Case study for education and reflection

Based on the education materials, we have designed a practical group assignment. The participants are expected to team up to analyse and discuss a case study. The objective of this assignment is to help you apply the lessons learned throughout the course in a practice case study. The expected outcome is to enable educators to better understand the complexities and the need for assessing fairness in AI systems. We refer to the case study "*Hiring by Machine*", developed by University Center for Human Values (UCHV) and the Center for Information Technology Policy (CITP) at Princeton and narrowed it down to the discussion on fairness, focusing on the social and legal aspects.

Role-playing: The group members will take on different roles as follows:

- Hiring manager
- HR Interviewer
- Technique Interviewer
- Data Scientist (who develops this AI recruitment system)
- Applicant (interviewee)

Background information. A bespoke resume vetting system (PARiS) was developed to help HR manage the influx of resumes. After weighing several options, the development team decided to implement a system that utilized natural language processing (NLP) and machine learning (ML) to identify markers in resumes that distinguished the best candidates. To train the system, HR provided the engineering team with dozens of resumes from current and previous employees who were deemed either exemplary or especially poor in terms of professional attributes and fit. The system would rate incoming resumes based on their match with the ideal profiles and discard those that were below a set threshold.

The HR was pleased with PARiS, as it dramatically reduced the number of hours needed to read each resume. While some members of the team were initially hesitant about delegating first-stage application sorting to an algorithm, skepticism about PARiS quickly abated as the system demonstrated its impressive capacity to learn. After only a few weeks of operation, the lists of candidates generated by PARiS consistently reflected those that would have been assembled by human HR agents, instilling confidence that the system had absorbed Strategeion's values. But PARiS was so much faster and more efficient than humans! Over time, growing trust in the system meant that the HR representatives felt less and less need to double-check PARiS' work, and they began shifting their energies elsewhere.

Case study: Hara, a promising and hard-working computer science student, received an automated rejection email within hours of applying for a job through its website. She was surprised to have been dismissed so quickly, as she was confident she was an ideal candidate for the company. Hara had strong academic qualifications and had carefully tailored her resume to highlight her civic commitments and experience working with nonprofit organizations that advocated for wheelchair users such as herself. Her ambitions to develop transparent, responsible tech solutions to improve the lives of those with disabilities seemed a perfect match with the company's mission to "leave no one behind."

Disappointed at her rejection, Hara wrote to the company requesting feedback on her application. She also published a blog post about her experience, promising to share any future response from the company. Her request eventually reached the HR department

Funded by
the European Union

where a representative reviewed her application and was equally puzzled by her rejection. Upon thoroughly examining her credentials, the HR representative concluded Hara was on par with the company's very best employees in terms of both interests and qualifications. Based on her resume alone, he believed she would be an excellent addition to the company, and couldn't understand why her application had been automatically discarded.

The HR representative flagged Hara's case for internal review. His supervisor, intrigued by the situation, decided to use the extra time freed up by the implementation of PARiS – the company's. automated hiring system - to convene a meeting with the system's engineers. The goal was to figure out why the system had rejected Hara's application. One initial concern raised during the meeting was that PARiS may have discriminated against Hara's disability status. However, the system's engineers reassured HR that the algorithm had been explicitly designed the algorithm to avoid discrimination against protected categories. Additionally, the company's policy of hiring ex-military personnel—many of whom were wheelchair users—meant that the system's training data was not biased against those with physical disabilities.

If her disability was not the issue, then what had caused PARiS to categorize her as a poor fit? What had the humans missed? After extensive investigation, the engineers discovered an unexpected factor: sports. PARiS had identified a strong positive correlation between participation in athletics and military service. Since veterans were overrepresented among the company's employees and often excelled in their roles, the system had learned to connect a history of playing sports with "good fit." While many of Strategion's ex-military employees are no longer active in sports, their resumes typically reflected past athletic involvement. Hara, however, had never been interested in sports. Having used a wheelchair her entire life, she also had no history of athletic activities.

Game playing Each role player is expected to contribute to the discussion from the perspective of their assigned role. The groups should follow the steps below to guide the discussion.

- State the problem
- Identify relevant factors and stakeholders
- Brainstorm possible solutions
- Evaluate solutions against the ethical requirements of fairness and make necessary adaption.
- Make a tentative choice
- Connect your choice to fairness metrics discussed earlier
- Review steps 1-6

Each group is required to submit an integrated report reflecting on the steps outlined above.

AEQUITAS

unbias AI

Consortium

UMEÅ UNIVERSITET

UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh

THE ADECCO GROUP

AKKODIS

SERVIZIO SANITARIO REGIONALE EMILIA-ROMAGNA POLICLINICO DI SANT'ORSOLA

PHILIPS

LOBA®

ALLAI.

PERIOD think tank

ARCIGAY Associazione LGBTI+ Italiana

EUROCADRES

ITI INVESTIGATE TO INNOVATE

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Universidad de La Laguna

Asociación Rayuela

Funded by the European Union. Views and opinions expressed are however those of the authors) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

www.aequitas-project.eu
info@aequitas-project.eu

Funded by
the European Union