Fundamental Rights Impact Assessment (FRIA) – Algorithmic hiring for XYZ

Example FRIA by

Sieuwert van Otterlo
Pavlo Burda

About this template/example

This template was created by the people of ICT Institute. You can find the latest version and other templates here: https://ictinstitute.nl/free-templates/

This template has been developed using the following sources:

- AI Act Guide v1.1 (2025) - https://www.government.nl/documents/publications/2025/09/04/ai-act-guide
- FRAIA 2021 - https://www.government.nl/documents/reports/2021/07/31/impact-assessment-fundamental-rights-and-algorithms
- EU High-Level Expert Group on AI: Ethics Guidelines for Trustworthy AI (2019) - https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 - https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419
- AI Act Articles 6, 14, 26, 27
- GDPR Articles 5, 9, 35
- Algorithmic hiring vendors: Pinpoint (https://www.pinpointhq.com/) and Workable Recruiting (https://www.workable.com/)
- P. Burda and S. van Otterloo, Fairness definitions explained and illustrated with examples, CRSJ, 2025 - https://ictinstitute.nl/fairness-definitions-explained-illustrated/
- AI in hiring, Wikipedia - https://en.wikipedia.org/wiki/Artificial_intelligence_in_hiring
- Fairness in ML - https://en.wikipedia.org/wiki/Fairness_(machine_learning)

Mandatory elements of a FRIA

The FRIA is necessary for high-risk AI systems (see AI Act Guide or AI Act Art. 3). A FRIA must contain, based on article 27 of the AI Act and the previously mentioned guidelines, the following elements:

- Organisation Details
- Description of the process (AI Act 27a)
- Description of the groups of people affected (AI Act 27c)
- Description of the timeline (AI Act 27b)
- Assessment of specific risks to fundamental rights (AI Act 27d)
- Description of human oversight measures (AI Act 27e)
- The measures to address the risks (AI Act 27f)

Contents

Organization details

@Note: Write down here on what day with whom has been spoken, which workshops have been done when, and what parties/which people were present.

Name and address of the organization: Secure Force BV, Jaarbeursplein 1, Utrecht

Author FRAIA:

Other involved and consulted experts: : e.g, DPO, CTO, System/IT dept, HRM owner

Plan of action for the FRIA execution: e.g., scheduling meetings/workshops, iterations

Description of the process

@Answer:
What does the AI system do and in which process step is it used?
What inputs does it use and what outputs does it produce?
What is the legitimate purpose (why is AI used instead of a manual process)?

Secure Force BV is a private physical security provider for industrial and commercial real-estate. The current problem the company is facing is dealing with a large amount of job applicants and the need to provide a quick, consistent and objective assessment to each candidate. Therefore, the company deploys a third-party machine-learning-based assistant (AI system) to support pre-screening job applicants for physical security guard positions.
The AI system predicts candidate suitability based on a test score (e.g., physical speed test) and contextual data (e.g., proximity to work location). The AI supports human recruiters in ranking applicants but does not replace them entirely (human-in-the-loop).

Commercial tools such as Workable and Pinpoint use AI systems to parse and score CVs, rank candidates, and in some cases evaluate video interviews, which illustrates that our hiring system falls squarely within the same class of AI-assisted screening tools that must be assessed for impact on fundamental rights.

The system's necessity is based on efficiency and consistency goals: faster shortlisting, reduced administrative workload and retaining a good accuracy in terms of hiring suitable candidates. The use of the AI system allows to make the hiring process more efficient by helping by helping HR staff to process a large amount of applications and hire suitable candidates more easily.

The main ML technique behind the AI system is a
[NEURAL NETWORK]
feed-forward fully connected neural network for binary classification with four input features, such as physical test-score, age, gender and residence place, and approximately 200 parameters.
[LOGISTIC REGRESSION]
logistic regression model for binary classification with four input features, such as physical test-score, age, gender and residence place, and output a probability for each class.

Alternatives (manual screening) were considered but are slower and less scalable for large recruitment volumes. The system is proportionate if fairness - impartial and just treatment without discrimination – together with other trustworthy AI requirements are maintained.

The hiring workflow:

1. Applicants submit the CV online and complete a physical test (speed and strength)
2. The AI system predicts suitability ('to hire'/'not to hire') based on previous and new data
3. HR reviews AI recommendations before final decisions
4. Logs stored according to data retention policy (6 months).

Note: the example problem of this FRIA is inspired by the example in the paper ""Fairness definitions explained and illustrated with examples": https://doi.org/10.54822/PASR6281

Definitions

@Answer:
Why your system qualifies as an AI system?
Why is your system a (not) high-risk?
Is your organization a provider or deployer? If deployer, who is the provider?

The automated hiring helper is a machine-based system with varying levels of autonomy, therefore is classified as an AI system according to the Article 3(1) of the AI Act.

The AI system profiles the individuals by processing personal data and can have influence on people's recruitment and access to work. Therefore, falls under the high-risk application area 'Employment, workers' management and access to self-employment' (AI Act Art. 6, Annex III). A FRAIA is therefore required under article 27 of AI Act.

Secure Force BV is a deployer since it will use the system. The system is provided by an external provider: [PROVIDER]. Secure force BV does not rebrand or re-sell the service. According to Article 3(4) of the AI Act, Secure Force BV is a deployer of the AI system.

Affected group of people

@Answer:
Which people are directly affected (e.g. applicants, employees)?
Which groups are particularly at risk (age, gender, location, other)?
Which internal users are affected (e.g. HR, managers)?

The main group affected by the AI hiring assistant are job applicants for physical security-guard roles, whose access to work, perceived fairness and privacy can be directly influenced by AI-supported screening. These people are at risk of incorrectly not being hired if the system would not work correctly.

Within the organisation, HR staff are the primary end users: they rely on the system's recommendations, interpret its outputs and remain responsible for final decisions.

Management is affected through staffing, compliance and reputational risk.

Timeline

@Answer:
From when to when will the system be used?
How often is it used (per case / day / campaign)?

How often will the model or configuration be updated?

The AI hiring system is used on an ongoing basis for all new applications to security-guard roles, typically several times per week depending on vacancy volume. It is intended to remain in operation for an initial period of [THREE YEARS], with an [ANNUAL] review by management.

Key performance and fairness metrics are monitored at least [QUARTERLY], or sooner if major changes occur in data, model or process. The model may be retrained or updated up to [TWICE PER YEAR], and any significant change triggers a review and, where necessary, an update of this FRIA.

Risks to fundamental rights

@Answer;
Which of the fundamental rights of the interested parties can be affected?
For each right, what could realistically go wrong?
How severe and how likely is this, based on your context and data/model results? Provide an explanation of why a severity is high, medium or low.

The hiring algorithm can influence several fundamental rights of applicants protected under the EU Charter, including the right to equality and non-discrimination, the right to privacy, the right of access to work and human dignity. Because the system supports employment decisions, its outcomes have a direct and personal effect on applicants. Other interested parties are not foreseen to be affected.

Fundamental right	**Risk** (What is the risk?)	**Risk severity** (severity and chance of event)
Equality & non-discrimination	Past data and model correlations could systematically disadvantage candidates based on age, gender, or potentially location treating them differently (e.g., bias, discrimination by proxy). [DERIVE FROM YOUR OWN DATA AND MODEL]	High because … Systemic exclusion and fewer opportunities for specific groups (e.g., women and older candidates). Use of historic data and proxy variables make bias likely (e.g., training data is unbalanced, test scores can correlate with age or gender). [APPLY YOUR OWN BIAS ASSESSMENT]
Privacy & data protection	Sensitive data processed and shared with vendor without clear necessity (e.g., training on/sharing sensitive attirbutes)	Medium because … Exposure of age, gender and location, and potentially coupled with other personal data [(NAMES, EMAILS)] can lead to identity or profiling risks.
Access to work and human dignity	Automated screening could unfairly exclude qualified candidates from jobs, especially if HR overseers are untrained or rely too	High because … Hiring outcomes directly affect access to work and livelihood, so wrongful exclusion has serious consequences for candidates. While human-in-the-loop review reduces

	heavily on biased or opaque model logic (e.g., untrained HR overseer, lack of transparency).	the likelihood, the risk remains if oversight is weak or [NO AI-RELATED TRAINING] poorly trained.
Right to explanation, fair procedure & effective remedy	No clear appeal mechanisms (e.g., lack of information on how to appeal). Complex or black-box model can make giving meaningful explanations harder (e.g., deep neural networks).	Low because … Applicants are explicitly informed in terms and conditions that AI assists the hiring process and may request an explanation or human review upon a decision. HR users have access to feature-importance explanation techniques to interpret recommendations. The simple nature of the model ([LOGISTIC REGRESSION/NEURAL NETWORK] with 200 parameters and 5 layers) makes it easy to derive explanations.
Societal and environmental effects	Reduce inclusion by excluding candidates by location. Unnecessary consumption of computing resources.	Low because … The work nature excludes remote working and is limited by commute distance. [LOGISTI REGRESSION] The model is lightweight with very low compute and energy use. [NEURAL NETWORK] Resources consumption minimal as no large datasets ('big data') are used, nor compute-intensive training is performed (such as LLMs). The system is run on locally/energy-efficient shared infrastructure.

Detailed analysis of bias, discrimination and fairness.

Since bias and discrimination is a major risk for algorithmic hiring , Secure Force BV asked two data scientists to do a thorough analysis of bias and fairness using various fairness metrics. The full analysis is in a separate paper (https://doi.org/10.54822/PASR6281).
A detailed analysis is available in Appendix 1.

Overall, the analysis shows that the current algorithm provides the most acceptable compromise between classification performance and fairness satisfiability. The chosen algorithm shows the least amount of bias and is most fair while still having acceptable performance.

Note that there is some biases. Thefore use of the system requires ongoing bias monitoring and human oversight.

Human oversight measures

@Answer to:
What kind of oversight is there?
How can affected people appeal, ask for human review, or complain?
Who is responsible (HR reviewers, system owner, DPO)?

The purpose of the AI system is to support human recruiters in ranking applicants, not to replace them. All AI outputs are advisory: a trained HR officer always reviews recommendations when screening applicants and can override them at any time.

Human oversight is implemented through several measures:

- Human-in-the-loop review and appeals: every hire/no-hire recommendation is checked by HR staff. Applicants are informed that AI assists the process and can request re-evaluation by HR, as well as an explanation of the decision. HRM owner is responsible.

- Pause, rollback and retraining: if anomalies or bias are detected, the system can be paused and the process reverts to full manual screening. If significant disparities are found (for example, a gap of more than ~10% across relevant fairness metrics), the model is retrained, adjusted or suspended. The AI system owner is responsible.

- Training and role definition: there remains a risk of weak oversight if staff are poorly trained. Therefore, HR reviewers using the system receive specific training on its purpose, limitations, fairness risks and override options. Their role explicitly includes handling applicant appeals and documenting overrides. HRM owner isresponsible.

- Escalation and complaints: in addition to HR review, a clear complaints procedure is in place: applicants can escalate concerns to the DPO and, if needed, to the supervisory authority (AP). If any, DPO/Legal is responsible.

Mitigation and Safeguards

@Action:
What are the measures that eliminate or mitigate the risks?
What do you do if problematic, unfair results or incidents are found?
Who decides and documents these actions?

[CHOOSE ONE, SHORT]
The risk and fairness assessment findings justify the mitigation measures and human-oversight controls described in the Risks to fundamental rights section and form the documented evidence for this assessment under Article 27(f) AI Act.

A detailed assessment of the trustworthy AI requirements according to the Ethics Guidelines For Trustworthy AI (2019) is available in Appendix 2.

Risk	**Mitigation Measure**
Age/Gender bias	Retrain model excluding sensitive attributes; test periodically or upon changes.

Insufficient transparency and explanation	Inform applicants of AI use and purpose.
Unclear appeal mechanism	Add human review request feature. Assign process owners (HRM, AI system, DPO/legal).
Insufficient human oversight	Assign trained/train HR reviewers.
Additional mitigations	
Data protection	Perform GDPR DPIA; minimize data retention (logs kept 6 months).

[CHOOSE ONE, LONG]
The risk and fairness assessment findings justify the mitigation measures and human-oversight controls described in the Risks to fundamental rights section and form the documented evidence for this assessment under Article 27(f) AI Act.

A detailed assessment of the trustworthy AI requirements according to the Ethics Guidelines For Trustworthy AI (2019) is available in Appendix 2.

Risk	**Mitigation measure**	**Residual risk**
Past data and model correlations could systematically disadvantage candidates based on age, gender, or potentially location treating them differently (e.g., bias, discrimination by proxy).	Use a model configuration that excludes age and gender from the production model (e.g. A1); perform regular fairness testing on age and gender; set thresholds for acceptable gaps (e.g. >10% triggers action); retrain, adjust or suspend the model if unfair patterns are detected.	Moderate- proxy effects (e.g. test scores) and small sample sizes can still create unequal outcomes between groups, even with monitoring; some disparity remains possible between audits.
Sensitive data processed and shared with vendor without clear necessity (e.g., training on/sharing sensitive attirbutes)	Perform a GDPR DPIA; apply data minimisation (only test results and necessary contact data in production); pseudonymise logs and limit retention (e.g. 6 months); put a Data Processing Agreement (DPA) in place with the supplier and avoid sharing sensitive attributes unless strictly necessary; evaluate getting ISO 27001 certified to control risks.	Low - controls reduce likelihood and scale of misuse. Breaches or misconfiguration can still occur.
Automated screening could unfairly exclude qualified candidates from jobs, especially if HR overseers are untrained or rely too heavily on biased or opaque model logic (e.g., untrained HR overseer, lack of transparency)	Implement/evaluate human-in-the-loop review for all important decisions; require training for HR reviewers on system limitations and bias; allow HR to override AI recommendations; define pause/rollback procedures and retraining if fairness gaps between groups exceed a set threshold.	Low - clear oversight procedures and trained staff reduce the chance of systemic exclusion. Minor risk remains due to time pressure, and huma errors.

No clear appeal mechanisms (e.g., lack of information on how to appeal). Complex or black-box model can make giving meaningful explanations harder (e.g., deep neural networks).	Inform applicants that AI assists the hiring process and that they can request human review and an explanation; ensure HR has access to feature-importance / explanation tools; keep the model sufficiently simple to allow case-level explanations; provide clear complaints channels via HR, the DPO and the supervisory authority (AP).	Low - explanations and appeal routes are available.
Reduce inclusion by excluding candidates by location. Unnecessary consumption of computing resources.	Limit use to roles where location is genuinely relevant and evaluate monitoring for location-based disparities; [LOGISTIC REGRESSION/SMALL NEURAL NETWORK] use a lightweight model trained on small datasets if possible; [BIG MODEL] run it on energy-efficient shared or local infrastructure to keep compute and energy use low.	Low - minor effects on inclusion given the limited scope and low compute footprint.

[CHOOSE ONE, DISCURSIVE]
This section translates the risks identified in the Risks to fundamental rights section into specific safeguards aligned with the EU HLEG Ethical Guidelines for Trustworthy AI (2019). Each subsection corresponds to one of the requirements and outlines the concrete measures Secure Force BV applies to ensure that the high-risk AI hiring system remains lawful, ethical, and trustworthy under Articles 9 and 27 AI Act.

A detailed assessment of the trustworthy AI requirements according to the Ethics Guidelines For Trustworthy AI (2019) is available in Appendix 2.

The mitigation measures follow the EU HLEG Ethical Guidelines for Trustworthy AI, ensuring that the Secure Force BV hiring assistant remains lawful, fair, and human-centred.

1. Human agency and oversight - All hiring decisions remain under human control. A HR officer reviewing AI recommendations when screening applicants shall be trained. They may override AI decisions, and handle applicant appeals.
2. Technical robustness and safety - The model is tested for accuracy and stability before use. A fallback to full manual screening is in place, and logs should be kept for six months to trace errors or anomalies.
3. Privacy and data governance - Only relevant test data are used. Sensitive fields (age, gender) are excluded from final models, and data are processed under GDPR (Art. 6 and 9) with retention limits. A (pre-) DPIA scan shall be carried out.
4. Fair procedure and transparency - Applicants are informed in terms and conditions upon applying that AI assists recruitment. HR users can access feature-importance explanations, and model versions versioning is planned for audit and traceability.
5. Diversity, non-discrimination and fairness - Group-fairness and error-rate metrics are checked upon re-training or changes. Retraining is required if bias/fairness metric exceeds [10%] and/or accuracy varies by [10%].

6. Societal and environmental effects - The AI system is used only for legitimate hiring, monitored for unintended exclusion of groups. The AI system resources consumption is minimal as no large datasets ('big data') are used, nor particularly compute-intensive training is performed (such as LLMs). The system is run on energy-efficient shared infrastructure.
7. Accountability - Ownership is assigned to [System Owner and DPO]. The system is in the scope of annual internal audit reviews for compliance [DEFINE YOUR OWN]; applicants can file complaints or request human review at any time [on the public website].

Conclusion

The hiring assistant qualifies as high-risk under article 3 and Annex III (Employment) of the AI Act. This FRIA identifies severe risks related to Equality & non-discrimination and Access to work and human dignity. Mitigation have been put in place to an acceptable residual risks level. Secure Force BV commits to ongoing fairness monitoring, re-assessment and huma oversight.

Appendix 1 – Detailed fairness assessment

@Note: provide a technical analysis of your AI system in terms of

Context

The main ML technique behind the AI system is a
[NEURAL NETWORK] feed-forward fully connected neural network for binary classification with four input features, such as physical test-score, age, gender and residence place, and approximately 200 parameters.
[LOGISTIC REGRESSION] logistic regression model for binary classification with four input features, such as physical test-score, age, gender and residence place, and output a probability for each class.
Give the model architecture, there is

Secure Force BV assessed the AI system by means of group-fairness metrics employing a [HISTORIC/SYNTEHTIC] dataset of candidate applications. The concept of fairness in ML - impartial and just treatment without discrimination - encompasses the ethical values highlighted in the Ethical Guidelines for Trustworthy AI and, therefore, can guide the assessments of risks and adherence to the guidelines. The dataset includes the relevant input features, the classification outcomes and the ground truth of successful hires.

The group-fairness metrics sourced from research and practice literature are:

Metric	**Measures Approximate Equality of …**
Group Fairness / Statistical Parity	Selection rates between protected (p) and unprotected (u) groups
Predictive Parity (Precision Parity)	Accuracy of positive predictions across groups
Predictive Equalit / (False-Positive Rate (FPR) Parity	False positive errors between groups
Equality of Odds (FPR=TPR)	Combined error rates balance

The employed dataset variables in the assessment are:

Feature	**Type**	**Description**
age	Sensitive	Candidate age (20–49 years)
gender	Sensitive	Male / Female
testresult	Feature	Combined physical strength and speed test (0–2)
livesnear	Potentially sensitive	1 = lives near job site
should_hire	Ground truth	1 = candidate suitable for hire

The AI system configurations are [IF MULTIPLE CONFIGURATIONS ARE CONSIDERED]:

Configuration	**Description**	**Main Features Used**	**Description**
A1	Weight only test score	Physical Test result	[LOGISTIC REGRESSION]/[NEURAL NETWORK] ML model predicting hire from physical-test score.
Human-expert baseline	Reference of human-in-the-loop	Subjective evaluation using manual expert judgment	Reference decision pattern.

The figures below illustrates the hiring outcomes for the assessed configurations (human expert reference and A1 algorithm).

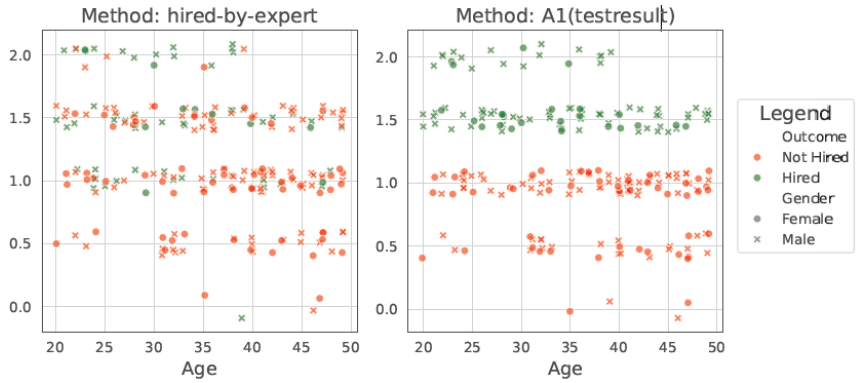

Figure 1.Scatterplot of hiring outcomes for hiring-by-expert and A1 against test score, age and gender. The markers are coded by shape (female/male) and colour (hired/not hired).

The figure below shows the distribution dataset features. It illustrates a clear unbalance towards male candidates in the bracket 30-40 years old.

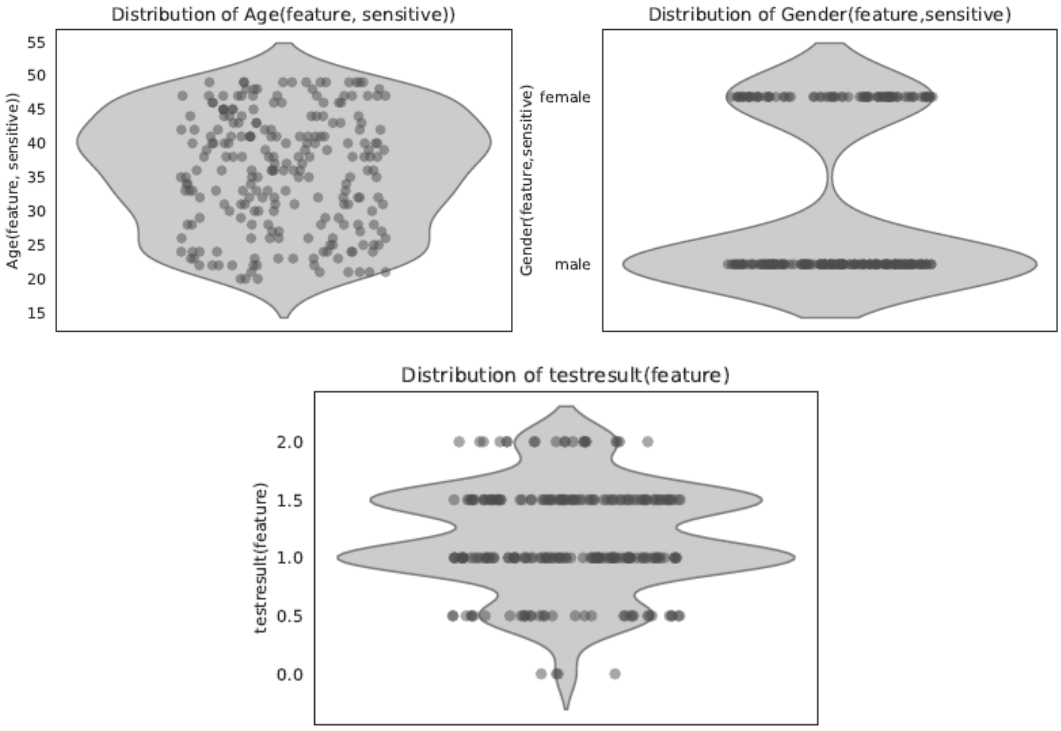

Figure 2. Distribution of Age, Gender and test result. There is notable prevalence of male applicants, slightly over 30 years old candidates.

Fairness results

The analysis results show that the current configuration (A1) is the most balanced. The reference results (the ground truth) are represented by the 'hiring expert' case (best achievable result). The table below reports the summary of the analysis on bias in with respect to the protected attributes (age bias is represented with 'over 40 years old' and gender bias as 'female').

Fairness results summary [CHOOSE ONE, SHORT]

Protected group (specific bias)	**Main finding**	**Fairness status**
Age > 40	Lower selection under configurations using age, and age and gender: systematically excluded; Hired-by-expert and A1: best performance	A1: best configuration
Gender = female	Lower selection under A1 and configurations using age, and age and gender: often excluded; Hired-by-expert, configuration using age and feature correction: best performance	Configuration with age and feature correction: best configurations

Fairness result summary: [CHOOSE ONE, LONG]

Protected Attribute	**Configuration**	**Statistical Parity**	**Pred. Parity (PPV)**	**Equality of Odds (TPR= FPR)**	**Pred. Equal. (FPR Parity)**	**Finding**
Age > 40	hired-by-expert	~✗ (moderate gap)	~✗	✓ (small gap)	✓	Human experts show mild age preference but within acceptable variance.
	A1 (test-only)	✗ (large gap)	✗	~✗	✓	Slight under-selection of older applicants; overall moderate parity.
	Configuration with age	✗	✗	✗	✗	Strong bias toward younger candidates.
	Configuration with age and gender	✗	✗	✗	✗	Strong bias toward younger candidates.
	Configuration with sensitive features corrected (positive discrimination)	✓	✗	~✗	✓	Improves representation, but is biased towards unprotected category.
Gender = Female	hired-by-expert	~✗	✓	✓	✓	Human expert slightly favour men in selection rate but maintain balanced error rates.
	A1 (test-only)	~✗	✓	~✗	~✗	Moderate under-selection of women; low FPR disparity.
	Configuration with age	✓	✗	✓	✓	Balanced gender outcomes overall.

	Configuration with age and gender	✗	✓	~✗	~✗	Strong bias toward male candidates.
	Configuration with sensitive features corrected (positive discrimination)	✓	✓	✗	✓	Fairer selection rate for women but biased towards unprotected category.

Accuracy and precision results

The evaluated system configurations have the following performance (accuracy, precision and recall):

Configuration	Precision	Accuracy	Recall
hired-by-expert	0.86	0.87	0.72
A1	0.59	0.74	0.66
Conf. with age	0.68	0.78	0.65
Conf. with age and gender	0.70	0.80	0.71
Conf. corrected features	0.54	0.69	0.49

- Hired-by-experts is best attainable result in terms of performance and mixed fairness that depends heavily on which metric and which group to assess.
- A1 has overall a modest performance and a moderate under-selection on age and gender, demonstrating an adequate trade-off in terms of performance and fairness.
- Conf. with age has reasonable accuracy/precision, comparatively good gender fairness, but strong age discrimination.
- Conf. with age and gender is technically strong on accuracy and some fairness metrics, but largely problematic because it disadvantages female and/or older candidates.
- Conf. corrected features is the worst on performance, but the most fair on groups (by design). Moreover, there is a bias against of the unprotected group (i.e., male and younger candidates) which remains a problematic.

From the candidate perspective, A1 (lower precision and relatively high recall) is the fairest for individuals: more people get a chance. From the business perspective, the reference standard the human expert is best in terms of performance and modest fairness. The other configurations tend towards higher performance but are low on group fairness. Corrected features configuration has the strongest group fairness, but is severely problematic under performance and still excludes the unprotected groups.

Summary of results

Empirical fairness testing shows that the AI system can influence individuals' access to work. Even with cases where the AI system does not use sensitive attributes (age and gender) explicitly, a bias by proxy variables (test score) still influences the fairness outcomes. On top of this, the results show that there is no one-size-fits-all metric and AI system configuration to satisfy all possible constraints. However, certain configurations satisfy fairness according to more metrics than other. Specifically:

- A1 and human-expert decisions appear to have better overall fairness metrics, with minor bias, and acceptable and best performances, respectively.

- Other configurations, with good performance, exhibit systematic discrimination risks: violating Statistical and Predictive parity, and Odds and FPR parity for gender.
- Corrected features configuration improves various metrics, but is discriminatory towards the unprotected groups (positive discrimination) and sacrifices accuracy.

The analysis, therefore, confirms the necessity of bias monitoring and human oversight before lawful deployment under Articles 9 and 27 AI Act. This effectively excludes the other configurations (even if the algorithm can be corrected for sensitive features, the performance is unacceptable).

Appendix 2 - Trustworthy AI requirements assessment

The AI system can affect the fundamental rights to equal treatment, privacy, and access to work of applicants. Because the system supports employment decisions, its outcomes have a direct and personal effect on applicants. Other interested parties are not foreseen to be affected. Below are defined the trustworthy AI requirements according to the Ethics Guidelines For Trustworthy AI (2019) together with the impact assessment relevant for the AI system.

Requirement	Assessment
Human Agency & Oversight	
1. Fundamental rights AI systems can support or harm fundamental rights; where risks exist, a prior fundamental rights impact assessment is required, and mechanisms for external feedback must be in place.	Analysed In section Impact on Fundamental rights. Risks to equality, privacy, and human dignity identified and mitigated. Feedback is possible through applicant appeals and DPO contact.
2. Human agency Users must be able to make informed, autonomous decisions, understand the system, and challenge its outputs. AI must not manipulate behaviour or subject individuals to solely automated decisions with significant effects.	Applicants are informed of AI use; HR staff trained to understand and challenge outputs. No automated decisions: every decision is reviewed by humans. No behavioural manipulation occurs.
3. Human oversight Appropriate oversight must prevent adverse effects; humans must be able to intervene, override, and decide when the system is used. Higher-risk systems require stricter oversight and testing.	Full human-in-the-loop: all recommendations reviewed by HR with override authority. System can be paused or disabled, and fallback to manual review exists. Oversight proportional to risk.
Technical Robustness & Safety	
4. Resilience to attack and security AI systems must be protected against vulnerabilities, including data poisoning, model leakage, adversarial attacks, and misuse. Security processes must prevent corruption, unintended applications, and malicious exploitation.	Model and data are stored in a secured environment with restricted access; no external API exposure. Adversarial risk extremely low due to internal-only use [and the model is based on LOGISTIC REGRESSION]. Regular IT security controls prevent tampering or data corruption.
5. Fallback plan and general safety AI systems must have safeguards that allow safe fallback in case of malfunction. Systems must minimise unintended	Full fallback to manual HR screening if anomalies, [bias drift], or errors are detected. Human override is built-in. AI output is advisory only, preventing harmful

consequences, ensure safe behaviour, and apply safety measures proportionate to their risk.	autonomous decisions. Safety controls proportional to the limited risk profile.
6. Accuracy Systems should make correct predictions, and their accuracy should be evaluated and monitored. When errors are unavoidable, the likelihood of errors should be indicated. High accuracy is important where human lives or rights are affected.	The used model configuration shows stable accuracy [(~0.74)]. [Accuracy monitored quarterly.] HR reviewers see confidence/probability scores, and final decisions remain human-made, reducing impact of occasional misclassification.
Privacy & Data Governance	
8. Privacy and data protection AI systems must guarantee privacy and data protection throughout their lifecycle. Data must not be used to infer or discriminate based on sensitive attributes and must not be used unlawfully or unfairly.	[GDPR DPIA completed.] Only job-relevant data (test score, contact info) processed; sensitive fields (age, gender) excluded from production model. Outputs and logs are pseudonymised and retained only as needed. Indirect correlations with sensitive attributes have been assessed. No profiling or inference of hidden attributes occurs.
9. Quality and integrity of data Data must be accurate, free from errors and socially constructed biases, and protected against corruption. All datasets and processes must be tested and documented at each stage of planning, training, testing and deployment.	The dataset is [regularly] reviewed for completeness and bias (see Appendix 1); in case, errors are corrected before training. Dataset, preprocessing steps, and model versions are documented. [Data integrity controls and validation scripts prevent accidental or malicious corruption.]
10. Access to data Clear protocols must define who may access data and under which conditions. Only qualified personnel with a legitimate need should have access.	Strict access control: only HR and authorised data analysts can access applicant data. Role-based permissions, logging, and confidentiality obligations in place. Internal audits verify adherence to access policies.
Transparency	
11. Traceability Datasets, data-gathering processes, labelling steps, and algorithms must be documented to enable full traceability of the AI system and its decisions. Traceability enables identification of errors and supports auditability and explainability.	All datasets, preprocessing steps, model versions, parameters, and training scripts are documented. [NEURAL NETWORK CASE Model explainability techniques like SHAP and LIME are available for explainability purposes. Decision logs link each recommendation to a model version and input features, supporting audits and error analysis]. [LOGISTIC REGRESSION CASE, Decisions within the logistic regression model are easily traceable given the simple model nature.]
12. Explainability AI processes and decisions must be explainable to humans, with explanations adapted to the stakeholder (e.g., applicant, HR staff, regulator). Trade-offs between accuracy and explainability should be considered.	Explanations available upon applicant request. [NEURAL NETWORK] HR reviewers and analysts can access feature-importance (SHAP) scores. Accuracy - explainability trade-offs documented; [LOGISTIC

	REGRESSION] simpler model chosen partly for clarity and explainability.
13. Communication Users must be informed that they are interacting with an AI system. Systems must be identifiable as AI, not appear human, and provide information on capabilities, limitations, and alternatives for human interaction.	Applicants are informed that AI supports screening in terms and condition upon applying [Users may opt for a full human review]. Internal documentation describes system capabilities and limitations. The system does not mimic human communication (no chatbot-like features).
Diversity, Non-Discrimination & Fairness	
14. Avoidance of unfair bias Training and operational data may contain historic or structural biases that can lead to direct or indirect discrimination. Bias must be identified and mitigated through data quality checks, transparent development processes, and diverse perspectives in system design.	Fairness assessment was performed considering sensitive attributes (age and gender) with findings documented in Appendix 1: overall moderate fairness within acceptable performance and data availability constraints [and no sensitive-attribute leakage]. Quarterly fairness audits, balanced training data, and HR override reduce risk.
15. Accessibility and universal design AI systems should be user-centric and accessible to all, regardless of age, gender, or ability. Universal design principles should be followed to support equitable access, including for persons with disabilities.	Applicant notifications and HR tools designed for accessibility; communication available in clear language; alternative manual processes available for candidates requiring accommodations.
16. Stakeholder participation Stakeholders who may be affected should be consulted throughout the AI system's lifecycle, and feedback should be regularly collected after deployment.	Survey-based Value Based Design (VBD) was performed to elicitate stakeholders' requirements prior to deployment. HR staff, works council, and DPO consulted during deployment. Annual review includes feedback from HR users and management. Applicants can provide feedback or request human review.
Societal & Environmental Well-being	
17. Sustainable and environmentally friendly AI AI systems should minimise environmental impact by limiting energy use during development and deployment, assessing the supply chain, and choosing less harmful options wherever possible.	[LOGISTIC REGREGSSION, the employed AI model does not rely on large datasets ('big data') nor complex system architectures (such as LLMs) and therefore do not compute-intensive in training nor usage.] Model is lightweight and trained quarterly on shared infrastructure. Energy use is minimal. No external supply-chain dependencies beyond standard compute resources.
18. Social impact AI systems may affect social relationships, wellbeing, or perceptions of social agency; such impacts should be monitored and considered.	Limited social impact: system only supports internal hiring decisions. Fairness assessment ensures it does not reduce workforce diversity or disadvantage particular groups. No behavioural or social interaction functions. VBD ensures that

	stakeholders' concerns are explored and considered.
19. Society and democracy Beyond individual impacts, AI's broader societal effects must be assessed, especially regarding institutions, democratic processes, and public trust.	Very low societal or democratic impact: system is used internally by a private employer for candidate screening and has no influence on political processes or public institutions.
Accountability	
20. Auditability AI systems must enable assessment of algorithms, data, and design processes. Independent auditing should be possible, especially where fundamental rights are affected.	All datasets, model versions, preprocessing code, and fairness metrics (see Appendix 1) are documented and accessible for audit. Internal audits performed yearly; system can be independently audited by regulators (AP) if required.
21. Minimisation and reporting of negative impacts Systems must allow negative impacts to be identified, documented, minimised, and reported. Whistle-blower protections and impact assessments help prevent harm.	Bias incidents or anomalies are documented and escalated to the DPO. Quarterly fairness testing in place. Whistleblowing channels exist through HR and DPO. Fundamental rights assessment updated when risks change.
22. Trade-offs Conflicts between requirements (e.g., accuracy vs. fairness) must be identified, evaluated, justified, documented, and revisited. Unacceptable trade-offs mean the system should not be deployed.	Accuracy - fairness trade-off documented in Appendix 1: test show a trade-off where discrimination avoidance leads to slightly lower accuracy. All trade-offs reviewed annually and documented in the AI governance register.
23. Redress Accessible mechanisms must exist for individuals to obtain redress when unjust harm occurs, with particular attention to vulnerable groups.	Applicants can request human review, receive explanations, or file complaints with the DPO or AP. HR reviews disputed decisions promptly. Special attention given to applicants who may be disproportionately affected.