

# Algorithmic foundations and ethics in AI: from theory to practice course

Toolkit for synchronous sessions

**CU5 | Case studies and projects**  
Support PowerPoint slides

---

# INDEX

- INTRODUCTION – 3
- ALGORITHMIC AND MACHINE LEARNING BIASES – 6
- ALGORITHMIC BIAS AND THEIR IMPLICATIONS IN REAL WORLD - CASE EXAMPLES - 16
- TOOLS TO DETECT ALGORITHMIC BIAS – 28
- PROJECT | AI IN RECRUITING - 35

# INTRODUCTION

---



IMAGE SOURCE | Freepik

---

## IN THIS COMPETENCE UNIT YOU WILL FIND THE FOLLOWING SUBJECTS:

- Presentation of various algorithmic and machine learning biases through case studies involving TikTok and facial recognition technologies.
- An overview of tools designed to identify algorithmic biases and their consequences.
- An in-depth project on the use of AI in recruitment, highlighting biases at various project phases.
- The role of fairness-aware approaches in machine learning and the implementation of bias auditing processes.

---

## AT THE END OF THE COMPETENCE UNIT, YOU SHOULD BE ABLE TO:

- Identify different types and causes of algorithmic bias, and examples of types of bias in machine learning data sets
- Evaluate algorithmic bias and their implications in real world
- Analyze different tools to detect algorithmic bias
- Analyze the project phases – Case “AI in recruiting”

# ALGORITHMIC AND MACHINE LEARNING BIASES

---



IMAGE SOURCE | Freepik

# Algorithmic and machine learning biases

## WHAT IS ALGORITHMIC BIAS AND WHY IS IT IMPORTANT TO CONSIDER?



IMAGE SOURCE | MISSING INFO

- In machine learning, algorithms use data sets or training data to learn patterns.
- Algorithmic bias occurs when the algorithm makes decisions that systematically disadvantage or discriminate against certain groups of people, for example, based on race or gender.
- Algorithmic bias can appear if a certain group of people is underrepresented in the data or in case existing societal biases are embedded within the data itself.
- It is crucial to understand the causes and implications of algorithmic bias in order to develop ethical and trustworthy AI solutions.

(Harvard Business Review, S. Friis & J. Riley, 2023)

# Algorithmic and machine learning biases

## Different types of algorithmic bias

**Algorithmic biases can be divided for example into three categories:**



1. **Pre-existing bias.** Bias in a system that exist independently and usually prior to the creation of the system. They can enter a system intentionally or unintentionally, whether through explicit efforts or unconscious actions, even with good intentions.



2. **Technical bias** arises from technical constraints or technical considerations. Sources of technical bias can be found in several aspects of the design process.



3. **Emergent bias** arises in a context of use with real users. This bias typically emerges sometime after a design is completed, as a result of changing societal knowledge, population, or cultural values.

Friedman and Nissenbaum (1996)



# Algorithmic and machine learning biases

## Different causes of bias

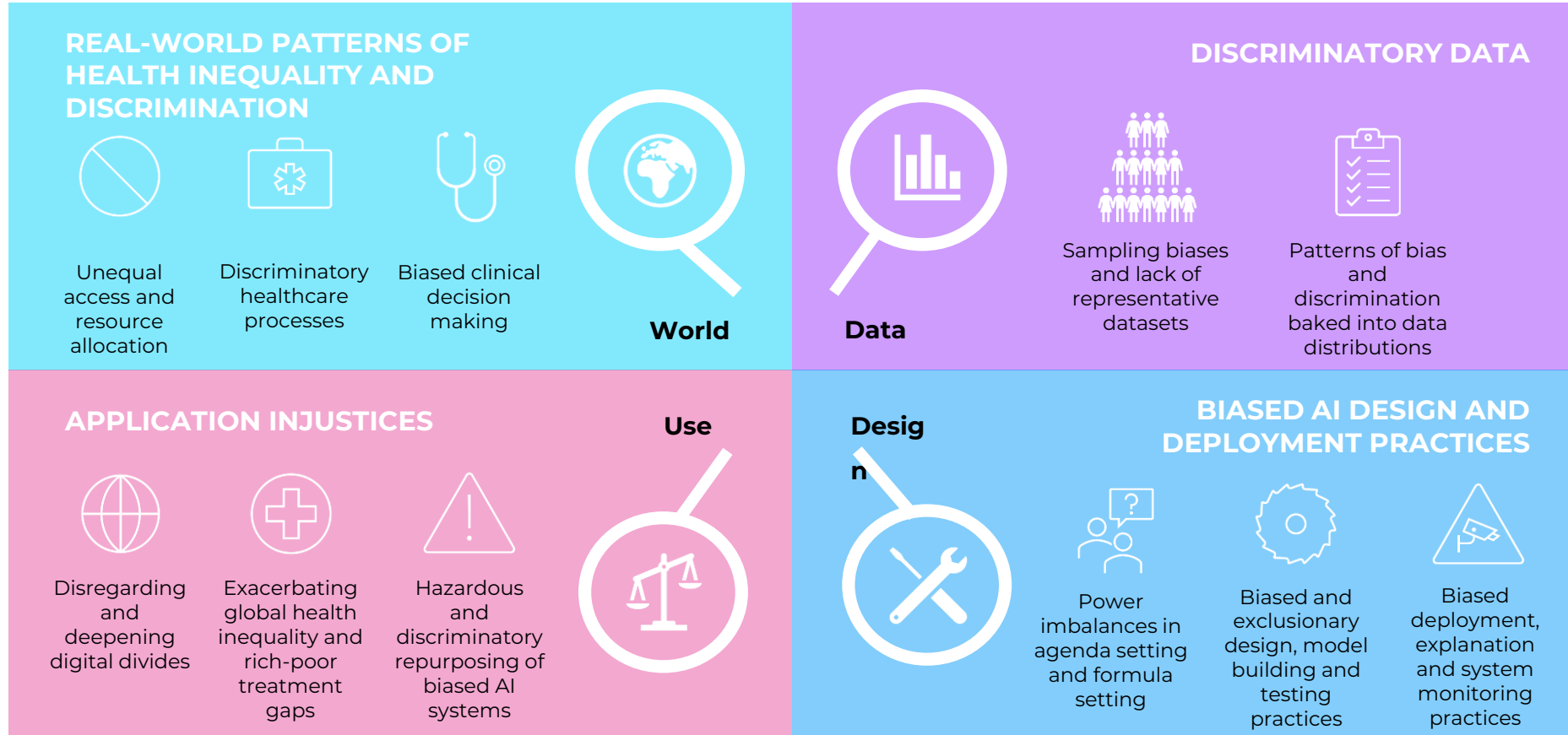


IMAGE SOURCE | Image redesigned (the original image: British Medical Journal)

Digital technologies tend to reinforce accelerate inequalities which can appear e.g. as discriminatory applications. This is a cause of discrimination in AI's decision making, which results from ethical gaps in the development, implementation or availability of these technologies.

# Algorithmic and machine learning biases

## Example of types of bias in machine learning data sets

Engineers train machine learning models by feeding them a set of training examples. The involvement of humans in curating and delivering the data can cause bias.

### **Let's analyse the following types of bias:**

- Reporting bias
- Group attribution bias
- Implicit bias
- Automation bias
- Selection bias

# Algorithmic and machine learning biases

## Example of types of bias in machine learning data sets

### Reporting bias

- Frequency of events, properties, or outcomes in a dataset does not reflect real-world frequency.
- Bias arises from a focus on documenting unusual or memorable circumstances.
- Ordinary situations may be neglected, as they are assumed to be self-evident.

### Example

In a sentiment analysis model training, it predicts if book reviews are positive or negative based on user submissions to a popular website. Most reviews in the training data express extreme opinions (either loving or hating a book) because people are less likely to review a book if they have a moderate reaction. As a result, the model struggles to correctly predict sentiment for reviews using subtle language to describe a book.

# Algorithmic and machine learning biases

## Example of types of bias in machine learning data sets

### Group attribution bias

Group attribution bias is the tendency to apply individual characteristics to an entire group they belong to. There are two main forms:

- **In-group bias:** preferring members of your own group or shared characteristics.

**Example |** Two engineers training a résumé-screening model for software developers may favor applicants from the same computer science academy they attended, assuming they are more qualified.

- **Out-group homogeneity bias:** stereotyping individuals in a group you don't belong to, viewing their traits as more uniform.

**Example |** Two engineers training a résumé-screening model may assume that all applicants who didn't attend a specific computer science academy lack expertise for the role.

# Algorithmic and machine learning biases

## Example of types of bias in machine learning data sets

### Implicit bias

Implicit bias occurs when assumptions are made based on personal mental models and experiences that may not apply universally.

**Example |** An engineer training a gesture-recognition model uses a head shake to indicate "no." However, in some regions, a head shake means "yes."

- A common type of implicit bias is **confirmation bias**, where model builders unconsciously process data to affirm existing beliefs.
- Sometimes, model builders may keep training until results align with their original hypothesis, known as **experimenter's bias**.
- **Example |** An engineer building a dog aggressiveness model associates hyperactivity with a childhood encounter with a toy poodle. Despite the model initially portraying toy poodles as calm, the engineer continues retraining it until it indicates that smaller poodles are more aggressive, thereby affirming their initial bias.

# Algorithmic and machine learning biases

## Example of types of bias in machine learning data sets

### Automation bias

Preference for results from automated systems over non-automated ones, regardless of their respective error rates.

#### Example

Software engineers at a gear manufacturer were excited to introduce a newly trained "innovative" model designed to detect faulty gears. However, their enthusiasm faded when a factory manager pointed out that the model's precision and recall rates were 15% lower than those achieved by human inspectors.

# Algorithmic and machine learning biases

## Example of types of bias in machine learning data sets

### Selection bias

Selection bias occurs when examples in a dataset are not chosen in a way that represents their real-world distribution. It can take several forms.

- **Coverage bias:** data is not selected in a representative way.

**Example |** An AI model is trained to recommend restaurants based on customer reviews. However, the model is only trained on the reviews of one popular restaurant chain, ignoring reviews of other restaurants. As a result, the AI model's recommendations may not accurately reflect the different tastes and preferences of all potential customers as it has not taken into account the opinions of customers who visited other restaurants.

- **Non-response bias (or participation bias):** data becomes unrepresentative due to gaps in participation during data collection.

**Example |** An AI-based movie recommendation system collects feedback from users of a particular streaming platform but ignores feedback from users of other platforms, resulting in their preferences not being adequately represented. As a result, the AI-generated movie recommendations may not effectively reflect the different tastes of all users due to the lack of feedback from users of different streaming platforms.

Sampling bias: proper randomization is not used during data collection.

- **Sampling bias:** proper randomization is not used during data collection.

**Example |** An AI model is trained to predict future sales of a new product by surveying people who bought it and people who bought a similar product about their buying habits. Instead of randomly choosing people, the surveyors picked the first 200 who replied to the survey, which might mean they were more excited about the product than most buyers.

# ALGORITHMIC BIAS AND THEIR IMPLICATIONS IN REAL WORLD - CASE EXAMPLES

---



IMAGE SOURCE | Freepik



# Algorithmic and machine learning biases

## CASE – TikTok's harmful algorithm

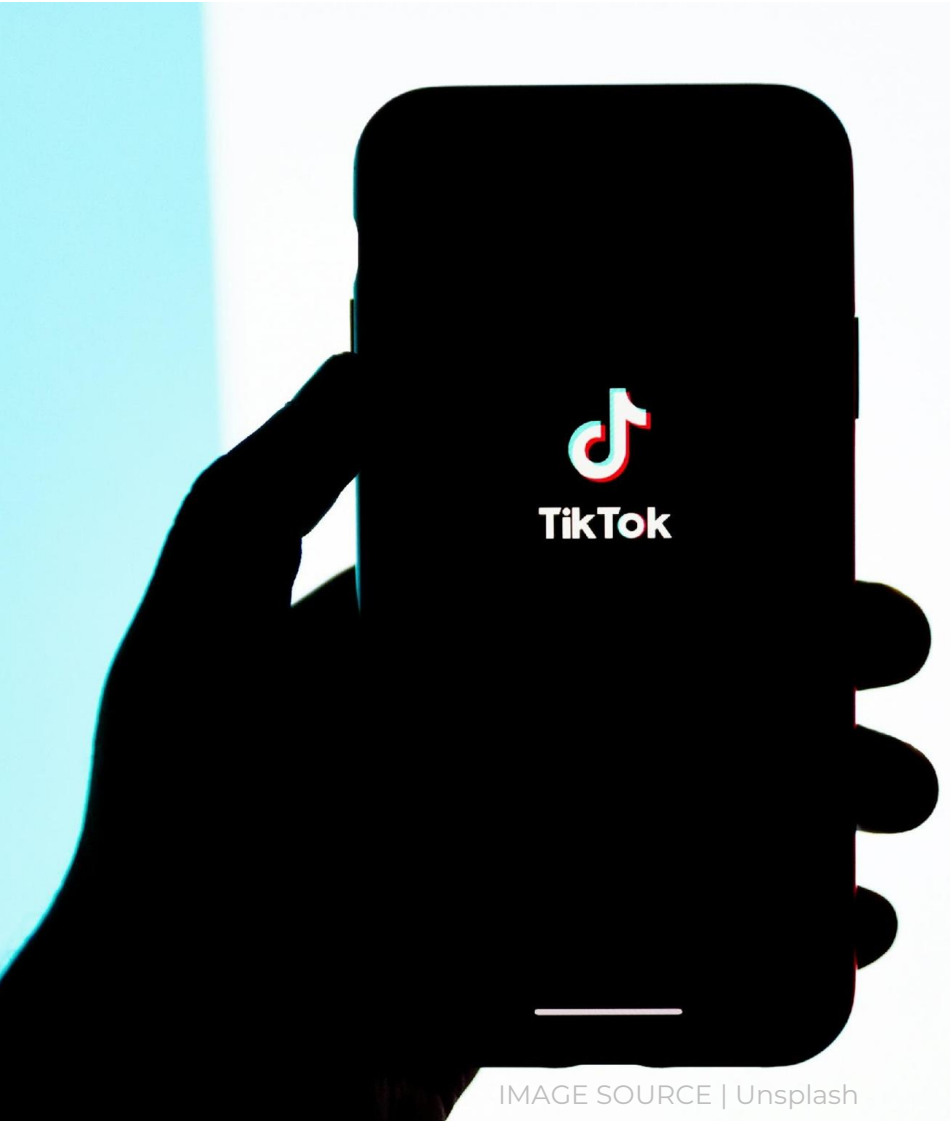


IMAGE SOURCE | Unsplash

Finland's national media company Yle investigated what kind of content TikTok's algorithm shows to a **13-year-old teenage girl suffering from depression and distorted body image**.

- Yle created a profile for **fictional character Ella** that represents the depressed 13-year-old girl.
- Yle's data scientist browsed TikTok with Ella's profile for a little over five hours.

Have a look on the videos that TikTok showed for **Ella on Yle's article**

Yle, November 11, 2023

# Algorithmic and machine learning biases

## CASE – TikTok's harmful algorithm

The results were alerting

### 1. HOUR - harmful content **7%** out of 100%

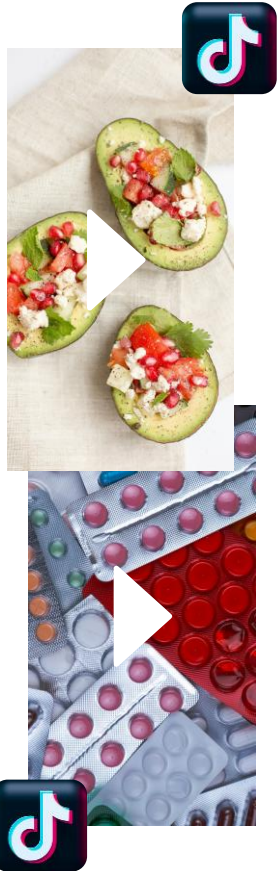
(Videos on mental health problems and food issues were counted as harmful content.)

- At first - cheerful videos about food, puppies, cats, skateboarding and jokes.
- Ella watches the food videos a little longer.
- After only 5 minutes Ella stops to look a video about a woman that is screaming.
- In just a moment, TikTok shows a video about antidepressants.
- More videos on mental health and light meals.

**"TikTok's algorithm quickly learns what topics a browser is interested in."**

### 2. HOUR - harmful content **28%** out of 100%

- The videos' tone gets darker.
- Occasional content on subjects such as hair care and animals.
- **Content related to suicide and self-harm is now starting to appear in the videos.**
- Among the videos, there is also content that, for example, encourages people to keep going despite difficulties.



Yle, November 11, 2023

# Algorithmic and machine learning biases

## **CASE – TikTok's harmful algorithm**

The results were alerting

---

"In one video, a woman with a mental disorder talks about how she sometimes feels like taking her own life."

Yle, November 11, 2023

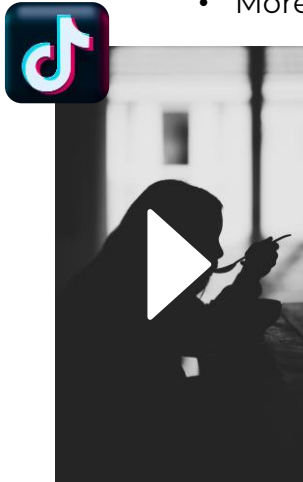
# Algorithmic and machine learning biases

## CASE – TikTok's harmful algorithm

The results were alerting

### 3. HOUR - harmful content **68%** out of 100%

- **Just before the start of the third browsing hour, Tiktok shows Ella the first video on eating disorders.**
- After the first video, more eating disorder content starts to appear, and Ella stops to watch it.
- More content on low-calorie foods and self-harm.



### 4. HOUR - harmful content **57%** out of 100%

- Slightly fewer videos about depression and self-harm.
- Instead, the Tiktok algorithm seems to have recognised that Ella is particularly interested in light meals and the toxic content that encourages her to limit her eating.
- More food and weight loss videos than before.

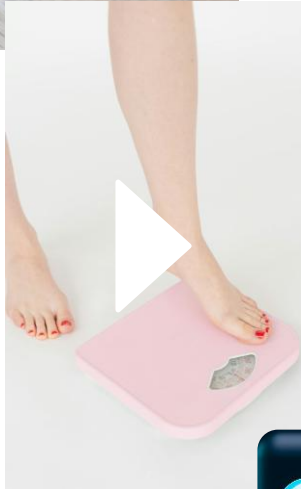
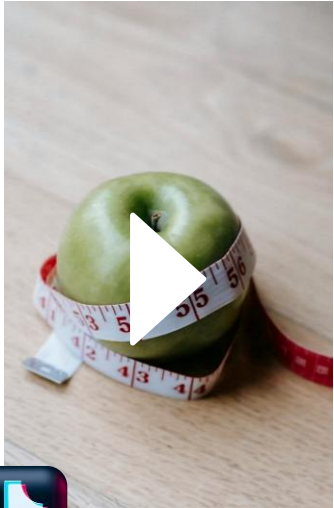


Yle, November 11, 2023

# Algorithmic and machine learning biases

## CASE – TikTok's harmful algorithm

The results were alerting

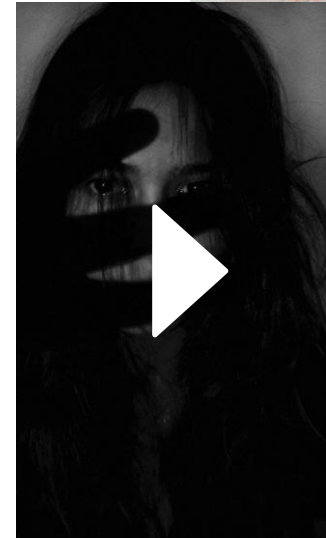
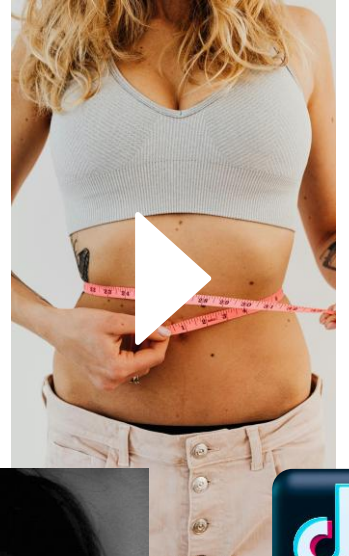


**5. HOUR** - harmful content **65%** out of 100%

- **Even more sad videos appear on Ella's screen.**
- Videos that glorify eating disorders and now directly give concrete advice on how to limit eating.

**THE END OF BROWSING** - harmful content **95%** out of 100%

- At the end of the browsing process, the Tiktok algorithm had shown Ella a huge number of videos related to depression, suicidal tendencies, low-calorie diets and eating disorders.



Yle, November 11, 2023

# Algorithmic and machine learning biases

## **CASE – TikTok's harmful algorithm**

The results were alerting

---

According to social psychologist, Suvi Uski, the algorithm shows the user content that interests them as much as possible, no matter how harmful it is. Uski says this is particularly problematic because, at the same time, Tiktok is by far the most addictive of the social apps.

Yle, November 11, 2023

# Algorithmic and machine learning biases

## Ethical concerns related to facial recognition

### Facial recognition – what and how?

- A facial recognition system **scans and identifies an individual's face by comparing it to a stored digital image or a frame from a video clip** within a database using an AI algorithm.
- The technology works **by identifying facial features** in an image **and comparing them to other** images or videos in the database using an AI algorithm.

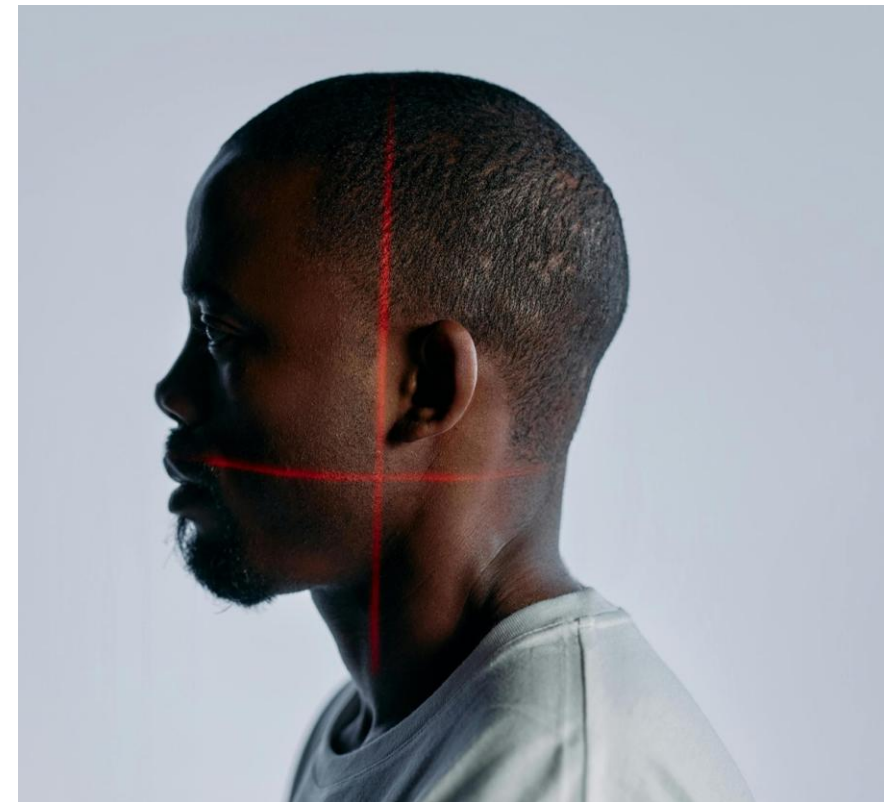


IMAGE SOURCE | Pexels

Shehmir Javaid, January 12, 2024)



# Algorithmic and machine learning biases

## Facial recognition – examples of its use



- 1 **Healthcare.** Automatically scanning the patient's face using face recognition to extract medical history and insurance information. Also used in diagnosing medical disorders.
- 2 **Retail.** Amazon uses face recognition for contactless payments. Can be used in personalizing the shopping experience, for example, for customer loyalty offers. Can also be used to improve retail security.
- 3 **Banking and finance.** Used in customer identification and verification. To speed up the customer onboarding process. And to reduce robberies or banking frauds.
- 4 **Law enforcement and security services.** In identifying and recording criminals and persons of interest. To detect fugitives by using surveillance cameras in a city, e.g. at junctions, train and bus stations and airports.



# Algorithmic and machine learning biases

## CASE – Unethical facial recognition data collection



Video on facial recognition at airports on the washington post Youtube channel

- ▶ Watch the video on facial recognition at airports.
- ? What ethical issues arise during the video?

Washington Post, Jun 10, 2019

# Algorithmic and machine learning biases

## Ethical concerns on facial recognition data collection

---

“FBI, ICE find state driver’s license photos are a gold mine for facial-recognition searches”

[Read the news article on The Washington Post on July 7, 2019](#)

Agents from the Federal Bureau of Investigation and Immigration and Customs Enforcement have utilized state driver's license databases as a rich source for facial recognition and scanned through millions of Americans' photos without their awareness or consent.

Drew Harwell, The Washington Post, July 7, 2019

# Algorithmic and machine learning biases

## Racial bias in facial recognition systems

---

“Asian and African American people were up to 100 times more likely to be misidentified than white men, depending on the particular algorithm and type of search.”

[Read the news article on The Washington Post on December 19, 2019](#)

Federal study confirmed that **many of the facial recognition systems have racial bias.**

Drew Harwell, The Washington Post, December 19, 2019

# TOOLS TO DETECT ALGORITHMIC BIAS

---



IMAGE SOURCE | Freepik

# Tools to detect algorithmic bias

## UNESCO's Ethical Impact Assessment

- Developed by UNESCO, 2023
- A tool to conduct an ethical assessment by evaluating and identifying AI systems' benefits and risks
- The tool consists of diverse questions that are ready to be filled in
- Both open and multiple-choice questions
- Ethics considered by design throughout the assessment
- Starting with scoping questions followed by questions related to UNESCO principles.

### The tool will help you answer questions such as:

- Who will interact with the AI solution?
- What are the roles and responsibilities in the AI team?
- Who are the stakeholders that will be impacted by the AI system?
- How is the safety and security considered?
- Any issues related to sustainability, privacy, transparency, explainability and accountability

[Find it here](#)

Figure 2: Positionality Matrix (developed by The Alan Turing Institute)



UNESCO, 2023

IMAGE SOURCE | UNESCO

# Tools to detect algorithmic bias

## UNESCO's Ethical Impact Assessment

### How the tool detects bias?

- Section Fairness, Non-Discrimination, Diversity focuses more in detail in bias bias that appear in data
- Bias are also addressed through paying attention on roles and responsibilities.

## 6.2. Procedural Assessment

Throughout this section, when responding to questions about testing with particular groups, the project team should consider especially – but not only – race, colour, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, and disability. Please specify if testing was conducted on groups that combine several of these criteria i.e. if the system has been tested in terms of intersectionality.

### 6.2.1. Preventing discriminatory outcomes:

6.2.1.1. Has the algorithm been tested with different groups?

6.2.1.1.1. Was there a difference in terms of accuracy rate (or any other performance metric used)? Please describe any difference of this kind.

6.2.1.1.2. Was there a discriminatory effect for particular groups?

### 6.2.2. Data quality and preventing discriminatory bias:

6.2.2.1. Are processes in place to test data against biases?

6.2.2.1.1. Have you undertaken an analysis of the data to prevent societal and historical biases in data?

6.2.2.1.2. Is the data well-balanced and does it reflect the diversity of the targeted end-user population?

6.2.2.1.3. Are there any differences you can foresee between the data used for training and the data processed by the AI system which could result in the AI system producing discriminatory outcomes or performing differentially for different groups?

6.2.2.1.4. Have you developed a process to document how data quality issues can be resolved during the design process?

6.2.2.1.5. Did you put in place educational and awareness initiatives to help AI designers and developers gain awareness of the possible bias they can introduce through the design and development of the AI system?

6.2.2.1. Are processes in place to test data against biases?



# Tools to detect algorithmic bias

## ALTAI

- Trustworthy AI assessment list
- Developed by the High-level expert group (AI HLEG) on artificial intelligence, The European Commission
- The Assessment List for Trustworthy AI (ALTAI)
- List of 7 requirements for trustworthy AI that are defined more specifically on the website
- Includes a recommendation on how to complete the ALTAI and who should be involved
- [Find it here](#)

## Sections of the ALTAI

- 📄 Human Agency and Oversight
- 📄 Technical Robustness and Safety
- 📄 Privacy and Data Governance
- 📄 Transparency
- 📄 Diversity, Non-Discrimination and Fairness
- 📄 Societal and Environmental Well-being
- 📄 Accountability

Is the **AI system** designed to interact, guide or take decisions by human **end-users** that affect humans ('**subjects**') or society? ⓘ \*

- ☐ Yes
- ☐ To some extent
- ☐ No
- ☐ Don't know

Did you put in place procedures to avoid that **end-users** over-rely on the **AI system**? ⓘ \*

- ☐ Yes
- ☐ No

# Tools to detect algorithmic bias

## ALTAI

### How the tool detects bias?

- The ALTAI tool's section Diversity, Non-discrimination and Fairness focuses on avoiding unfair bias.
- The ALTAI tool assesses algorithmic bias through targeted questions, evaluating input data and algorithm design for biases, while promoting inclusivity, accessibility, and stakeholder participation to ensure fairness and transparency in AI systems.

### Avoidance of unfair bias

Did you establish a strategy or a set of procedures to avoid creating or reinforcing **unfair bias** in the **AI system**, both regarding the use of input data as well as for the algorithm design?**?** \*

- ☐ Yes  
☐ No

Did you consider a mechanism to include the participation of the widest range of possible stakeholders in the **AI system's** design and development?**?** \*

- ☐ Yes  
☐ No

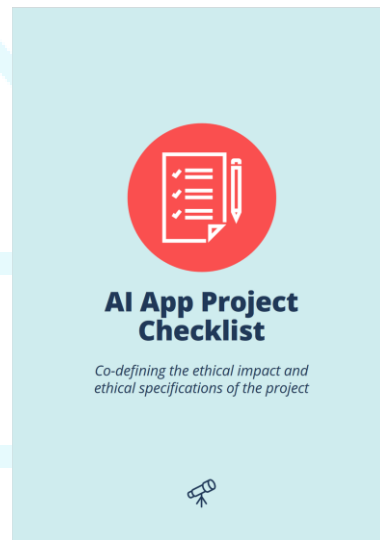


# Tools to detect algorithmic bias

## Ethical Toolkit for the Development of AI Applications

- Developed by Mario Sosa Hidalgo, (TU Delft Industrial Design Engineering)
- Considers ethics in development of AI applications
- Innovative toolkit that has an active, concrete and visual approach

[Find it here](#)



### AI APP PROJECT CHECKLIST

Name of the Project

1 Goal

2 Team Accountable for App

3 Context of the Project

4 Stakeholders Impacted

5 Type of Impact

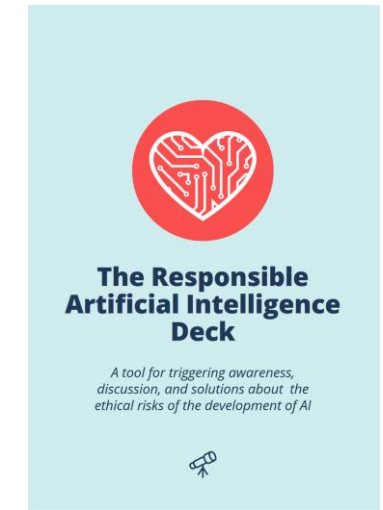
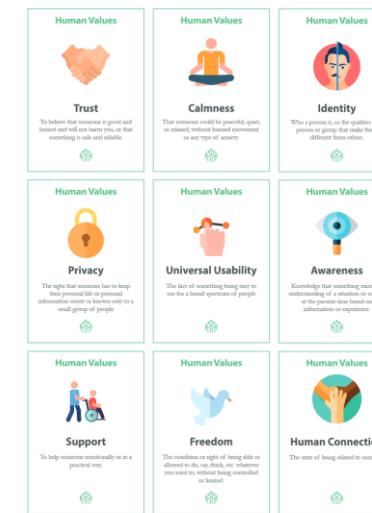
6 AI Algorithm & Learning Model used

7 Data Type

8 Data Source

9 Ethical Principles Checklist (Order by relevance for project: High(+)/ Med(0)/ Low(-))

| High(+)        | Med(0)               | Low(-)           |
|----------------|----------------------|------------------|
| Data Privacy   | Honest Communication | Human Well-Being |
| Data Safety    | Accountability       | Governance       |
| Explainability | Fairness             | User Safety      |
| Transparency   | Value Alignment      |                  |



Mario Sosa Hidalgo, 2019

# Tools to detect algorithmic bias

## Ethical Toolkit for the Development of AI Applications

### How the tool detects bias?

Mario Sosa's tool addresses bias and recommends developers to ensure to make sure that the AI system developed is as unbiased as possible reminding of strategies for fairness failures.



Top icons by Flaticon.com

Mario Sosa Hidalgo, 2019

# PROJECT | AI IN RECRUITING

---



IMAGE SOURCE | Freepik

# Project | AI in recruiting



## PROBLEM DEFINITION & BUSINESS UNDERSTANDING



## Case study | Background

A global industrial company faces inefficiencies and uncertainties in its recruitment process, which puts a significant strain on the HR department by consuming time and resources.

Anticipating future growth, the company also recognizes the urgent need to hire a variety of specialists, including engineers, marketers, finance professionals, coders and project managers.

Fierce competition for competent employees and the high costs associated with recruitment errors emphasize the importance of successful recruitment. The situation increases the pressure on the HR team, whose task is to ensure the success and smooth running of the recruitment process.

As a result, the company is exploring the development of an artificial intelligence solution tailored to its recruitment needs.

When integrating AI into the recruitment process, the primary goal is to improve efficiency, objectivity and the overall success rate of recruitment.

**As a first step, the pros and cons of an AI recruitment solutions are considered.**

# Project | AI in recruiting

## AI recruitment | PROS and CONS

### PROS

- **Efficiency and Time Management:** AI streamlines tasks, focusing recruiters on strategy, speeding up hiring
- **Enhanced Candidate Quality:** AI analyzes data to select best-fit candidates, reducing turnover
- **Improved Applicant Experience:** AI provides timely updates, improving communication
- **Cost Reduction:** AI cuts recruitment costs by reducing manual work in routine tasks
- **Increased Fairness:** AI reduces bias, enhancing objectivity in hiring  
-> Diversity enhancement
- **Insightful Social Media Analysis:** AI evaluates social media for cultural fit
- **Global Recruitment:** AI overcomes language barriers, accessing wider talent pools

### CONS

- **Reduced Personal Engagement:** AI makes recruitment more mechanical, possibly deterring candidates
- **Bias Risk:** Improper AI design or biased data can compromise diversity and fairness perpetuating social inequalities
- **Data Security:** Personal data handling by AI raises privacy and cyberattack risks
- **High Investment Costs:** AI in recruitment demands significant financial and resource investments
- **Human Judgment:** AI reliance might miss skilled candidates due to data issues
- **Legal Compliance:** Recruitment AI must adapt to changing laws to maintain fairness
- **Cultural Assessment:** AI may fail to evaluate cultural fit or teamwork effectively

([www.recruiter.com](https://www.recruiter.com), 2023; Bursell & Roumbanis, 2024; Arivu Recruitment, 2023; Köchling & Wehner, 2020, Albaroudi et al., 2024, Sheard, 2022; Albassam, 2023).



# Project | AI in recruiting



## PROBLEM DEFINITION & BUSINESS UNDERSTANDING

### Understanding the problem

The company begins background research to develop an AI solution, forming an internal team dedicated to identifying key challenges in the existing recruitment process. The team conducts interviews with a range of stakeholders, including the HR department, newly hired employees, and managers who have recently onboarded new team members. The team also gathers data on the current costs of recruitment and analyzes the success rate of these initiatives.

Understanding the perceptions of potential employees regarding the use of AI in recruitment is crucial for the company. It aims to avoid any damage to its reputation or brand. To achieve this, the company is committed to ensuring that the adoption of artificial intelligence not only gains widespread acceptance but also adheres to ethical and legal standards.



# Project | AI in recruiting



## PROBLEM DEFINITION & BUSINESS UNDERSTANDING



## Understanding the users of the system

One of the interviewed persons is Lead Recruiter Maria, who has several years of experience in the company. She has seen firsthand the accumulation of recruitment tasks and the growing pressure it causes on the HR department, which intensifies when hiring global talent.

Maria is open-minded and interested in the possibilities of AI to help with recruitment tasks, even though she is not deeply familiar with artificial intelligence technology.



### TOOL

**Personas** are detailed user profiles in design that embody key traits of target user groups, avoiding stereotypes. They include behaviors, needs, and demographics to foster empathy and guide design teams in creating user-centric solutions. They're useful for transcending conventional market segments for more inclusive design outcomes. (Kumar, 2013; Stickdorn et al., 2018a, 2018b)

# Project | AI in recruiting

## Maria – LEAD RECRUITER

#motivationdetermines  
#goodtalents



IMAGE SOURCE | Image made by VAMK

### DEMOGRAPHICS

Age: 37  
Location: Helsinki, Finland  
Nationality: Guatemalan  
Education: MBA Human Resource Management  
Job: Lead recruiter  
Family: Lives alone with her cat

### WORK GOALS

Finding new talents  
Employee satisfaction  
Fast recruiting process

### HOBBIES



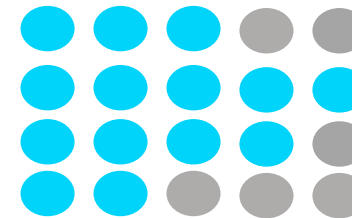
Swimming  
Climbing  
Reading

### FRUSTRATIONS

- ✓ High work load
- ✓ Finding a good talent, who backs away just before job start
- ✓ Too slow recruiting process making candidates go elsewhere
- ✓ Not enough time to keep the candidates informed during the recruiting process
- ✓ Feels pressure to find the perfect candidates

### SKILLS

Social Media  
HR software  
Interviewing  
Benchmarking



Maria has high hopes, but she's also worried about the impact AI-driven recruiting will have on her work. She is concerned about how the introduction of AI may change the functions of the HR staff in the future: What new capabilities and competencies does she need to possess? Will her contribution be less valued? Is her expertise still needed? And to what extent can the recruitment process be automated?

**Empathetic**

**Open minded**

**Outgoing**

**Team player**

**Perfectionist**



# Project | AI in recruiting



## PROBLEM DEFINITION & BUSINESS UNDERSTANDING



## Understanding the recruitment journey

To gain a comprehensive understanding of the recruitment process, the company develops a recruitment journey involving Maria and the recruitment team. Collaborating with line managers, they navigate through each stage of the journey to identify significant frustrations and time-consuming steps. Additionally, they explore the potential integration of AI at each stage to enhance the recruitment journey's efficiency and effectiveness



### TOOL

**A journey map** is a design tool that outlines a user's step-by-step service experience, including interaction stages, touchpoints, and challenges. It also captures emotional responses to highlight the quality of the experience. Journey maps can represent current or future interactions and range from broad overviews to detailed snapshots. (Kumar, 2013; Stickdorn et al., 2018a)

# Project | AI in recruiting

## Maria's – recruiting journey



### IDENTIFYING HIRING NEEDS

Need for a new person is identified by line manager and informed to Maria



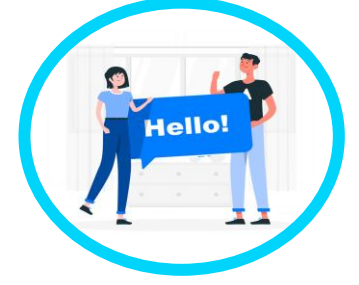
### TALENT SEARCH

Maria needs to identify and attract new talents



### INTERVIEWS

Maria and line manager interview good candidates



### INTRODUCTION AND INDUCTION

Maria sends the induction kit to employee and induction continues when the employee arrives



### PREPARING THE JOB DESCRIPTION

Job description is created to clarify the specific skills needed



### SCREENING AND SHORTLISTING

Maria needs to identify the right applicants from a big pool of candidates



### EVALUATION AND OFFER

Selection and background check with reference. Offer is handed over to candidate.

# Project | AI in recruiting

## How AI can be utilized in different stages



### IDENTIFYING HIRING NEEDS

- Analyzing external factors like market demands, competitor actions, and HR metrics
- Assessing historical labor trends, demographics, and skill requirements alongside internal evaluations
- Tracking metrics like turnover rates and time-to-fill positions for recruitment efficiency



### TALENT SEARCH

- Enabling tailored communication on digital platforms and social networks
- AI bots pinpointing both active and passive job seekers across platforms and past applicant databases
- Employing text-mining strategies to enhance job listing appeal with candidate-attracting phrases



### PREPARING THE JOB DESCRIPTION

- Mitigating Human Bias in Job Description Creation
- Adjusting language in advertisements and monitoring impact on application volume and quality



### SCREENING AND SHORTLISTING

- Initial Screening by analyzing semantics and predicting long-term potential and negative behaviors
- Video Interviews by assessing verbal and non-verbal cues to evaluate personality and cognitive skills
- Social Media Analysis by examining online behaviors for cultural fit
- Resume Analysis with NLP techniques enriching resume evaluations

Recruiter.com; Chen, 2022; Hunkenschroer & Luetge, 2022, Lawton ; Albouradi 2024; Marr 2023; Fritts & Carbera,2021; Vivek, 2023

# Project | AI in recruiting

## How AI can be utilized in different stages



### INTERVIEWS

- AI interviewing techniques used mostly in the pre-screening phase



### EVALUATION AND OFFER

- Analyzing candidate's previous job positions to create a tailored offer that the candidate is likely to accept.



### INTRODUCTION AND INDUCTION

- Automating document creation
- Monitoring the onboarding progress
- Tailoring educational paths

Recruiter.com; Chen, 2022; Hunkenschroer & Luetge, 2022, Lawton ; Albouradi 2024; Marr 2023; Fritts & Carbera,2021; Vivek, 2023

# Project | AI in recruiting



## DESIGN PHASE



## Recognizing the personal biases

After a clear problem statement, the company has chosen to proceed with an AI-driven solution. You have been appointed as the lead AI developer responsible for overseeing the implementation of this chosen AI technology.

The company acknowledges the importance of assembling a diverse, multi-stakeholder group to assure a variety of perspectives for the success of the project.

In an effort to address potential unconscious biases within the team, the company has mandated that all members participate in an exercise modeled after Alan Turing called “Positionality Matrix



### TOOL

**Alan Turing positionality matrix** is designed to illuminate the team's personal biases, demonstrating how socioeconomic backgrounds, cultural contexts, and individual life experiences can influence one's judgment and decision-making.

# Project | AI in recruiting

## Alan Turing test | Reflect on purpose, positionality and Power



### Personal Characteristics & Group identification

How do I identify?  
Age, race & ethnicity, disability status, religion, gender, sexuality, marital status, parental status, linguistic background



### Education, Training & Work Background

How have I been educated and trained?  
Schools attended, level of education, opportunities for advancement and professional development, employment history



### POSITIONALITY MATRIX

To what extent do my personal characteristic, group identifications, socioeconomic status, educational, training & work background, team composition & institutional frame represent sources of power and advantage of source of marginalization and disadvantage? How does this positionality influence my (and my team's) ability to identify & understand affected stakeholders and the potential impacts of my project?



### Institutional Frame and Team Composition

What does my institutional context and team composition look like?  
Authority structure within my project team, wider policy-ownership and power hierarchies in my organization, levels of decision-making autonomy, opportunities to voice concerns & objections, team diversity, culture of inclusion or exclusion



### Socioeconomic Status

What is my socioeconomic history?  
Socioeconomic status growing up, social mobility over time, present status, socioeconomic aspiration

# Project | AI in recruiting



## DESIGN PHASE

### Stakeholder map

On top of the internal stakeholder group taking part in the project, company also needs to consider other related stakeholders.

Therefore, company created a stakeholder map following G.J. Millers definition of the AI development process stakeholder mapping defining the development and usage stakeholders as well as the external stakeholder.



#### TOOL

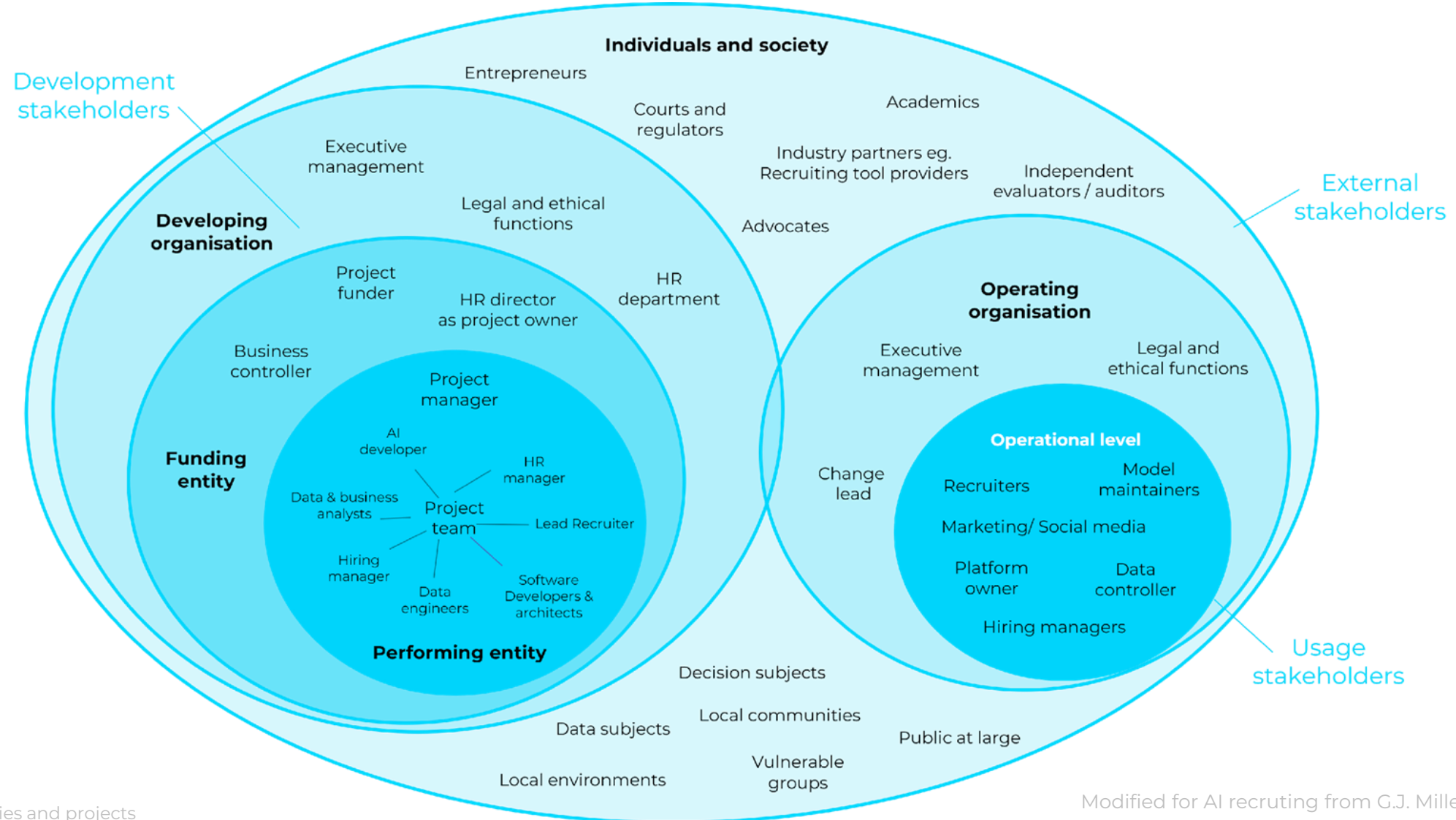
**A stakeholder map** is a visual tool used to identify and display the roles and relationships of all stakeholders involved in a project. It can range from a simple quadrant showing levels of influence and engagement to a detailed matrix detailing interactions among stakeholders. Stakeholder map helps to clarify stakeholder dynamics and is crucial for understanding and managing network relations. (Giordano et al., 2018; SDT, n.d.-b)





# Project | AI in recruiting

## Stakeholder map AI recruiting project





# Project | AI in recruiting



## DESIGN PHASE

### Stakeholder analysis matrix

Based on the map, the project manager needs to organise the stakeholders into groups in order to know, how to keep the stakeholders informed and to see what is their influence on the project.



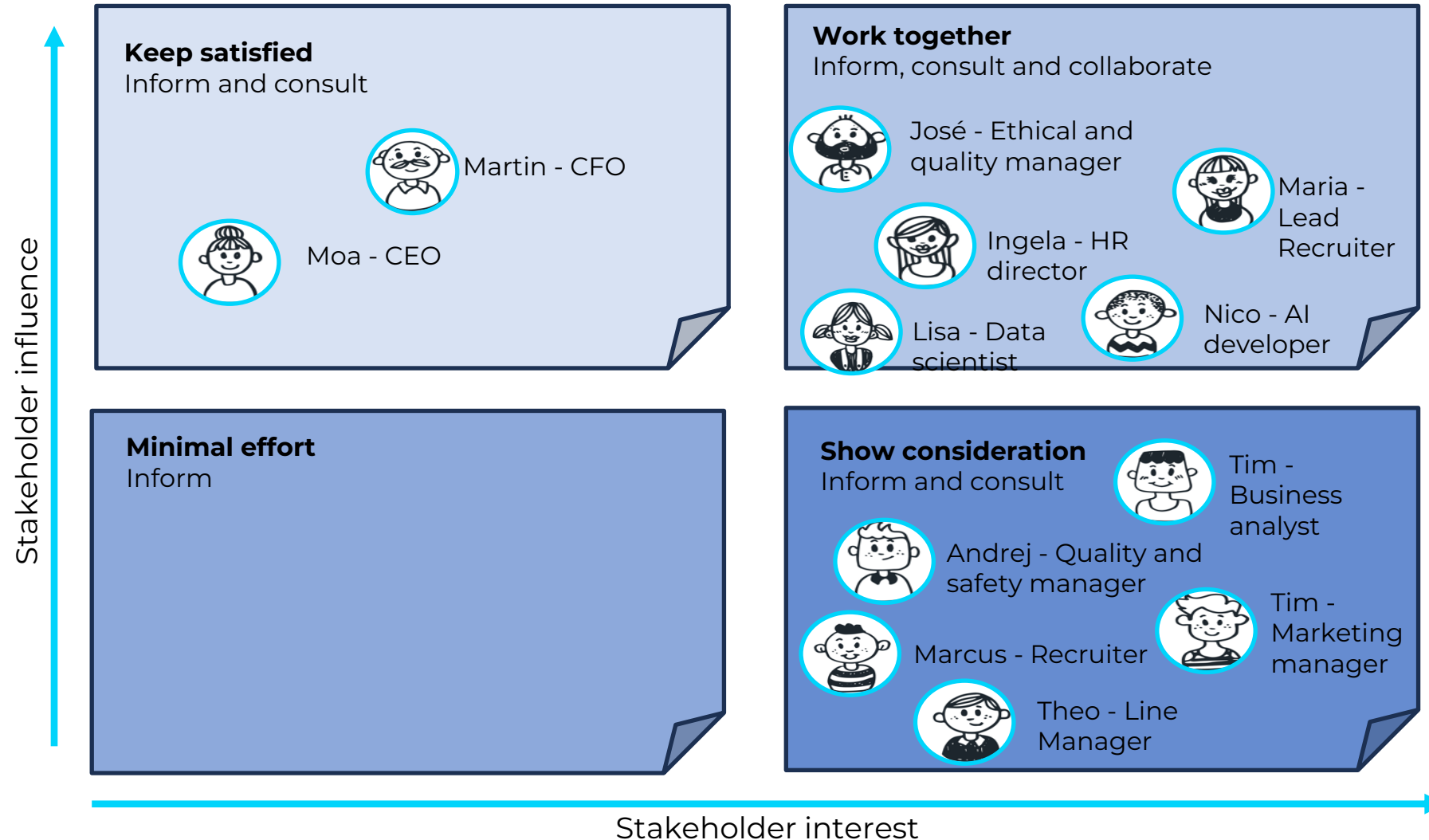
#### TOOL

The **stakeholder analysis matrix** organizes project stakeholders by influence and interest in a quadrant grid to inform engagement tactics. It guides how to satisfy high-influence, low-interest stakeholders; closely manage those with high interest and influence; keep informed those interested but less influential; and minimally engage those with little of either. (Hoory & Botorff, 2022; Wallbridge, 2023)



# Project | AI in recruiting

## Stakeholder analysis matrix



# Project | AI in recruiting



## DESIGN PHASE



## Biases in recruitment stages - Workshop

When the stakeholders are clear, company decides to have an ethical workshop going through all the possible biases in different hiring stages.

Stakeholders from the "Work together" field (previous slide) are invited to join.

What kind of post-its would you have filled in?



### TOOL

A **workshop** is a collaborative design tool where a group of participants engages in interactive activities to address a specific problem or work on a project. Typically facilitated by a leader, workshops can vary in duration from a few hours to several days, depending on the objectives and complexity of the tasks at hand. This method is effective for generating ideas, solving problems, and fostering teamwork and innovation within a group. (Wirtz, 2022)

# Project | AI in recruiting



## Biases in identifying hiring needs

### **ANCHORING BIAS**

The person, whose position is to be filled is affecting on expectations and prediction on hiring needs

LOW performer -> Prediction: need of more resources  
HIGH performer -> Prediction: need of less resources

### **STATUS QUO-BIAS**

There might be resistance in changing the type of roles in the organisation even if business needs might state otherwise

### **DEMOGRAPHIC BIAS**

Predictive models might not take demographic shifts, such as aging population, into account

### **FEEDBACK LOOP BIAS**

AI algorithm is learning from it's own past decisions ie. If hiring need data is biased, it might end up creating bigger and bigger bias, based on it's early predictions.

### **INDUSTRY BIAS**

In case of too narrow industry view, the broader future need might be overlooked.

# Project | AI in recruiting



## Biases in preparing the job description

AI created job descriptions are mostly generated to **reduce the human bias** in recruiting -> AI can identify language leading to biases

### **HISTORICAL BIAS**

Job description can be trained based on old job descriptions, which might be biased

### **DEMOGRAPHIC BIAS**

AI can suggest terms such as "ambitious" or "confident leader" targeting the descriptions to certain demographics.

### **ANCHORING BIAS**

In case job descriptions are created based on high-performers core competences, the job descriptions might create bias

# Project | AI in recruiting



## Biases in talent search

### **GENDER, RACIAL and AGE bias**

Pre-selection criteria for the add selected by employer excluding certain demographics.

### **Cost-efficiency in ad-delivery**

can lead into showing job adds to the ones, who most likely click on the add instead to ones, who would be successful in the role.

The **employer's selection of the chosen platform** can be biased, since some platforms can be more used by certain applicant groups.

### **Social media usage differs greatly among people;**

while some might not have time to use eg. LinkedIn at all, some might give false picture of themselves, while others have more possibilities to impact on their digital footprint.

### **User behaviour is reinforcing the algorithms to replicate the patterns**

ie. If certain type of white male open the job ads more frequently, the add will be visible for similar users more frequently.

If **recruiters communicate more with candidates from certain platforms**, the algorithms may increase ad-visibility in those platforms increasing the bias further.

Sheard, 2022; Naakka 2018; Lawton, 2022; Chen 2022; Köchling & Wehner, 2022

# Project | AI in recruiting



## Biases in screening and shortlisting

In general, AI can generate a fairer recruitment environment by reducing unconscious bias, which humans tend to have based on demographics.

If training data exists based on a pool of existing employees or narrowly defined group, it can lead to unintentional discrimination against underrepresented groups.

Aggregation bias can occur if false generalisations are made about entire populations of people; eg, "young people have tech skills" -> leaves older people outside.

Using candidate's digital footprint as part of screening; when algorithm is built on preferences such as "energetic" or "active", might discriminate against disabled individuals.

# Project | AI in recruiting



## Biases in interviewing

AI techniques in interviewing are used mostly in the pre-screening phase.

Final interviews should be done in human-to-human interaction.

In video assessments emotional recognition software can interpret eg. intonations in different languages differently.

Some candidates might not feel confident in front of camera.

Dehumanization?  
Undermining the employee – employer relationship

The use of facial recognition technology can raise privacy concerns and bias against certain groups\*

\* See the example of facial recognition biases

Sheard, 2022; Naakka 2018; Lawton, 2022; Chen 2022; Köchling & Wehner, 2022



# Project | AI in recruiting

## Offer creation & onboarding



To be considered; at what stage is the human-to – human interaction needed? Onboarding is giving a picture of the company.



### **Onboarding**

Training data can be based on historical data and thus carry biases.

### **Onboarding**

AI algorithm creator's personal biases might affect on the educational offering.

Offering tools often use historical data in anticipating the industry standards -> can lead into biases.

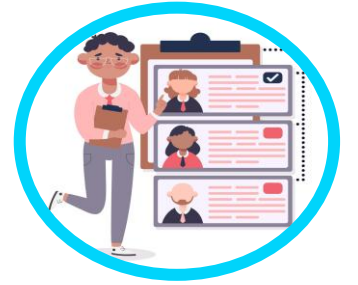
# Project | AI in recruiting



## DATA COLLECTION AND PREPARATION



## Screening and shortlisting the candidates; dataset definition



While the company is looking into all aspects and possibilities in AI recruiting, currently the company is intensively evaluating AI on the screening and shortlisting phase. This involves an in-depth analysis of resume scanning and the alignment of candidates to suitable job vacancies.

By deploying AI tools to automate the resume screening process, the goal is to quickly and precisely pinpoint candidates who fulfill the job requirements. The adoption of such technology is expected to lead to a more strategic use of HR resources, enhance the caliber of candidates making it to the shortlist, and, by aligning candidates' skills and backgrounds with the company's requirements, increase the overall effectiveness of the hiring process.

First, the company needs to define the datasets, which are the basis of the recruiting system and from which the algorithms take the data from.

## Training AI

- Training datasets play a crucial role in the development of algorithms. These datasets are the foundation on which algorithms are built. The purpose of these datasets is to serve as the "ground truth" or factual basis that instructs the algorithm on how to interpret and process large quantities of data.
- They teach the algorithm to understand the relationships between input variables (the information fed into the system) and an output variable (the prediction or decision the system makes based on the inputs).
- Significant challenge arises when the training data itself contains biases. These biases can be a reflection of historical inequalities, prejudices, or simply the result of unrepresentative or incomplete data collection.
- When an algorithm is trained on biased data, it learns these biases and perpetuates them in its predictions and decisions. This phenomenon is commonly described as the "bias in, bias out" problem. It means that if the input data (bias in) is biased, the output of the machine learning model (bias out) will also be biased.
- This can lead to unfair, discriminatory, or flawed outcome affecting which candidates are selected or rejected based on biased criteria rather than their true qualifications or potential.

# Project | AI in recruiting

## Training AI on company's internal and external data

### Internal data

- **Employee Performance Data & Historical Employment Data**
  - Data collected internally on employee performance and past hiring decisions.
- **CVs Submitted Over Time & CVs and Resumes**
  - Historical CVs and resumes submitted directly to the company for job applications.
- **False Negative Cases' Data**
  - Information about applicants wrongly rejected, derived from the company's recruitment processes.
- **Keywords and Phrases from CVs**
  - Specific words or phrases from CVs collected through the company's application process.
- **Educational Background Information**
  - Educational details obtained from candidate applications to the company.

### External Big Data

- **Social Media**
  - Online platforms where individuals share personal interests, experiences, and professional achievements. Social media can provide insights into a candidate's personality, skills, and professional network, making it a valuable source for identifying potential talents.
- **Professional Websites**
  - Websites specifically designed for professional networking and career development, such as LinkedIn. These sites contain detailed profiles including work history, education, skills, endorsements, and professional accomplishments, offering a rich source of information for talent identification.
- **Internet Searches**
  - Broad online searches that companies may conduct to find publicly available information about a candidate. This could include news articles, personal blogs, publications, or any web presence that provides additional context about a person's qualifications, achievements, and character.

Sheard, 2022; Fernandes, 2021; Ma, 2024

# Project | AI in recruiting



## MODEL DEVELOPMENT AND TRAINING

### Job advertisement; key word selection for algorithm

The company has recently published a job description and received hundreds of applications in response.

AI screening and shortlisting prototype will scan applicants' CVs for keywords and other information believed to be relevant to successful hires, such as experience, job titles, previous employers, universities and degrees. Based on this, the system creates a structured applicant profile, and all candidates are scored and ranked.



# Job advertisement and related key words



## WE ARE LOOKING FOR..

...Communications assistant to be responsible for the creation of content such as media releases, blogs, and social media posts on behalf of our company. You will also be monitoring media and campaign coverage and attending internal and external events.

### RESPONSIBILITIES

- Develop and manage diverse content, including media releases, blogs, and social media posts, aligned with the company's strategic goals; monitor media and campaign coverage.
- Support the implementation of internal and external communications strategies and assist in managing the company's image.
- Organize marketing events and provide comprehensive administrative support, including maintaining event calendars and updating contact lists.
- Compile and prepare presentations and reports while tracking project progress and media exposure.

### REQUIREMENTS

- Holds a Bachelor's degree in communications, marketing, or related field, with excellent verbal and written communication abilities.
- Proficient in social media strategies and media relations, with a creative and innovative approach.
- Strong organizational skills and attention to detail, capable of multitasking and maintaining positive interpersonal relationships.
- Skilled in using office management and design software, such as Photoshop and InDesign, and knowledgeable in various social media platforms.

"Bachelor's degree in communications"  
"Bachelor's degree in marketing"  
"Content creation"  
"Media releases"  
"Blogs"  
"Social media posts"  
"Monitoring media" and "campaign coverage"  
"Supporting implementation of communications strategies"  
"Managing company's image"  
"Organizing marketing events"  
"Providing comprehensive administrative support"  
"Maintaining event calendars"  
"Updating contact lists"  
"Compiling presentations" and "reports"  
"Tracking project progress" and "media exposure"  
"Proficient in social media strategies" and "media relations"  
"Creative and innovative approach"  
"Strong organizational skills" and "attention to detail"  
"Capable of multitasking"  
"Maintaining positive interpersonal relationships"  
"Skilled in using office management" and "design software"  
"Photoshop" and "InDesign"  
"Knowledgeable in various social media platforms"

# Project | AI in recruiting

## MODEL DEVELOPMENT AND TRAINING



## Comparing two applicant's CVs towards the key word phrasing

Next, we will focus on analyzing two applications to understand how the AI application prototype evaluates candidates during the screening phase. This analysis will help to identify if there are any biases influencing the selection process.

The company organizes a workshop to explore the details of AI-driven CV screening, using a case study from their own operations.

During the workshop, they analyze two CVs side-by-side to uncover how differences in wording and phrasing can inadvertently lead to overlooking qualified candidates. This exercise aims to illuminate the potential biases embedded within their AI tools and encourages critical thinking about refining the system to fully recognize and appreciate the breadth of an applicant's capabilities.



CV1

CV2



# JOHN DOE

Communications assistant

## EXPERIENCE

**Communications assistant**

**ABC Corporation**

**2019- Present**

- Developed and managed diverse content, including media releases, blogs, and social media posts
- Supported the implementation of comprehensive communications strategies
- Organized multiple marketing events and maintained event calendars
- Compiled presentations and tracked project progress, ensuring alignment with strategic goals

## EDUCATION

**University of communications**

**Bachelor's degree in Marketing**

**2014-2018**

## SKILLS SUMMARY

- Proficient in Photoshop and InDesign
- Skilled in social media strategies and media relations
- Excellent verbal and written communication abilities
- Strong organizational skills and attention to detail

## About Me

A dedicated communications professional with a Bachelor's degree in Marketing, seeking the role of Communications Assistant to leverage extensive experience in content creation, social media management, and event coordination to contribute to the company's strategic goals.



+123-456-7890



hello@doeland.com

123 Anywhere St., Any City



# JANE DOE

Digital Content Creator

## EDUCATION

**CREATIVE UNIVERSITY**

**BSc in Communications Studies 2018**

## WORK EXPERIENCE

**Digital Content Creator; Creative Media Co**  
**2019 -Present**

- Spearheaded content initiatives across digital channels, crafting engaging articles and dynamic social engagements
- Drove strategy for public relations efforts, enhancing brand visibility
- Led the logistics for promotional and networking gatherings, ensuring smooth execution
- Synthesized data into compelling reports and visuals for team updates

## COMPETENCIES

- Advanced user of digital design tools and content management systems
- Crafted engaging online dialogue and cultivated media partnerships
- Articulate communicator, both in written formats and oral presentations
- Excelling in project orchestration and meticulous in administrative tasks

## PROFILE

Passionate communicator with an academic background in Communications Studies and hands-on experience in crafting engaging narratives and managing digital platforms. Eager to bring my toolkit of creative dissemination and stakeholder engagement to the Communications Assistant position.

## CONTACT ME



(123) 456-7890



Jane@doe.com



123 Anywhere St.,  
Any city, State,  
Country 12345

This bolding helps to see the connections between specific wording in CVs and keyword-based AI screening tools.



# JOHN DOE

Communications assistant

## EXPERIENCE

Communications assistant  
ABC Corporation  
2019- Present

- Developed and managed diverse content, including **media releases, blogs, and social media posts**
- Supported the implementation of comprehensive communications strategies
- Organized multiple **marketing events** and **maintained event calendars**
- Compiled **presentations** and tracked **project progress**, ensuring alignment with strategic goals

## EDUCATION

University of communications  
**Bachelor's degree in Marketing.**  
2014-2018

## SKILLS SUMMARY

- **Proficient in Photoshop and InDesign**
- Skilled in **social media strategies** and **media relations**
- Excellent verbal and written communication abilities
- **Strong organizational skills** and **attention to detail**


About Me

A dedicated communications professional with a **Bachelor's degree in Marketing**, seeking the role of Communications Assistant to leverage extensive experience in **content creation, social media management**, and event coordination to contribute to the company's strategic goals.

+123-456-7890

hello@doeland.com

123 Anywhere St., Any City



# JANE DOE

Digital Content Creator

## EDUCATION

CREATIVE UNIVERSITY  
BSc in Communications Studies 2018

## WORK EXPERIENCE

Digital Content Creator; Creative Media Co  
2019 -Present

- Spearheaded **content** initiatives across digital channels, crafting engaging articles and dynamic social engagements
- Drove strategy for public relations efforts, enhancing brand visibility
- Led the logistics for promotional and networking gatherings, ensuring smooth execution
- Synthesized data into compelling **reports** and visuals for team updates

## COMPETENCIES

- Advanced user of digital design tools and content management systems
- Crafted engaging online dialogue and cultivated media partnerships
- Articulate communicator, both in written formats and oral presentations
- Excelling in project orchestration and meticulous in administrative tasks

## PROFILE

Passionate communicator with an academic background in Communications Studies and hands-on experience in crafting engaging narratives and managing digital platforms. Eager to bring my toolkit of creative dissemination and stakeholder engagement to the Communications Assistant position.

## CONTACT ME

(123) 456-7890  
Jane@doe.com  
123 Anywhere St.,  
Any city, State,  
Country 12345

This table illustrates the impact of wording and phrasing differences in two CVs and how they can affect the outcome of an AI screening process. Despite possessing relevant qualifications, the second CV may be overlooked due to its descriptions not aligning closely with the specific keywords established by the company for screening. This demonstrates the importance of matching the language in a CV to the keywords expected by an AI system.

| ASPECT                        | JOHN DOE   | JANE DOE  |
|-------------------------------|--|---|
| Job titles and terminology    | Uses standard terms like "Communications Assistant"                  | Uses titles like "Digital Content Creator"  |
| Academic background           | Explicitly mentions "Bachelor's Degree in Marketing"                 | States "BSc in Communications Studies"  |
| Describing tasks and skills   | Direct match with keywords like "Content creation", "Media releases" | Uses varied language like "crafting engaging narratives", "spearheaded content initiatives" |
| Technical and software skills | Specifies "Photoshop" and "InDesign"                                 | General mention of "digital design tools and content management systems"                    |
| Direct keyword matches        | Closely matches many specified keywords                              | Uses different phrasing that may not match the specific keywords set for screening          |

# Project | AI in recruiting



## MODEL EVALUATION & TESTING



## Testing the fairness of the solution

Evaluating the algorithm prototype revealed inconsistencies in CV analysis, prompting refinements for a more comprehensive understanding of CV attributes. The model is now prepped for a testing phase prior to real-world deployment.

While several aspects require assessment, the focus here will be on testing the application's fairness to determine the extent of any biases incurred during the process. Fairness metrics are employed to measure the level of bias present in the application's decisions.



### TOOL

There are different tools to examine, report and mitigate discrimination and bias in machine learning models. Check different fairness tools from the link.

[AI Auditing Tools: Empowering Systems with Best 6 Solutions - HyScaler](#)



# Project | AI in recruiting

## Testing the fairness

To create systems that are equitable for all and free from biases, like gender or racial discrimination, it's important to evaluate the solution's performance across diverse population segments.

While there are various methods to assess a solution's fairness, in this context of recruitment, we will focus exclusively on binary classification tests.

Defining fairness in mathematical terms has proven challenging, leading to a lack of consensus on standard formulations. However, most definitions of fairness coalesce around several key concepts:

- Unawareness
- Demographic parity
- Equalized odds
- Predictive rate parity
- Individual fairness
- Counterfactual fairness

**Demographic parity, predictive rate parity and equalized odds** are typically grouped under the umbrella of "group fairness." This concept will be explored further in this section, particularly through the lens of AI in recruiting.

# Project | AI in recruiting

## Confusion matrix

One of the binary classification methods, is the confusion matrix.

Confusion matrix summarizes the predictions made by an algorithmic model in comparison to the actual outcomes, on which it was trained. It displays the counts of both correct and incorrect predictions, providing a clear view of the model's performance and its explainable **accuracy**.

Confusion matrix in the recruitment example

|        | Unqualified   | Qualified   |
|--------|---|---|
| Reject | True Negative (TN)<br>= rejected unqualified candidates | False Negative (FN)<br>= rejected, qualified candidates |
| Hire   | False Positive (FP)<br>= hired, unqualified candidates  | True Positive (TP)<br>= hired, qualified candidates     |

(Wortman Vaughan & Wallace, 2022; Zhong, 2018, Saplicki, 2022)

# Project | AI in recruiting

## Accuracy

In the recruiting example, we illustrate the evaluation of fairness using gender as the demographic variable for assessment assuming there are 100 female and male applicants.

| MEN    | Unqualified    | Qualified      |
|--------|----------------|----------------|
| Reject | <b>15 (TN)</b> | 5 (FN)         |
| Hire   | 20 (FP)        | <b>60 (TP)</b> |

| WOMEN  | Unqualified    | Qualified      |
|--------|----------------|----------------|
| Reject | <b>60 (TN)</b> | 20 (FN)        |
| Hire   | 5 (FP)         | <b>15 (TP)</b> |

Accuracy

$$\frac{(TP + TN)}{(TP + FP + TN + FN)}$$

In both male and female categories, 75 applicants are classified correctly  
-> ie. **accuracy is similar in both demographics**



# Project | AI in recruiting

## Demographic parity

Building on the data in the confusion matrix, demographic parity is a criterion for fairness where the probability of a favourable outcome- in our example hiring- is not influenced by a protected characteristic- like gender.

| MEN    | Unqualified    | Qualified      |
|--------|----------------|----------------|
| Reject | 15 (TN)        | 5 (FN)         |
| Hire   | <b>20 (FP)</b> | <b>60 (TP)</b> |

| WOMEN  | Unqualified   | Qualified      |
|--------|---------------|----------------|
| Reject | 60 (TN)       | 20 (FN)        |
| Hire   | <b>5 (FP)</b> | <b>15 (TP)</b> |

In the example scenario of recruiting, if 80 out of 100 male applicants and only 20 out of 100 female applicants are hired, demographic parity is not met. However, this might be acceptable if the male applicants are indeed more qualified than the female applicants, like it seems to be in this case.



While enforcing group level fairness; this can be unfair to individuals since it could drop out otherwise qualified candidate just to achieve the demographic parity

# Project | AI in recruiting

## Predictive parity

Unlike demographic parity, predictive parity considers both the classifier's decisions and the true outcomes. It assesses whether the likelihood of an applicant being suitable for a position is consistent across different groups, provided the AI has selected them for hiring. The expectation is that this probability should be nearly identical for all groups considered.

| MEN    | Unqualified | Qualified |
|--------|-------------|-----------|
| Reject | 15          | 5         |
| Hire   | 20          | <b>60</b> |

| WOMEN  | Unqualified | Qualified |
|--------|-------------|-----------|
| Reject | 60          | 20        |
| Hire   | 5           | <b>15</b> |

In the recruiting example the system meets the criteria for predictive parity as three-quarters of the hired applicants from both groups are indeed qualified .

Castelnovo et al., 2022; Wortman Vaughan & Wallace, 2022

# Project | AI in recruiting

## Equalized odds; False positive rate & False negative rate

The metric known as the **false positive rate** asserts that the likelihood of an unqualified candidate being erroneously hired should be consistent across all demographics.

| MEN    | Unqualified | Qualified |
|--------|-------------|-----------|
| Reject | 15          | 5         |
| Hire   | 20          | 60        |

| WOMEN  | Unqualified | Qualified |
|--------|-------------|-----------|
| Reject | 60          | 20        |
| Hire   | 5           | 15        |

The metric known as the **false negative rate** asserts that the likelihood of and qualified candidate being erroneously rejected should be consistent across all demographics.

| MEN    | Unqualified | Qualified |
|--------|-------------|-----------|
| Reject | 15          | 5         |
| Hire   | 20          | 60        |

| WOMEN  | Unqualified | Qualified |
|--------|-------------|-----------|
| Reject | 60          | 20        |
| Hire   | 5           | 15        |

**Equalized odds** combines these two attributes by satisfying both false negative rate balance and false positive rate balance.

In the recruiting example the solution **does not meet the criteria**, since the unqualified male applicants are much likely to be hired than the unqualified female applicants AND the qualified female applicants are much likely to be rejected than the male applicants.

# Project | AI in recruiting

## Considerations around the fairness metrics

- A system cannot achieve predictive parity, false positive rate balance, and false negative rate balance all at once due to mathematical constraints.
- Trade-offs between different fairness objectives are inherent and must be managed thoughtfully.
- Fairness considerations must be contextualized within the specific application they are applied to.
- Not all aspects of fairness are quantifiable, acknowledging the socio-technical complexity of fairness.

# Project | AI in recruiting

## Individual fairness

A



Bachelor degree  
1 year related work experience

B



Master Degree  
1 year related work experience

C



Master degree  
No related work experience

- Individual fairness differs from group-based fairness (the previously discussed metrics) criteria by focusing on the treatment of individuals.
- This concept asserts that individuals with similar characteristics should receive comparable outcomes.
- Determining a metric to measure individual similarity is complex.
- Consider three job candidates; is A more capable of doing the job than C?

# Project | AI in recruiting



## MODEL DEPLOYMENT



## Informing applicants about using AI in recruiting

Following careful internal evaluations, the company is set to launch its screening tool selectively on open roles to gather initial user and applicant feedback.

It's crucial for applicants to be continuously aware that they're engaging with an AI-driven recruitment system, fostering trust. Awareness about the advantages of AI tools should be established upfront, enhancing applicants' willingness to use interactive technology like chatbots. Additionally, refining AI recruitment processes and ensuring clarity about the system's workings are key to its adoption and success among candidates.

Also, it should be made sure that there is a human overseeing and intervening when needed.

# Project | AI in recruiting

## Informing applicants about using AI in recruiting | Tool examples

### UNESCO's Ethical Impact Assessment

An example from Unesco's Ethical impact assessment tool presented earlier

#### 10.2.1. System Awareness:

- 10.2.1.1. Are users made fully aware when they are interacting with an AI system, as opposed to a human being?
- 10.2.1.2. Are individuals (directly or indirectly) impacted by the AI system made fully aware of when a decision (that impacted them) was informed by or made on the basis of an AI system or AI algorithms?
  - 10.2.1.2.1. Are they made aware of the extent to which they are impacted, including the rationale, benefits and limitations of the decision(s)?
- 10.2.1.3. Have appropriate explanations been put in place to help users and other impacted individuals understand the decision-making process or how the system works when required?
- 10.2.1.4. Have appropriate explanations been put in place to help the government bodies in charge of regulation understand the decision-making process or how the system works when required?
- 10.2.1.5. Has the decision to adopt the AI system been documented and communicated online?
- 10.2.1.6. Can the AI system make any decisions which the physical persons or legal entities in charge of the system lack expertise or competence to critique, modify or override?

### Eccola Cards; a method for implementing ethically aligned AI systems

Data

#### #7 Privacy and Data

**Motivation:** Privacy is a rising trend in the wake of various recent data misuse reveals. People are now increasingly conscious about handing out personal data. Similarly, regulations such as the GDPR now affect data collection.

**What to Do:** Ask yourself:

- What data are used by the system?
- Does the system use or collect personal data? Why? How is the personal data used?
- Do you clearly inform your (end-)users about any personal data collection? E.g., ask for consent, provide an opportunity to revoke it etc.
- Have you taken measures to enhance (end-user) privacy, such as encryption or anonymization?
- Who makes the decisions regarding data use and collection? Do you have organizational policies for it?

**Practical Example:** Rather than collecting and selling data, appealing to privacy can also be profitable. Regulations are making it increasingly difficult to collect lots of personal data for profit. Privacy can be an alternate selling point in today's climate.

**ECCOLA**

©10-4001-20200415

Unesco, 2023; Vakkuri et al. 2021



# Project | AI in recruiting



## OPERATION AND MONITORING

### After deployment

While internal assessments of fairness were conducted before launch, the company must also be ready for independent external audits.

It's also advantageous to continuously seek feedback post-deployment and establish ongoing engagement channels for stakeholders. This involves keeping workers informed and involved in consultations and participative processes during the entire AI system implementation within the organization.



# THANK YOU

Project number: 2022-1-ES01-KA220-HED-000085257



The European Commission's support for the production of this publication does not constitute of the contents, which reflect the views only of the authors , and the Commission cannot be held responsible for any use which may be made of the information contained therein.

