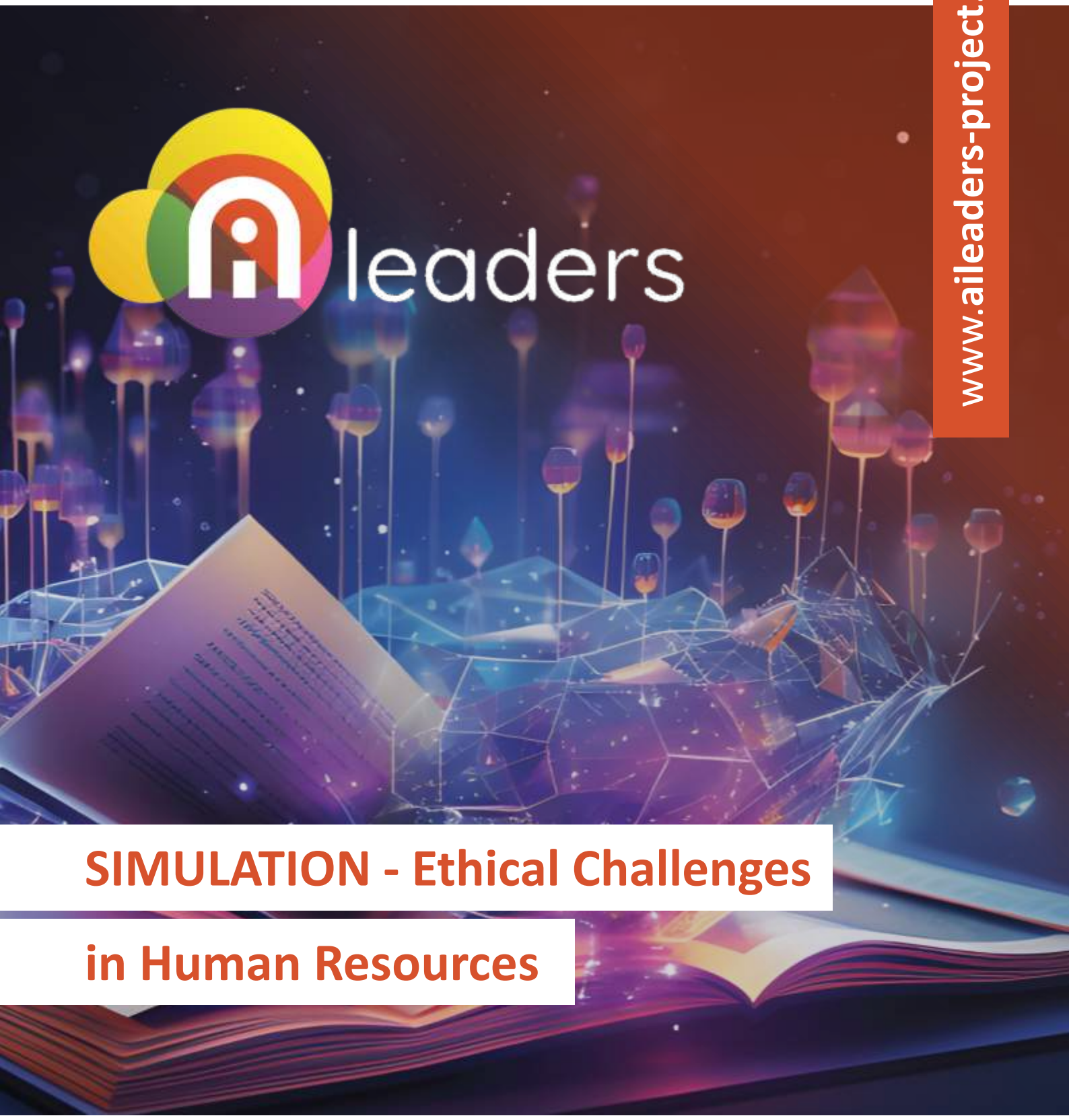
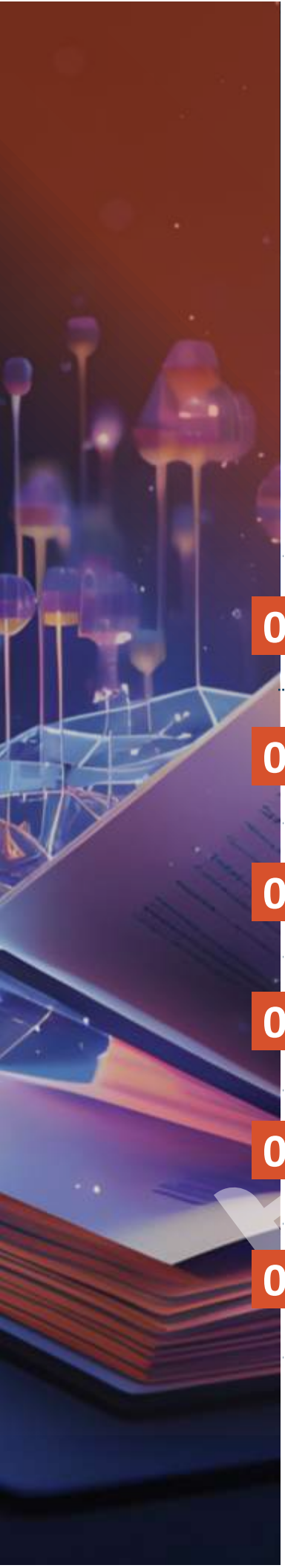
www.aileaders-project.eu



SIMULATION - Ethical Challenges in Human Resources

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



SIMULATION - Ethical Challenges in Human Resources

01  Abstract 3

02  Introduction 4

03  Tools Presentation 6

04  Simulation Execution 7

05  Conclusions 8

06  References 10

01 Abstract

Type of OER

Demo/Simulation using Google Colab (RandomForest vs FairGBM Classifier)

Goal/Purpose

To provide students with a practical and critical exploration of how algorithmic decision-making in recruitment can reproduce structural inequalities by comparing outputs from fairness metrics. This simulation encourages reflection on the ethical dimensions of AI in Human Resources and promotes the development of fairer, more inclusive predictive models.

Expected Learning Outcomes:

By the end of the simulation, students will be able to:

01 Detect and interpret algorithmic bias in AI-driven hiring;

02 Reflect on the ethical implications of automated recruitment systems

03 Use fairness metrics to evaluate model outcomes;

Keywords:

- Machine Learning
- Classification
- Human-Resources
- Biases
- Fairness

Suggested Methodological Approach:

Problem-Based Learning

02 Introduction

Glossary

01 **Accountability** means the organization is responsible for its hiring decisions and their outcomes, ensuring consequences exist for unfair practices.

02 **Transparency** involves making the recruitment process and selection criteria clear and understandable to all candidates.

03 **Fairness** ensures that all applicants are assessed impartially based on their qualifications for the role, free from any bias or discrimination.

04 **Social bias** in hiring is the unfair tendency to favor or reject candidates based on prejudiced assumptions about their social group (like age, gender, or race) instead of their actual skills and qualifications.

Recruitment processes are increasingly expected to be supported by Artificial Intelligence (AI) technologies, particularly in the early stages of hiring - such as CV screening, candidate ranking, and shortlisting applicants for interviews. While these systems offer the potential for greater efficiency and scalability, they also raise important ethical concerns, particularly regarding fairness and bias (Horodyski, 2023).

Long before the adoption of AI, research into traditional recruitment and selection practices had already documented persistent patterns of discrimination (Breaugh, 2013; Hebl et al., 2020). These include biases related to race, ethnicity, and minority status (e.g., Allen & Vardaman, 2017; Hiemstra et al., 2013; Veit & Thijsen, 2019; Zschirnt & Ruedin, 2016); gender (e.g., Ellemers, 2018); social class (e.g., Henderson, 2018); age (e.g., Czopp et al., 2015; Zaniboni et al., 2019); and educational background (e.g., Daly et al., 2000).

Recruitment processes are increasingly expected to be supported by Artificial Intelligence (AI) technologies, particularly in the early stages of hiring—such as CV screening, candidate ranking, and shortlisting applicants for interviews, which can raise important ethical concerns, particularly regarding fairness and bias (Horodyski, 2023).

Long before the adoption of AI, research into traditional recruitment and selection practices had already documented persistent patterns of discrimination (Hebl et al., 2020). These include biases related to race, age, ethnicity, and minority status (e.g., Allen & Vardaman, 2017; Veit & Thijsen, 2019; Zaniboni et al., 2019).

The concern now is that AI technologies, rather than mitigating these biases, may entrench or even amplify them. Emerging studies suggest that algorithmic decision-making in human resources can replicate historical inequities embedded in the data used to train such systems (e.g., Rigotti & Fosh-Villaronga, 2024; Seppälä & Małecka, 2024). As a result, the use of AI in recruitment raises critical questions about accountability, transparency, and fairness in hiring practices.

In this context, the present simulation exercise is designed to engage students with the ethical challenges of AI-driven recruitment and candidate screening. The objective is to provide hands-on experience in identifying potential biases within hiring algorithms and exploring fairness-aware strategies to support more inclusive and equitable decision-making.

Through the use of fairness metrics, students will analyse whether machine learning models used to predict shortlisting decisions reflect—or reproduce—historical patterns of discrimination, with particular attention to ethnicity, gender, and socio-economic background.

By confronting these challenges in a controlled, simulated environment, students will develop both the technical competence and ethical sensitivity essential for responsible AI design and deployment in human resource management.

03 Tools Presentation

This simulation addresses a binary classification task in AI-based recruitment: predicting whether a candidate should be shortlisted for a mid-level management role based on their profile.

The dataset used is synthetic but modelled on real-world recruitment data, publicly available via Kaggle Platform. Candidate attributes includes Gender, Race/Ethnicity, Education Level, Years of Experience, Salary Expectations, Recruitment Source

Although the dataset is artificial, it realistically simulates common hiring decisions and patterns, making it ideal for studying fairness in machine learning. Of particular concern are the ways in which sensitive variables such race/ethnicity may influence model predictions. Studies have shown that recruitment processes may systematically disadvantage candidates from certain ethnic backgrounds, even when qualifications are equal (Zschirnt & Ruedin, 2016; Veit & Thijsen, 2019; Hiemstra et al., 2013).

This simulation focuses on one **main question**: *Do ML algorithms disproportionately "recommend" the rejection of job applications submitted by Black and Women?*

By investigating this question, students critically assess whether AI models used in candidate screening exhibit unintended bias—and how these patterns relate to empirical findings in academic literature.

04 Simulation Execution

01 Access the Simulation Notebook

Go to https://tinyurl.com/k63793wp

02 Run the Code

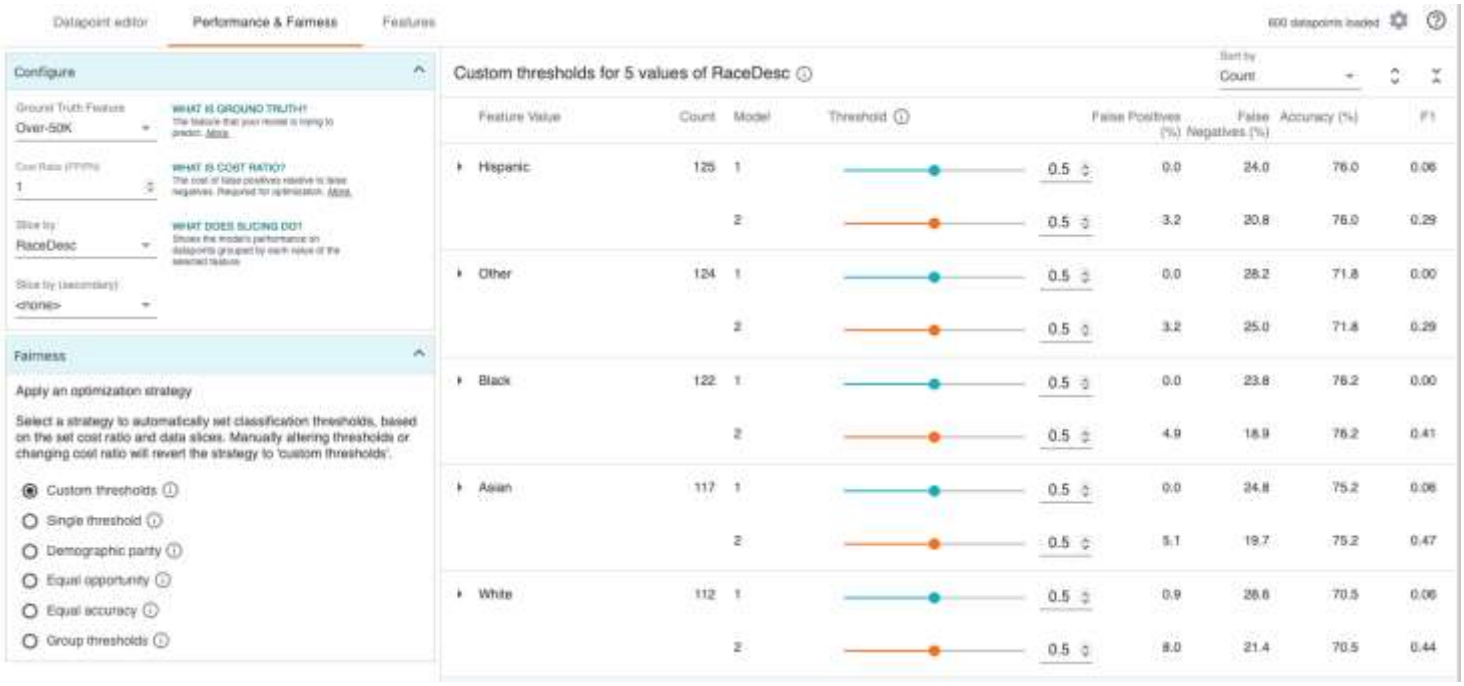
- To execute all the cells, click on "Runtime" tab and select "Run All". Wait some time (1 min).
- Ensure that outputs load correctly and that all models are successfully trained.

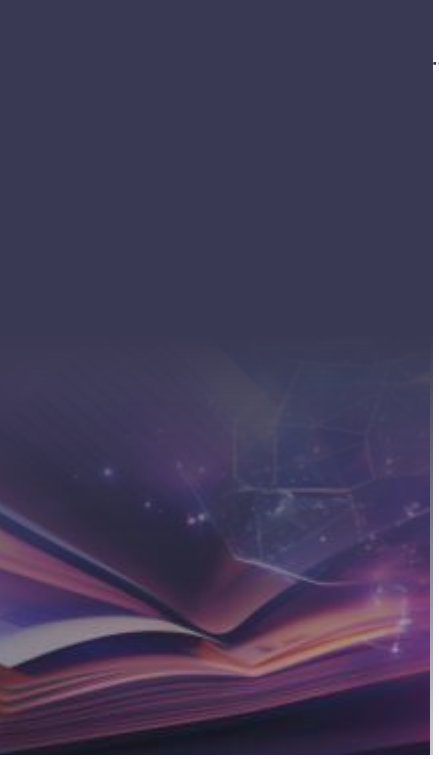
03 Explore the Dataset

- On the What If Tool click on "Performance and Metrics" and then select "Over 50k" feature on the "Ground Truth Feature" and choose "GenderCode" for the "Slice by"
- Review the dataset and inspect key variables. Pay particular attention to demographic features such as race/ethnicity and gender, and their potential correlation with shortlisting outcomes.

04 Analyse Fairness

Apply fairness metrics to analyse differences in rejection accuracy rates across demographic groups, namely race/ethnicity.





05 Compare Results

Compare outcomes across groups, particularly with respect to ethnicity. Determine whether equally qualified candidates from different ethnic backgrounds receive unequal treatment. Reflect on how this mirrors known patterns in hiring discrimination.

06 Reflect on Ethical Implications

Discuss whether and how the model reflects or reinforces social bias. Evaluate the effectiveness of fairness-aware interventions. Consider broader ethical implications for the use of AI in recruitment and talent management.

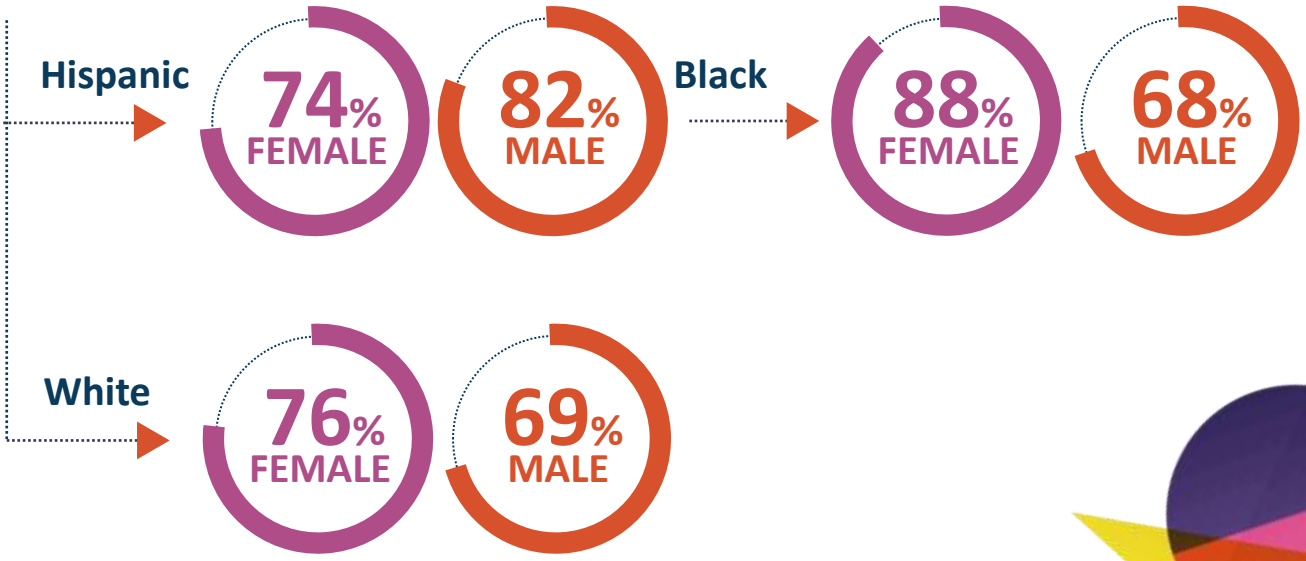
05 Conclusion

This simulation illustrates how machine learning models, if left unchecked, can embed and reproduce historical biases within automated recruitment pipelines. While the dataset used is synthetic and the model intentionally simplified, the findings mirror real-world concerns surrounding fairness in hiring—particularly with regard to gender and ethnic disparities.

The results reveal significant inconsistencies in predictive accuracy across demographic groups. Specifically, the accuracy of recruitment rejection predictions is notably higher for Black women. In this context, higher accuracy corresponds to a lower rate of false negatives, meaning that applications from Black women are more frequently and correctly classified as rejections

Output Summary

Accuracy in Predicting Recruitment Rejection Decisions — i.e., True "Negatives" (%)



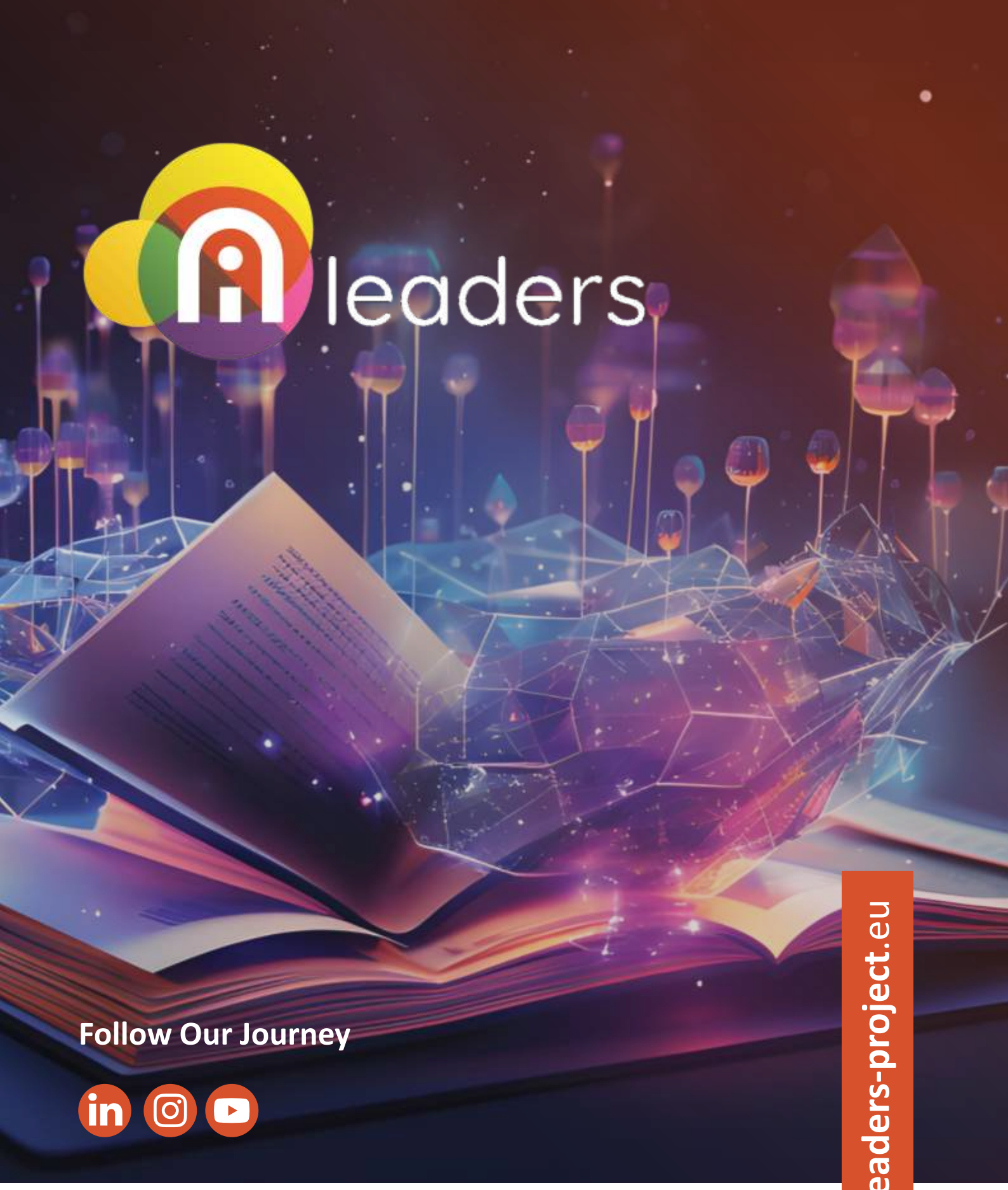
This suggests that the model disproportionately "recommends" rejecting candidates from this group, when comparing to those in other groups, highlighting a troubling bias in the algorithm's decision-making process.

These findings reignite critical discussions about blind spots in algorithmic hiring, particularly the ways in which structural inequalities become encoded into training data and subsequently learned by AI systems. The observation that predictive accuracy varies across marginalised groups—and even within protected categories—underscores the complexity of assessing fairness: greater accuracy does not necessarily indicate more equitable outcomes.

By applying fairness metrics and examining outcomes at the group level, students are invited to engage critically with the socio-technical dimensions of algorithmic bias. This exercise not only provides hands-on experience with bias detection and mitigation techniques, but also emphasises the ethical responsibility of developers and analysts in designing fair and accountable AI systems. In the context of human resource management, fairness is not merely a legal or reputational concern—it is a foundational ethical obligation.

06 References

- Allen, D. G., & Vardaman, J. M. (2017). Recruitment and Retention Across Cultures. Annual Review of Organizational Psychology and Organizational Behavior, 4(1), 153–181.
- Breaugh, J. A. (2013). Employee Recruitment. Annual Review of Psychology, 64(1), 389–416.
- Czopp, A. M., Kay, A. C., & Cheryan, S. (2015). Positive Stereotypes Are Pervasive and Powerful. Perspectives on Psychological Science, 10(4), 451-463.
- Daly, M. C., Büchel, F., & Duncan, G. J. (2000). Premiums and penalties for surplus and deficit education: Evidence from the United States and Germany. Economics of Education Review, 19(2), 169-178.
- Ellemers, N. (2018). Gender Stereotypes. Annual Review of Psychology, 69(1), 275-298.
- HeModern Discrimination in Organizations. Annual Review of Organizational Psychology and Organizational Behaviorbl, M., Cheng, S. K., & Ng, L. C. (2020)., 7(1), 257–282.
- Henderson, D. (2018). The Effects of Social Class on Perceptions of Job Applicants' Suitability for Employment. Academy of Management Proceedings, 2018(1), 13748.
- Hiemstra, A. M. F., Derous, E., Serlie, A. W., & Born, M. P. (2013). Ethnicity Effects in Graduates' Résumé Content. Applied Psychology, 62(3), 427–453.
- Horodyski, P. (2023). Recruiter's perception of artificial intelligence (AI)-based tools in recruitment. Computers in Human Behavior Reports, 10, 100298.
- Rigotti, C., & Fosch-Villaronga, E. (2024). Fairness, AI & recruitment. Computer Law & Security Review, 53, 105966.
- Seppälä, P., & Małecka, M. (2024). AI and discriminative decisions in recruitment: Challenging the core assumptions. Big Data & Society, 11(1), 20539517241235872
- Veit, S., & Thijsen, L. (2019). Almost identical but still treated differently: hiring discrimination against foreign-born and domestic-born minorities. Journal of Ethnic and Migration Studies, 1–20.
- Zaniboni, S., Kmicinska, M., Truxillo, D. M., Kahn, K., Paladino, M. P., & Fraccaroli, F. (2019). Will you still hire me when I am over 50? The effects of implicit and explicit age stereotyping on resume evaluations. European Journal of Work and Organizational Psychology, 28(4), 453-467.
- Zschirnt, E., & Ruedin, D. (2016). Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. Journal of Ethnic and Migration Studies, 42(7), 1115–1134.



Follow Our Journey

www.aileaders-project.eu

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.