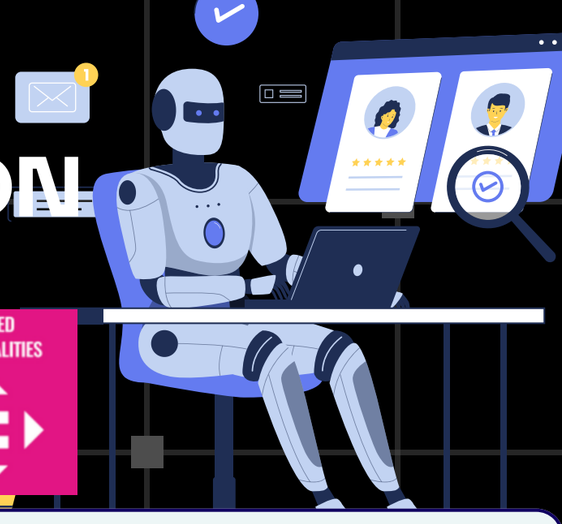
UTILISING SYNTHETIC DATA FROM LLM FOR DATA BIAS DETECTION AND MITIGATION IN RECRUITMENT SYSTEMS

DONGHYEOK LEE
MSC. HUMAN CENTRED AI
TECHNOLOGICAL UNIVERSITY DUBLIN



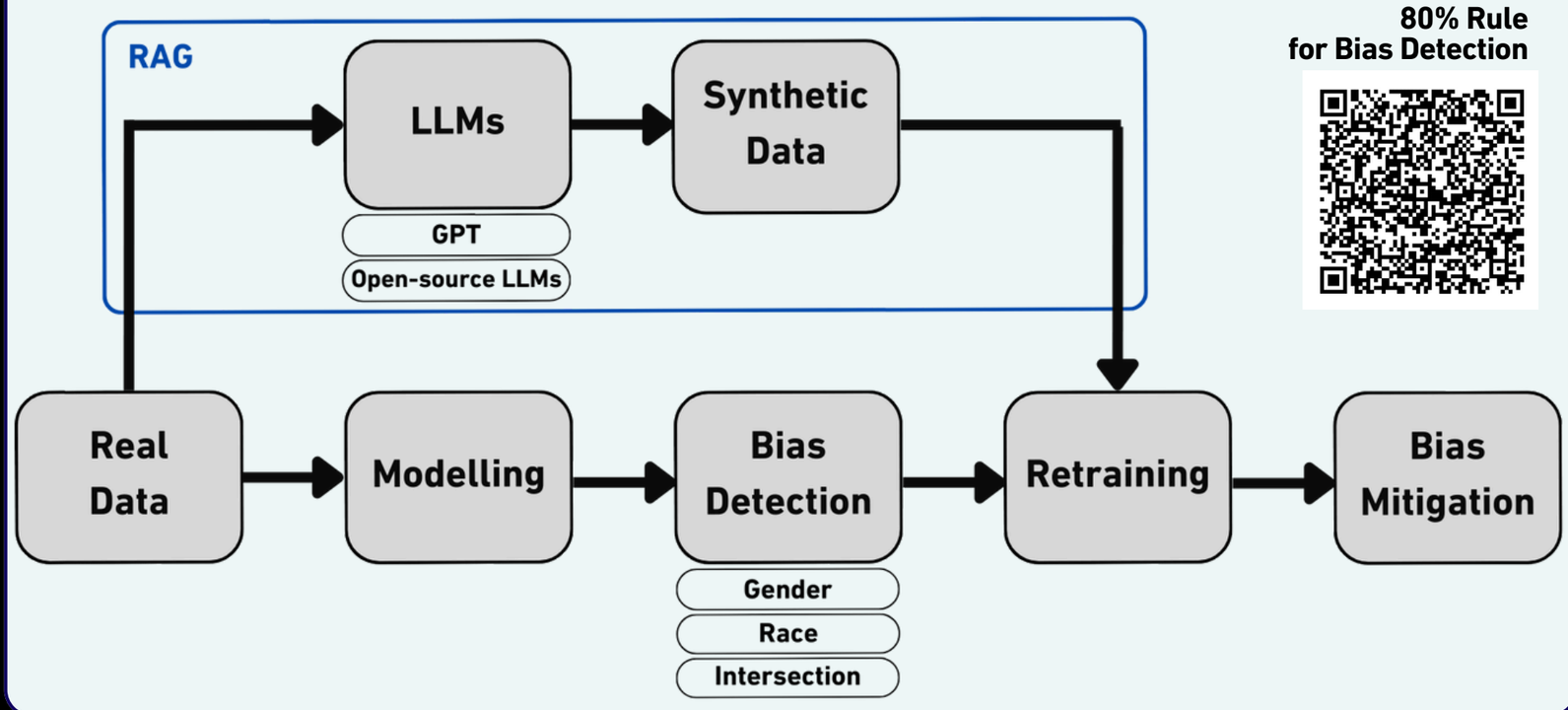
ABSTRACT

In the current landscape, **diversity** and **inclusion** are highly emphasised. This research **develops a methodology to identify and mitigate biases in AI recruitment systems**. **EU AI Act** classifies **AI recruitment systems as high-risk areas**, imposing numerous limitations on the use of real customer data to effectively address biases within these systems (**GDPR**). The process includes identifying biases, generating synthetic data with GPT, mitigating these biases, and comparing results before and after mitigation. Using GPT, the research **creates high-quality, diverse synthetic data to retrain AI systems**, addressing ethical and privacy concerns of using real data. **The goal is to establish a comprehensive framework for ensuring fairness and ethical practices in AI-driven recruitment, promoting a diverse and inclusive employment environment.**

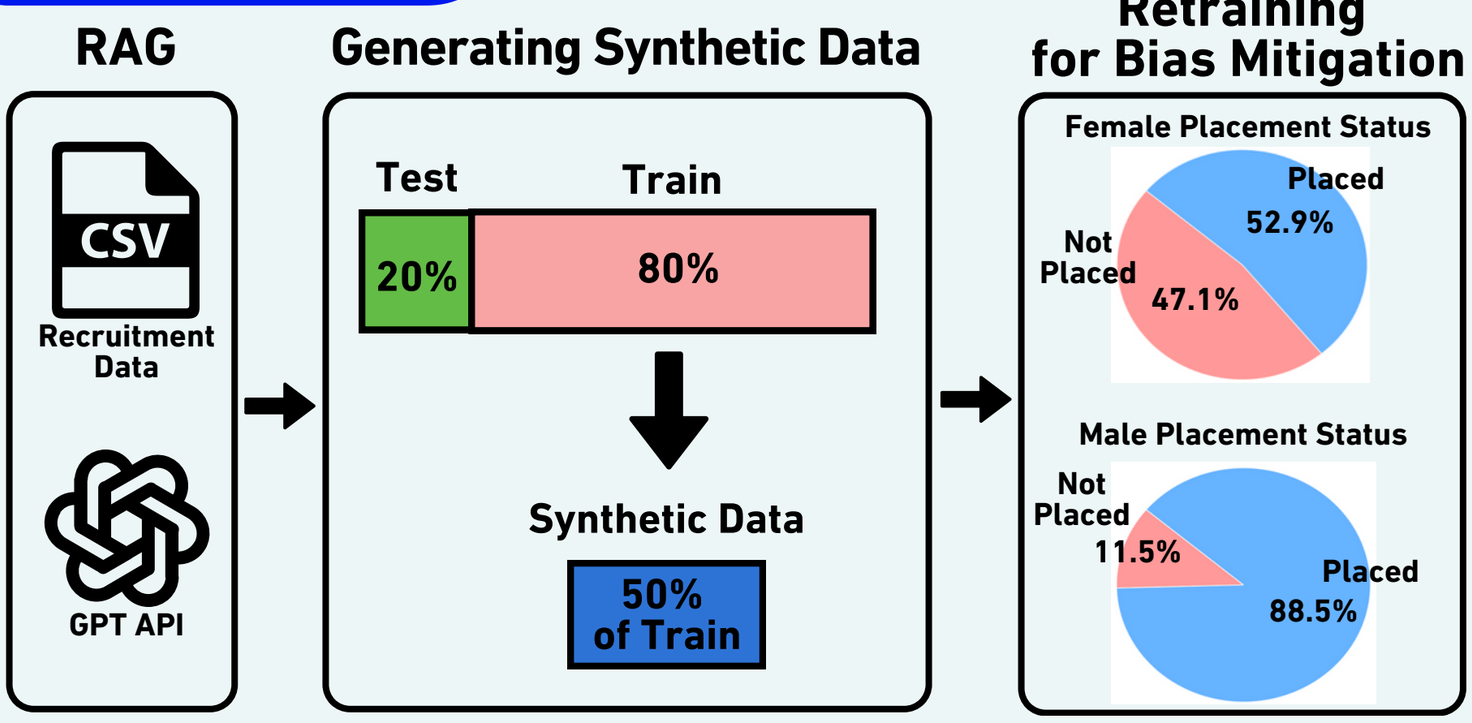
INTRODUCTION

As the **use of synthetic data**, free from data privacy issues, **emerges as a hot topic**, this research **focuses on developing a de-biasing pipeline for recruitment**, where **diversity and inclusion are particularly emphasised**. When biases are detected in existing models, this pipeline can **quickly** and **cost-effectively reduce those biases**. It enables **data privacy protection**, reduces data collection costs, and allows for **rapid bias resolution.** This approach utilises LLMs for bias mitigation.

DE-BIASING PIPELINE



METHODOLOGY



RESULT

Before retraining

	Testset Count	Placed Count	Placed Rate	80% Rule	Bias Detection
Female	17	9	52.94%	70.77%	✓
Male	26	23	88.46%		
Total	43	32			

After retraining

	Testset Count	Placed Count	Placed Rate	80% Rule	Bias Detection
Female	17	13	76.47%	70.77%	
Male	26	23	88.46%		
Total	43	36			

CONCLUSION

The development of a de-biasing pipeline using synthetic data and LLMs represents a significant **step forward in creating fairer AI systems**. By addressing biases in recruitment, we can **promote a more diverse and inclusive employment environment.** This research not only **contributes to the academic understanding of bias mitigation** but also **provides practical solutions** for organizations striving for ethical AI.

IMPROVEMENTS

- Generate synthetic data considering correlations in real data. **(High Quality Synthetic Data)**
- Utilise advanced prompt engineering techniques. **(Accurate Synthetic Data)**
- Compare the results with open-source LLMs. **(LLM Performance Comparison)**
- Identify better real data that includes race column.

Acknowledgements. The HCAIM (The Human-Centred AI Master's Programme) Project Is Co-Financed By The Connecting Europe Facility Of The European Union Under Grant №CEF-TC-2020- 1 Digital Skills 2020-EU-IA-0068. This Research Was Created As Part Of The HCAIM Project