numbers.
- **Global Justice League:** Students don't just study history; they act as "**Global Reformers**" by applying SDG 5 to fix real-world economic and legal inequalities.
- **The Sticky Note Feedback Loop:** A fast-paced, interactive finale where their opinions—the Plus, Delta, and Dot—directly shape how the "lab" runs next time.
- **Breaking the Binary**: The excitement of discovering the "Human Edge"—finding the complex parts of gender and identity that a simple 0 or 1 in code can never fully capture.

**Success indicators.** How do we know it worked? What shows participants learned?

- Critically **identify specific gender biases** in AI outputs and trace them back to the "Data Gap" in Large Language Models.
- **Connect social justice to data**: Articulate how SDG 5.3 and 5.a (land rights and safety) provide the necessary legal "dataset" for a fair society.
- **Evaluate the "Binary Trap"**: Discuss why reducing human identity to height and weight measurements fails to capture the complexity of gender empowerment.
- **Synthesize ethical solutions**: Propose specific ways to "de-bias" an algorithm or improve a dataset to better reflect global equality goals.
- Ability to reformulate the key message: "**AI is a mirror of our past data, but through the lens of SDG 5, we can build tools that reflect a more equal future**."

Co-funded by the European Union





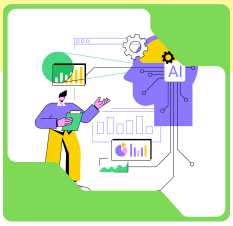
# Appendix - Example of material (duplicate if needed)

**Appendix 1**

Step 1 - Gender Equality

**Content of the appendix**


Co-funded by the European Union

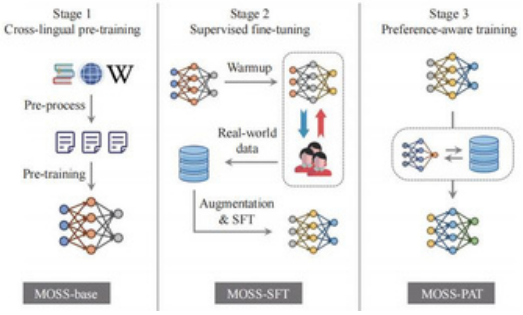


# Appendix - Example of material (duplicate if needed)

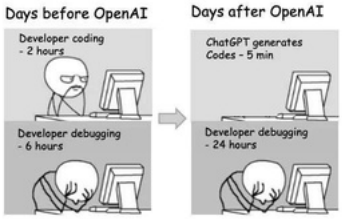
**Appendix 1**

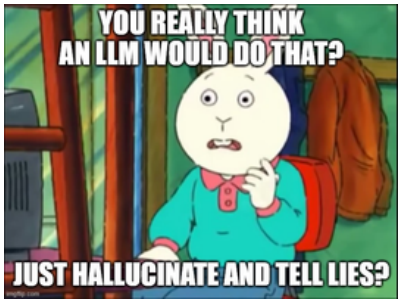
Step 2 - BIAS in LLMs

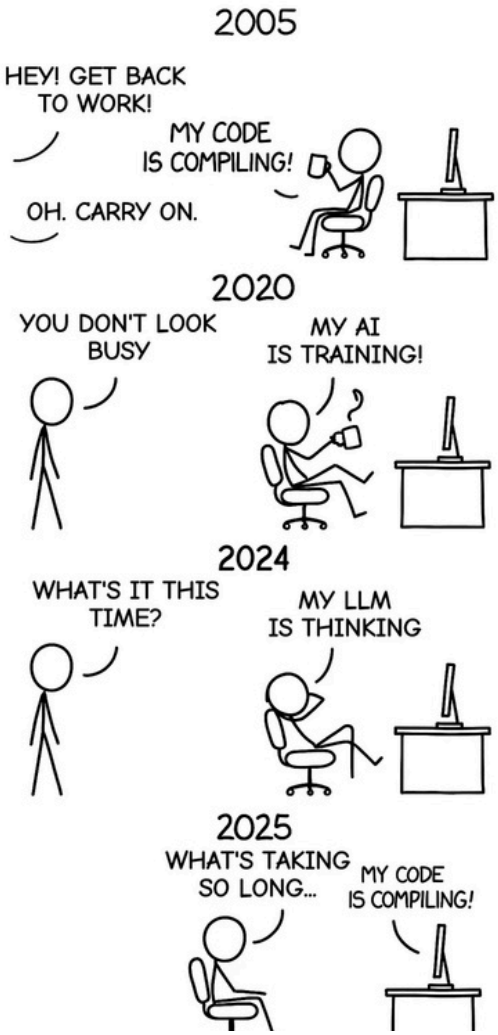
## Content of the appendix



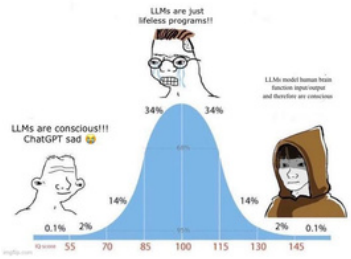




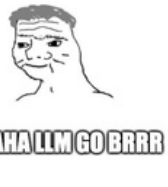



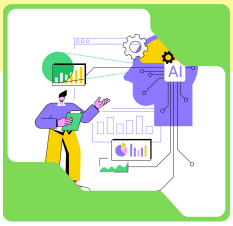









Co-funded by the European Union
L.A.B
UAB Universitat Autònoma de Barcelona
Perlatecnica Research and Technological Education ETS
DIGITALE WOLVEN





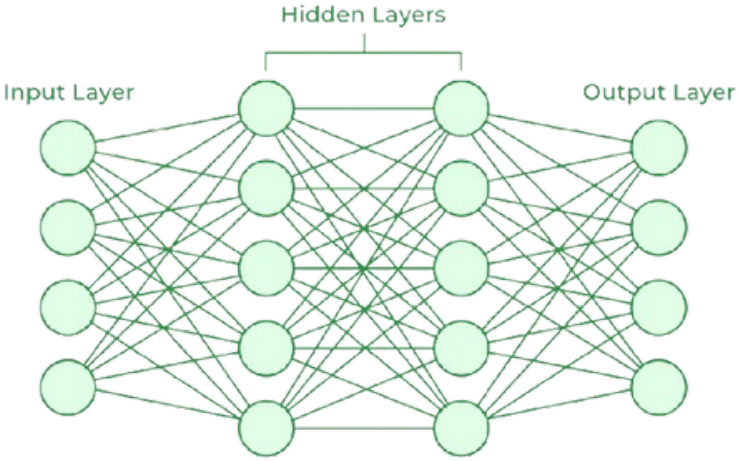
# Appendix - Example of material (duplicate if needed)

**Appendix 1**

Step 3 - Neural Networks: Classifications

## Content of the appendix

How Neural Network works?
Neurons:


People telling me AI is going to destroy the world

My neural network