

DEVELOPING STRATEGIES TO **MINIMIZE GENDER** **BIAS IN AI LANGUAGE** **MODELS**

MARENOSTRUM5 ACC

Led by



C. TORRES
Leader (PI)



A. FREIRE
PhD student

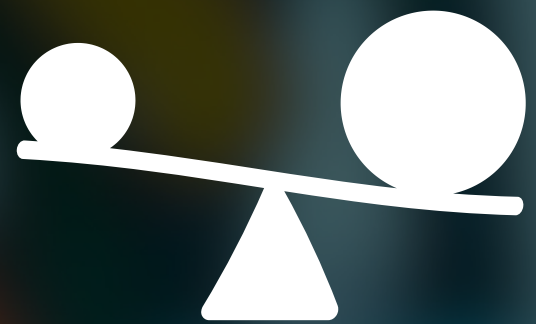
Universidad Técnica Federico Santa María (Chile)

1 IS AI REPRODUCING GENDER STEREOTYPES?



Artificial Intelligence (AI) learns from massive amounts of **human-written text**

These texts often contain **social biases and stereotypes**, especially about **gender**



As AI is increasingly used in recruitment and decision-making, **these biases can be amplified**

Understanding how bias is encoded in AI and **how to reduce it** is key to building **fairer AI systems**



HOW CAN WE MINIMIZE GENDER BIAS? 2

Some stereotypes are **deeply embedded** in the relationships AI learns from language



For example, models may **associate some professions** more strongly **with one gender** than another

To reduce bias, these relationships must be **rebalanced during the learning process**



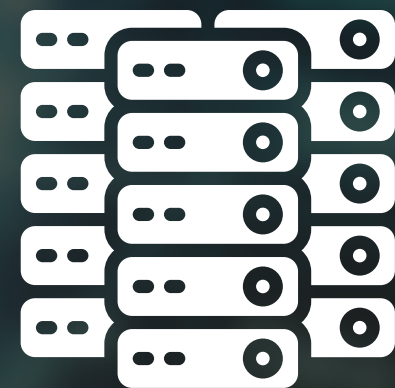
The challenge is to minimize bias while preserving **the model's ability to understand language**

3 THE ROLE OF HPC RESOURCES



To **evaluate different bias mitigation strategies**, models must be trained and tested under **multiple conditions**, requiring to process a huge amount of texts

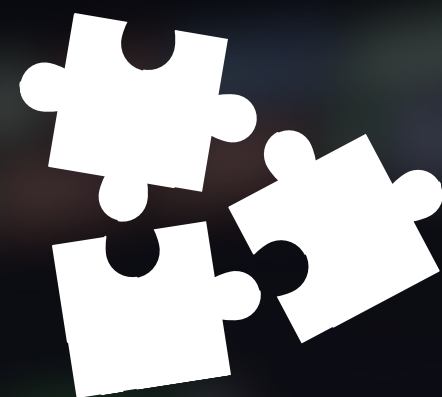
Such experiments require large-scale **GPU computing**, high **memory capacity**, and extensive **storage resources**



HPC infrastructure is essential for carrying out this type of research

RESULTS ON GENDER BIAS MITIGATION 4

Bias can be reduced, but there is **no single solution that works in all cases**



Different mitigation strategies show **different effectiveness** depending on **when they are applied**

Some methods act **during training**, while others are applied **later during prediction or model use**



The study provided a **comparative framework for evaluating bias mitigation strategies** across different AI language models



INFRAESTRUCTURAS CIENTÍFICAS
Y TÉCNICAS SINGULARES



EuroCC
SPAIN RES



Barcelona
Supercomputing
Center

Centro Nacional de Supercomputación



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA