2025

AI Toolkit

WORLD EMPLOYMENT CONFEDERATION

The Voice of Labour Market Enablers

WORLD EMPLOYMENT CONFEDERATION
The Voice of Labour Market Enablers

AI Toolkit

CONTENTS

For any questions related to this toolkit, please contact Beatrice Miano, Public Affairs Advisor, World Employment Confederation.

WORLD EMPLOYMENT CONFEDERATION
The Voice of Labour Market Enablers

Introduction

Artificial Intelligence (AI) has the potential to transform the HR industry by automating tasks, improving efficiency, and providing valuable insights. However, it is essential to ensure that AI is used responsibly to mitigate the risks of bias, discrimination, and lack of transparency.

The World Employment Confederation (WEC) recognizes the importance of responsible AI use and therefore, published in March 2023 a Code of Ethical Principles in the Use of Artificial Intelligence. This toolkit has been created in this context: to provide guidance to HR professionals on how to implement and use AI ethically and in compliance with relevant regulations such as the EU AI Act, US AI legal requirements and the International AI Treaty.

This toolkit is a collaborative endeavour between the WEC Digitalization Taskforce and the WEC Data Protection Taskforce. It leverages the expertise of WEC members to deliver practical guidance and best practices and aims to make ethical principles more tangible and applicable by HR professionals.

This work originated from the adoption of the EU AI Act, considering its comprehensiveness as a regulation. The utilisation of AI in the context of recruitment and employment falls under the AI Act's High-Risk category, which necessitates the implementation of risk mitigation measures, processes, and procedures with implications for providers and deployers of AI systems. The toolkit, however, also includes key concepts coming from US federal and state laws.

This toolkit provides specific guidance, actions and templates while focusing upon the risks, requirements and controls related to:

- inclusivity, bias, and discrimination
- transparency and human oversight of AI Systems (i.e. recommendations, decisions)

We hope that this toolkit is a valuable resource to organisations looking to build or procure AI systems in the recruitment and HR domains. We will monitor the legislative developments and update accordingly over the coming years. We also invite readers to share any future improvements or suggestions with the World Employment Confederation so that we can continue to update and improve this toolkit.

WORLD EMPLOYMENT CONFEDERATION
The Voice of Labour Market Enablers

The EU AI Act: Overview & Glossary of Key Concepts

As mentioned, this toolkit is intended to provide guidance for different jurisdictions towards the topics mentioned above. Since the AI Act has extraterritorial effects, we want to provide a high-level context into the EU AI Act.

The diagram below depicts the three main aspects of the EU AI Act that are also relevant to the use of AI in geographies throughout the globe. Understand that for all geographies outside of the EU, the information below was only a guide on how to utilize high-risk AI systems:

01

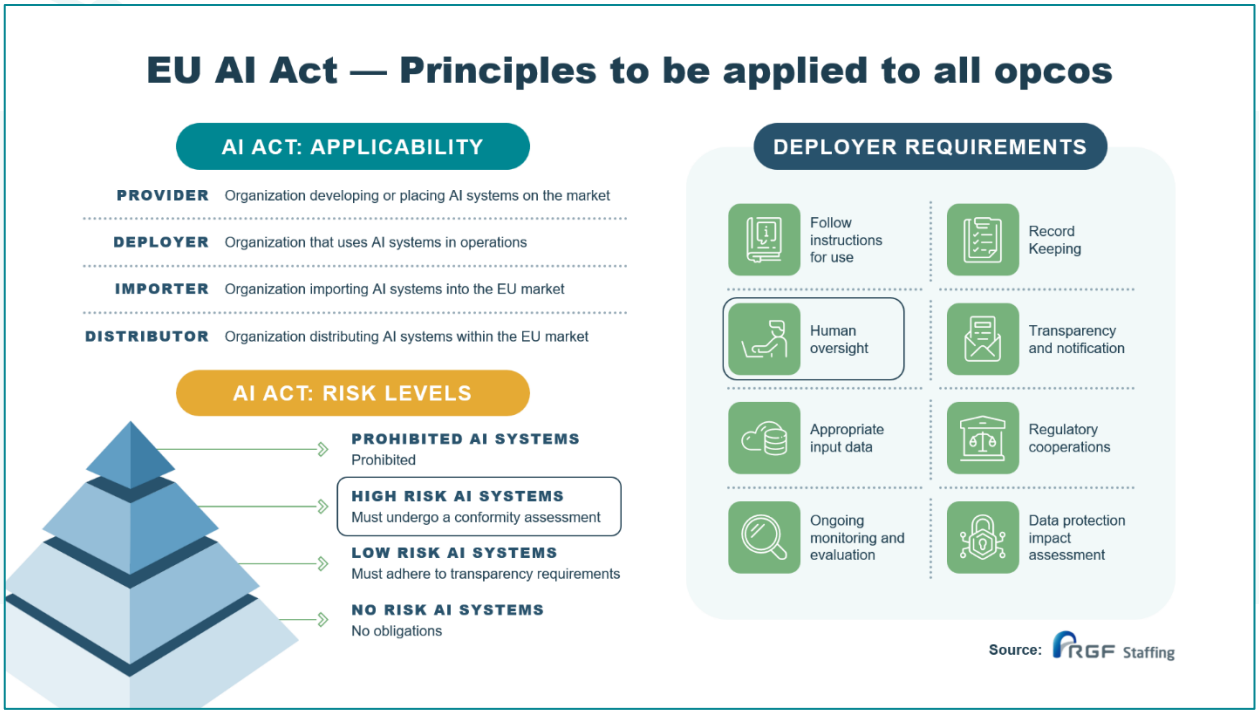
AI Act Applicability: There are four types of delivery and use of AI and are defined in the glossary below. The areas are Providers, Deployers, Importers, and Distributors.

02

Risk Levels of AI: There are four areas of AI Risk, the two we are most concerned about are Prohibited AI Systems and High-Risk AI Systems. Both risk levels are defined in the glossary below.

03

Deployer Requirements: Most agencies within the staffing industry will be Deployers of AI systems. The chart depicts the main requirements that "**deployers"** need to follow to be compliant with the EU AI Act. One key element is that when deployers of AI use High-Risk AI tooling, human-oversight is required during the process.



WORLD EMPLOYMENT CONFEDERATION
The Voice of Labour Market Enablers

GLOSSARY

AI model
A mathematical representation of a real-world phenomenon or process, which is trained on data and can be used to make predictions or decisions. AI models are often trained using machine learning techniques.

AI system
A machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

Deployer
A natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity.

Distributors
Organizations that distribute AI systems within the EU Market.

General-purpose AI model
An AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are released on the market.

General-purpose AI system
An AI system which is based on a general-purpose AI model, which has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems.

High-Risk AI Systems
These are AI systems that, due to their intended purpose, have a significant potential to cause harm to health, safety, or fundamental rights. The Act identifies two main categories of high-risk AI systems:

01 **AI systems as safety components or products:** These are AI systems that are used as part of a product or are themselves products, and are covered by Union harmonisation legislation listed in Annex I. These systems are classified as high-risk if they are required to undergo a third-party conformity assessment under the relevant legislation.

02 **AI systems used in specific high-risk areas:** These are AI systems that are not covered by the Union harmonisation legislation listed in Annex I but are nevertheless considered to be high-risk due to their use in specific sensitive areas. These areas are defined in Annex III, and include:

Biometrics, in so far as their use is permitted under relevant Union or national law:

- Remote biometric identification systems (except for systems intended solely for biometric verification)
- Biometric categorisation systems based on sensitive or protected attributes (e.g., race, gender, religion, political opinion, sexual orientation)

- Emotion recognition systems

Critical infrastructure:

- AI systems intended to be used as safety components in the management and operation of critical digital infrastructure, road traffic, or in the supply of water, gas, heating, or electricity.

Education and vocational training:

- Systems used to determine access or admission to educational institutions, or to assign students to specific institutions or programs.
- Systems used to evaluate learning outcomes.
- Systems used to assess the appropriate level of education for an individual.
- Systems used for monitoring and detecting prohibited behaviour of students during tests.

Employment, workers management, and access to self-employment:

- Systems used for recruitment and selection of personnel.
- Systems used to make decisions affecting terms of work-related relationships.
- Systems used to allocate tasks based on individual behaviour or personal traits.
- Systems used to monitor or evaluate the performance and behaviour of individuals in work-related relationships.

Access to and enjoyment of essential private services and essential public services and benefits:

- Systems used by public authorities to evaluate the eligibility of individuals for essential public assistance benefits.
- Systems used to evaluate the creditworthiness of natural persons or to establish their credit score, except for AI systems used for the purpose of detecting financial fraud.
- Systems intended for risk assessment and pricing in relation to natural persons in the case of life and health insurance.
- Systems intended to evaluate and classify emergency calls by natural persons, or to be used to dispatch or establish priority in the dispatching of emergency first response services.

Law enforcement:

- Systems used by law enforcement authorities, or by Union institutions, bodies, offices or agencies in support of law enforcement authorities, to assess a natural person's risk of becoming the victim of criminal offences.
- Systems used by law enforcement authorities, or by Union institutions, bodies, offices or agencies in support of law enforcement authorities, as polygraphs or similar tools.
- Systems intended to be used by law enforcement authorities or on their behalf or by Union institutions, bodies, offices or agencies in support of law enforcement authorities to evaluate the reliability of evidence in the course of the investigation or prosecution of criminal offences.
- Systems intended to be used by law enforcement authorities or on their behalf or by Union institutions, bodies, offices or agencies in support of law enforcement authorities for assessing the likelihood of a natural person of offending or re-offending.
- Systems intended to be used by law enforcement authorities, or by Union institutions, bodies, offices or agencies in support of law enforcement authorities for the profiling of natural persons in the course of the detection, investigation or prosecution of criminal offences.

Migration, asylum, and border control management:

- Systems intended to be used by competent public authorities as polygraphs or similar tools.
- Systems intended to be used by competent public authorities to assess a risk (e.g., security risk, a risk of irregular migration, or a health risk).
- Systems intended to be used by competent public authorities to assist in the examination of applications for asylum, visa or residence permits and for associated complaints.
- Systems intended to be used for detecting, recognizing or identifying natural persons, with the exception of the verification of travel documents.

Administration of justice and democratic processes:

- Systems intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts, or to be used in a similar way in alternative dispute resolution.

- Systems intended to be used for influencing the outcome of an election or referendum or the voting behaviour of natural persons. (This does not include AI systems to the output of which natural persons are not directly exposed, such as tools used to organise, optimise or structure political campaigns from an administrative or logistical point of view.)

Importers
Organizations that import AI systems into the EU Market.

Non-personal data
Data other than personal data as defined in Article 4, point (1), of Regulation (EU) 2016/679.

Personal data
Personal data as defined in Article 4, point (1), of Regulation (EU) 2016/679: any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

Prohibited AI Systems
The EU AI Act prohibits the development, deployment, and use of AI systems that:

- **Employ subliminal techniques:** These systems aim to influence a person or group of people without their conscious awareness, using techniques like subliminal messaging or deception.
- **Exploit vulnerabilities:** These systems exploit the vulnerabilities of individuals or specific groups of people due to their age, disability, or social or economic situation.
- **Engage in social scoring:** These systems categorise individuals based on their social behaviour, which can lead to discriminatory outcomes and the exclusion of specific groups.
- **Use profiling for risk assessments:** These systems assess the likelihood of individuals committing a criminal offence based solely on profiling, without considering objective evidence.
- **Create facial recognition databases through untargeted scraping:** These systems scrape facial images from the internet or CCTV footage without a specific purpose.
- **Infer emotions in workplaces or education institutions**: These systems infer the emotions of individuals in workplaces or education institutions without being intended for medical or safety purposes.
- **Categorise individuals based on sensitive attributes:** These systems categorise individuals based on their race, political opinions, trade union membership, religious beliefs, sex life, or sexual orientation.
- **Use real-time remote biometric identification:** These systems are prohibited except in specific cases for law enforcement, such as searching for victims of abduction or trafficking, preventing a terrorist attack, or identifying a suspect of a serious criminal offence.

Provider
A natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge.

Risk
The combination of the probability of an occurrence of harm and the severity of that harm.

Special categories of personal data
The categories of personal data referred to in Article 9(1) of Regulation (EU) 2016/679, Article 10 of Directive (EU) 2016/680 and Article 10(1) of Regulation (EU) 2018/1725: personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union

membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation

Systemic risk

A risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain.

WORLD EMPLOYMENT CONFEDERATION
The Voice of Labour Market Enablers

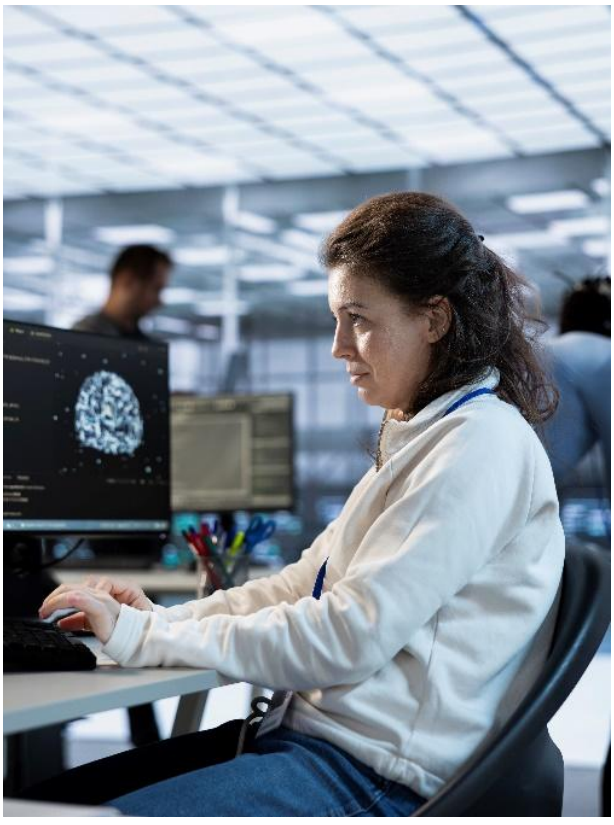
Transparency And Human Oversight

INTRODUCTION TO THE CONCEPTS OF TRANSPARENCY & HUMAN OVERSIGHT IN AI SYSTEMS

In an era when Artificial Intelligence (AI) systems are increasingly woven into our daily lives, the principles of transparency and human oversight are essential.

Transparency allows for comprehension of AI system functionality, encompassing the data utilized, algorithms employed, and decisions made. Transparency is a foundational principle for trustworthy AI, enabling traceability, accountability, and the protection of fundamental rights. It is essential for building public trust and facilitating the responsible development and use of AI technologies across various sectors.

Human oversight, sometimes used interchangeably with concepts like 'human-in-the-loop', involves active human monitoring and intervention in AI systems to guarantee proper functionality and adherence to ethical and legal standards. While this principle is globally relevant, this guide uses the concept of human oversight as defined by the European Union's AI Act



Transparency and human oversight become particularly crucial when AI systems are applied in "high stakes" contexts. Recruitment and employment is undoubtedly a high stakes context. This is because there could be material harm to people's lives if the systems are inaccurate, biased, or not robust. In our industry, different types of AI, both generative and non-generative systems, are increasingly used to automate tasks such as resume screening, candidate selection, interview scheduling, and even conducting interviews. When AI systems are involved in hiring and employment processes, they can:

- Influence career prospects and thereby have a long-term impact on people's economic outcomes
- Affect livelihoods
- Impact workers' rights

Because of these known risks, society expects companies that use these systems to do so in a way that is trustworthy. Being transparent, and training human experts to understand the limitations of these systems and ensuring that such high-stakes decisions are not fully automated, goes a long way in establishing that trust.

This is why the World Employment Confederation together with Charlotte Eijkelkamp, Research Intern at Randstad Global, created a Guide to help a broad range of stakeholders implement transparency and human oversight in recruitment.

HOW TO IMPLEMENT TRANSPARENCY AND HUMAN OVERSIGHT: A GUIDE

OVERVIEW

The Guide aims to provide practical guidance on implementing transparency and human oversight in recruitment and employment AI systems. It is designed for a broad range of stakeholders, including:

Employers and HR professionals who are responsible for implementing and using AI systems in the hiring process.

AI developers and vendors who design and build AI systems for recruitment and employment.

Policymakers and regulators who are responsible for setting standards and guidelines for the ethical use of AI.

Job seekers and advocacy groups who are concerned about the potential impact of AI on employment opportunities and fairness. (Note: this stakeholder group is unlikely to be direct users of the framework, but their interests were kept in mind during the creation of the framework.)

This Guide is **organisation-agnostic and vendor-agnostic**. You should be able to use it regardless of what type of industry you operate in; it is relevant as long as your organisation uses AI in recruitment or HR processes. Similarly, if you are procuring these AI systems from a vendor, and want to ensure that the vendors' systems are transparent and can be overseen by humans in a meaningful way, use these guidelines to help you ask vendors the right questions during the procurement process.

The Guide can be used to:

- Assess the transparency and human oversight of AI systems
- Implement best practices for responsible AI use

The Guide is available on the website of the World Employment Confederation in Excel format. It was developed, based on work done by Charlotte Eijkelkamp as part of her master thesis at Randstad Global for Leiden University.

WORLD EMPLOYMENT CONFEDERATION
The Voice of Labour Market Enablers

HOW TO USE THE GUIDE

The Guide is divided into five sections, namely:

01 Transparency – *Generic*
02 Transparency – *Application Specific*
03 Human Oversight – *Generic*
04 Human Oversight – *Application Specific*
05 Processes and Applications

Generic: how to achieve a basic level of transparency or human oversight, regardless of the type of AI system and context in which AI is applied

Application-specific: how to achieve transparency or human oversight in specific contexts, i.e. when AI systems are used in different parts of the recruitment and employment process. The framework distinguishes between the following 'applications' of AI:

- Sourcing
- Screening
- Interviewing
- Selection
- Onboarding
- Training & skills development
- Performance management
- Advancement & career paths
- Retention
- Salary evaluation & employee benefits

HOW TO FILTER THE SECTIONS

The guide on both Transparency and Human Oversight can be filtered according to different framework users in the ''Relevant User'' column:

- **Deployer:** according to the EU AI Act definition
- **Provider:** according to the EU AI Act definition
- **Developer:** a subgroup of the Provider, who has hands-on technical involvement in the creation of the system.

How to filter Transparency guidelines

The granularity of information that needs to be made transparent, and how that should be visualised or communicated, is very different depending on who the audience or receiver of that information is. You cannot be transparent about the use of AI without knowing whom you need to be transparent towards. This is why the ''Audience'' column can be filtered among:

- Candidates
- Employees
- Users
- Auditors

How to filter Human Oversight guidelines?

For this specific section, you may need to identify different actors who needs to do the human oversight. This is why the ''Actor'' column can be filtered between:

- **System user:** Active user of the AI system, who makes decisions based on, or with the input of, the AI system.
- **Supervisor:** User of the system that is not only able to oversee individual decisions but is able to oversee bigger trends. In practice, this actor can be, depending on the implementation of the system, the same as the 'system user'.

Use these different dimensions to filter down the guidance relevant to your organisation's situation, to help you achieve transparency and meaningful human oversight in your specific context

WORLD EMPLOYMENT CONFEDERATION
The Voice of Labour Market Enablers

Inclusivity And Bias

INTRODUCTION

The principle of non discrimination constitutes the cornerstone of all just societies. For decades, legislative texts have highlighted the importance of non discriminatory practices, expanding on various grounds including race, sex, religion, age, and sexual orientation, recognizing the inherent dignity and equality of all individuals. This principle is deeply embedded in every aspect of HR, from recruitment and hiring to performance evaluation and promotion. However, the progress made in creating inclusive workplaces faces new challenges in the digital age.

The increasing use of AI in the HR business, mainly for purposes of recruiting, managing talent and conducting performance analyses, poses a new challenge to the established principle. While potentially offering efficiency and relative objectivity in their outcomes, AI systems can perpetuate and even amplify existing biases if not carefully designed and implemented, and therefore reinforce the discrimination against targeted social groups.

In order to understand why these systems can become non inclusive and biased, imagine them in a form of a mirror; they reflect the data sets they are trained on. If the data sets in use have been trained based

on structural, individual and institutional biases, the algorithmic biases that will occur will pledge for non-inclusive results. The impact of such results can be detrimental in the recruitment and staffing processes as it can create decisions that perpetuate and amplify existing inequalities, leading to discriminatory outcomes that unfairly disadvantage certain groups. Therefore a set of good practices needs to be established within every organization, in order to enhance the HR processes by leveraging AI, without compromising their ability to be inclusive.

THE PERILS OF AI BIAS

AI can perpetuate existing biases since AI systems are trained on large scale datasets. As a result, these systems may have the same biases as those sources and the society that produced them. The use of discriminatory AI systems and their resulting harms have been well documented in the HR industry, particularly in the case of Amazon and its recruiting tool.

CASE STUDY

In 2018, Reuters reported that Amazon had abandoned an AI-powered recruiting tool it had been developing for multiple years.[1] The tool was designed to automate the initial screening process for potential job candidates by analyzing their resumes. However, during the training phase, the AI system exhibited significant gender bias against female candidates.

The AI tool was trained on resumes submitted to Amazon over the span of a decade, primarily from male candidates due to the male-dominated nature of the tech industry. As a result, the system learned to penalize CVs that listed educational backgrounds or extracurricular activities associated with women. Consequently, the AI downgraded the ranking of candidates based on their gender, effectively discriminating against qualified female applicants.

This case study highlights the significant legal ramifications of AI bias in employment, underscoring its potential to foster discriminatory practices with adverse legal and societal consequences. Notably, the EU's groundbreaking AI Act mandates substantial penalties for such violations, potentially reaching €30 million or 6% of a company's global annual turnover, whichever is greater.

LEGAL FRAMEWORKS
for addressing AI bias in employment

Growing awareness of the potential for AI bias in employment has spurred the development of legal frameworks worldwide to protect individuals from algorithmic discrimination. These frameworks generally

EUROPEAN UNION

The EU has been proactive in establishing comprehensive protection against discrimination. Directives 2000/43/EC and 2000/78/EC prohibit discrimination on grounds of race, ethnic origin, religion, belief, disability, age,

UNITED STATES

U.S. federal law prohibits employment discrimination based on race, color, religion, sex (including sexual orientation and gender identity), national origin, disability, and veteran status. The Office of Federal

ASIA-PACIFIC

The Asia-Pacific region, despite its diverse legal landscape, is also witnessing a rise in anti-discrimination measures related to AI in employment. For example, in South Korea, the Personal Information Protection Act

[1] https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

prohibit discriminatory practices based on protected characteristics and establish mechanisms for redress.

and sexual orientation. Directive 2006/54/EC ensures equal treatment for men and women in employment. These directives notably shift the burden of proof to the respondent in cases with prima facie evidence of discrimination.

Furthermore, the landmark AI Act designates AI systems used in employment as "high-risk." This classification mandates strict obligations for providers and deployers regarding risk management, data governance, and transparency, with significant penalties for non-compliance.

Contract Compliance Programs (OFCCP) handles complaints from applicants and employees alleging discrimination.

Complementing federal legislation, a wave of state-level AI regulations is emerging. Colorado, for example, has enacted comprehensive legislation requiring developers and deployers of high-risk AI systems to exercise reasonable care to prevent algorithmic discrimination. This trend signifies a growing focus on AI bias at the state level.

was amended in 2020 to prohibit discrimination by automated decision-making systems, including those used in recruitment. Similarly, India's draft Digital India Bill is expected to include provisions addressing algorithmic bias and discrimination in various sectors, including employment.

These legal developments across various jurisdictions demonstrate a global movement towards mitigating the risks of AI bias in employment. By establishing clear prohibitions against discrimination and imposing obligations on AI system developers and deployers, these frameworks aim to ensure fairness and equality in the workplace.

THE PROMISE OF AI IN EMPLOYMENT

Since AI is machine-based, it is often touted as a possible solution to the human bias problem. The presumption is that bias can be consciously excluded when training an AI system, for example, by not including attributes such as age, gender, nationality etc. (‘discriminatory attributes’) in the datasets used for it, to eliminate any influence those attributes may have on the AI’s output. However, the Amazon example demonstrated that AI is sufficiently versatile, even in the absence of discriminatory attributes, to still produce discriminatory outcomes by indirectly inferring from other information in the dataset. This leaves recruiters with little choice but to review the AI outcomes to make sure there are no mistakes or unfair decisions that can cause harm to individuals.

Despite these challenges, AI still offers significant potential for fostering fairer hiring practices. It can serve as a diagnostic tool, revealing existing societal biases through rigorous pre-deployment testing and monitoring. This iterative process enhances awareness of the need for bias detection, an approach that is not readily applicable to individual human recruiters. By acknowledging and addressing the limitations of AI, we can leverage its strengths to create a more equitable recruitment process.

WORLD EMPLOYMENT CONFEDERATION
The Voice of Labour Market Enablers

STRATEGIES FOR MINIMIZING BIAS OUTCOME IN RECRUITMENT PRACTICES

Notwithstanding the challenges in building an AI that is risk free, AI can significantly help recruiters manage the large number of job opportunities and even larger number of applicants while at the same time minimizing AI-risks.

Although the common narrative is that the datasets to train AI must be free of bias to ensure a bias-free AI outcome, there are no details on how this can be achieved. In the absence of technical guidance, we advocate using AI's known features to achieve fair outcomes and biases that are minimized as much as possible.

Some controls to consider are:

Bias Risk	Description	Mitigation Measures
Data Bias	Training data reflects existing societal biases, leading to discriminatory outcomes.	**Diverse and representative datasets:** Ensure training data includes a wide range of demographics and backgrounds.
		Data preprocessing: Clean and transform data to remove or mitigate biases.
		Synthetic data generation: Create artificial data that reflects desired diversity and fairness.
		Data Augmentation: Increase the representation of underrepresented groups by creating synthetic data points or carefully duplicating and modifying existing ones.
		Reweighting: Adjust the weight given to different data points during training to counter imbalances in the dataset.
		Statistical Parity Measures: Use statistical techniques to ensure the model's predictions are equally distributed across different demographic groups.
Algorithmic Bias	AI models amplify or perpetuate biases present in the data.	**Fairness-aware algorithms:** Employ algorithms designed to minimize bias and promote fairness.
		Regular audits: Conduct ongoing audits to identify and address potential biases in the model's outputs.
		Explainable AI (XAI): Use XAI techniques to understand how the model makes decisions and identify potential sources of bias.

Algorithmic Bias	AI models amplify or perpetuate biases present in the data.	**Adversarial Training:** Train the model with adversarial examples to identify vulnerabilities and improve fairness.
		Regularization Techniques: Add constraints to the model's training process to discourage it from learning patterns that lead to biased outcomes.
		Counterfactual Analysis: Assess how the model's predictions would change if certain input features were different, helping to identify and mitigate biases.
		Ensemble Methods: Combine multiple AI models trained on different subsets of the data or with different algorithms to reduce the impact of individual model biases.
		Bias Statements: For high-risk AI, best practice is to perform bias testing on the algorithm (or request from provider) and publish results, to confirm non-bias.
Human Bias	Human biases influence the design, development, and deployment of AI systems.	**Diverse development teams:** Include individuals from different backgrounds and perspectives in the AI development process.
		Bias awareness training: Educate developers and users about potential biases and how to mitigate them.
		Human oversight: Maintain human review and intervention in critical decision-making processes.
		Structured Interviews: Use standardized interview questions and scoring systems to reduce the influence of interviewer bias.
		Unconscious Bias Training: Educate hiring managers and recruiters about unconscious bias and its impact on decision-making.
		Diversity and Inclusion Goals: Set specific goals for increasing diversity in hiring and track progress towards those goals.
Confirmation Bias	AI systems reinforce existing beliefs and preferences, leading to a lack of diversity.	**Blind recruitment techniques:** Remove identifying information from resumes to reduce unconscious bias.
		Diverse candidate slates: Ensure a diverse pool of candidates is considered for each position.
		Objective evaluation criteria: Use clear and measurable criteria for evaluating candidates.

Confirmation Bias	AI systems reinforce existing beliefs and preferences, leading to a lack of diversity.	**Skills-based Assessments:** Use assessments that focus on measuring candidates' relevant skills and abilities, rather than relying on resumes or subjective evaluations.
Similarity Bias	AI systems favor candidates who resemble those already successful in the company.	**Anonymized profiles:** Remove information that could reveal a candidate's background or identity.
		Skills-based assessments: Focus on evaluating candidates' skills and abilities rather than their background.
		Diversity goals: Set targets for hiring individuals from underrepresented groups.
		Blind Hiring Tools: Utilize software that removes identifying information from resumes and applications, allowing recruiters to focus on qualifications.

This table provides a comprehensive overview of the key bias risks and mitigation measures in AI-driven recruitment processes. It is essential to tailor these measures to specific recruitment needs as such needs differ from company to company and even between recruitment teams.

Additional Considerations:

Transparency

Be transparent about how AI is being used in the recruitment process and provide candidates with information about how their data is being used.

Accountability

Establish clear lines of responsibility for ensuring fairness and mitigating bias in AI-powered recruitment systems.

Continuous Monitoring

Regularly evaluate the performance and effectiveness of AI systems and make adjustments as needed to ensure fairness and equity in the hiring practices and mitigate bias.

By implementing these measures, organizations can help to ensure that AI is used responsibly and ethically in the recruitment process, promoting fairness and equal opportunities for all candidates.

Even with the emergence of Generative AI ("GenAI"), a special category of AI capable of generating contents, the strategies outlined above to manage bias still apply. For instance, when using a GenAI to assess the suitability of candidates, the system should strictly evaluate the resume against the job description (and not current employees' resumes) because the job description represents the public communication of what the ideal candidate should be. It is only fair that GenAI evaluates the received resumes against the criteria communicated publicly. Another example proposes using GenAI to scrutinize job descriptions after training the GenAI to look out for bias attributes or sentences deemed to be biased by referring to relevant laws/regulations/guidelines as training datasets. Notwithstanding methods such as prompt steering and using correct datasets, it is essential that human recruiters still review the GenAI's output to pre-empt 'hallucinations' that have gotten human users into trouble.

WORLD EMPLOYMENT CONFEDERATION
The Voice of Labour Market Enablers

THE ROLE OF HR PROFESSIONALS IN REVIEWING AND ADDRESSING AI OUTCOMES

By account of the above, HR professionals are expected to play a bigger, not lesser, role using AI to complement the human recruiter (human-centric approach). Although AI is touted to possess human-like intelligence, it is far from fully replicating how a human thinks, feels and makes decisions. Contrary to common misconceptions, humans do not make decisions solely based on rules. Besides rules, humans base their decisions on common-sense reasoning, intuitive understanding, and contextual awareness which are absent in AI. For instance, a human recruiter may attempt to understand why an applicant has a gap in their resume, e.g., the applicant may have been made redundant during the COVID pandemic, but the AI may not consider this aspect and reject the resume outright because it replicates the datasets it is subject to. AI, despite its perceived flexibility in many domains such as recruitment, suffers from a lack of understanding of the real world unlike humans. It operates based on patterns learned from datasets without comprehending the underlying concepts. At this stage of AI technology, it is not possible to impart contextual considerations to AI to make it behave more human-like.

For these reasons, it is still important to involve HR professionals to review AI decisions, primarily to check for alignment to human values, eliminating bias and preventing unfair decisions. Maintaining a record of human interventions and annotating AI outcomes is also an important step to creating reinforced training datasets to further mature AI to be more human-like in decision-making and this can be only achieved with the participation of knowledgeable human HR professionals.

NECESSARY TRAINING AND EDUCATION FOR HR PROFESSIONALS

HR professionals and those involved in recruiting have a role to play in ensuring AI is used fairly and ethically in recruitment, especially keenly aware of bias in datasets and AI outcomes. For this to happen, everyone with a role leveraging AI should be trained in recognizing bias (both conscious and unconscious) so that it becomes easy to detect and tackle. Organisations should invest in establishing a data governance programme to enable standards and practices to be implemented that help ensure data reliability and consistency, to know the source of data used to train the AI and do the necessary pre-processing to make sure the datasets are, and remain, relevant and free of irrelevant data that could alter the intended output of the AI, among others. Investing in the right tools and technologies to support data governance is also essential. Regular bias training and workshops for employees are crucial to empower their teams to recognise and deal with their own biases, as humans are the only reliable entity who can proactively identify and intervene to prevent biased data from entering into the AI ecosystem. This people-focused approach, combined with well-defined processes and appropriate technology, creates a robust framework for ethical and responsible AI development.

Finally, by facilitating open discussions on bias, employers can help create a conducive environment where employees feel comfortable sharing their experiences and insights. This will ultimately lead to more equitable hiring processes with minimised bias[2], allowing HR organizations to leverage AI's full potential without compromising their commitment to a positive and inclusive candidate experience.

[2] S. Wachter, "Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond," *Yale Journal of Law & Technology,* p. 688, 2024.