Department for Science, Innovation & Technology

RESPONSIBLE AI IN RECRUITMENT

Guidance

March 2024

Department for Science, Innovation & Technology

Contents

This guidance has been developed with feedback and contributions from the ICO and EHRC, alongside:

Department for Science, Innovation & Technology

01

Executive summary

Adopting Artificial Intelligence (AI)-enabled tools in HR and recruitment processes offers the automation and simplification of existing processes, promising greater efficiency, scalability, and consistency. However, these technologies also pose novel risks, including perpetuating existing biases, digital exclusion, and discriminatory job advertising and targeting.

Tools for trustworthy AI, including AI assurance mechanisms and global technical standards, can play a vital role in managing these risks and building trust. For example, the Department for Science, Innovation and Technology's (DSIT) Public attitudes to data and AI tracker survey (Wave 1) demonstrates that strong, well-communicated governance and assurance mechanisms can increase public willingness to share data for a variety of uses.

This guidance identifies potential ethical risks of using AI in recruitment and hiring processes. It further outlines how AI assurance mechanisms can provide organisations with the tools, processes and metrics to evaluate the performance of AI systems, manage risks, and ensure compliance with statutory and regulatory requirements. It is intended for organisations seeking to procure and deploy AI systems in their recruitment processes. The guidance is written for a non-technical audience and assumes a minimal understanding of AI and data-driven technologies, and is appropriate for organisations with or without a comprehensive AI strategy. Through this guidance, readers will gain an understanding of:

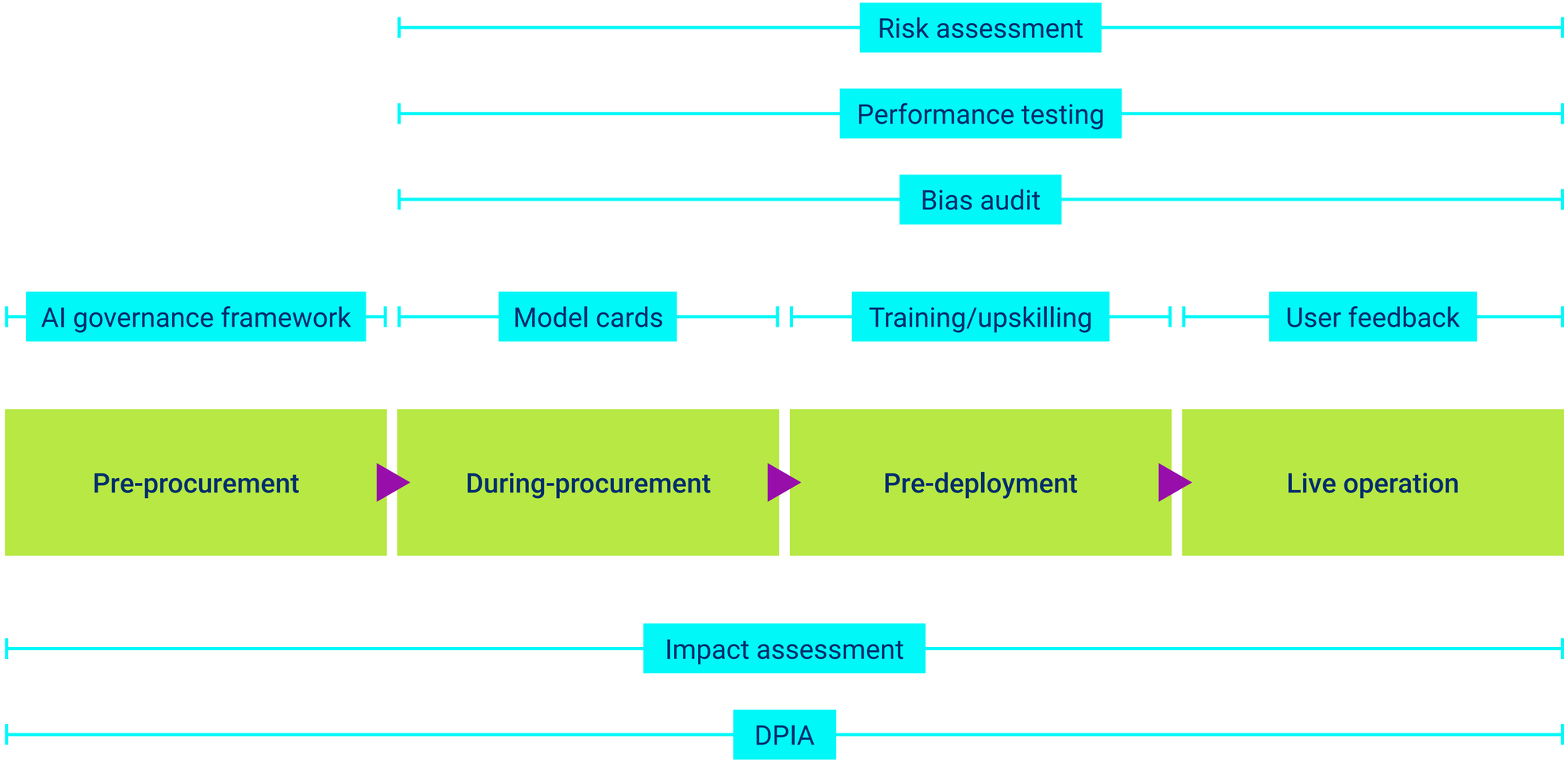
- **key considerations:** Core areas organisations should consider when procuring and deploying AI responsibly in recruitment
- **assurance mechanisms:** Actions which aim to directly address these considerations and support alignment with the UK government's AI regulatory principles
- **justified trust in suppliers:** Examples of what constitutes acceptable evidence of a supplier's claims around their system
- **key risks:** Examples of the key risks arising from use cases of AI in recruitment

In the HR and recruitment sector, organisations are using a range of AI-enabled technologies across every stage of the recruitment process, including:

At all stages there is a risk of unfair bias or discrimination against applicants. Additionally, inherent to these technologies is a risk of digital exclusion for applicants who may not be proficient in, or have access to, technology due to age, disability, socio-economic status or religion. An overview of some of the specific risks posed by these systems is provided in annex A.

This guidance outlines a range of considerations that should be considered by all organisations seeking to procure and deploy AI in recruitment. Alongside this we outline options for mechanisms that may be used to address concerns, actions or risks identified as a result of these considerations.

While the assurance mechanisms are segmented by the stage of the procurement/deployment lifecycle, many should be repeated across stages as outlined in the diagram below:



Department for Science, Innovation & Technology

02

Introduction

Introduction

In December 2021, DSIT's Responsible Technology Adoption Unit (RTA) (formerly, the Centre for Data Ethics and Innovation) co-published the Data-driven tools in recruitment guidance with the Recruitment and Employment Confederation (REC), an industry body for UK-based recruiters.

Since publication, there have been two important developments relevant to AI in recruitment:

(1) advancement of the UK's AI regulatory ecosystem, with publication of the UK government's outcomes-based AI governance framework: A pro-innovation approach to AI regulation and the government response to the AI white paper consultation

(2) government engagement with HR and recruitment organisations to understand current attitudes towards AI governance and tools for trustworthy AI. DSIT published findings from this engagement in its Industry Temperature Check, which identified a need for clearer guidance on how HR and recruitment organisations can, and should, use AI assurance mechanisms to support the responsible procurement and use of AI systems

In response to these advancements, this guidance outlines how AI assurance mechanisms can support the responsible procurement of AI systems in HR and recruitment, to support the implementation of UK's wider approach to AI governance.

As AI becomes increasingly prevalent in the HR and recruitment sector, it is essential that the procurement, deployment, and use of AI adheres to the UK Government's AI regulatory principles, outlined in 'A pro-innovation approach to AI regulation'. These principles are:

Safety, security and robustness

Appropriate transparency and explainability

Fairness

Accountability and governance

Contestability and redress

AI assurance will play a critical role in the implementation and operationalisation of these principles. The principles identify specific goals – the 'what' that AI systems should achieve, regardless of the sector in which they are deployed. AI assurance techniques and standards (commonly referred to as "tools for trustworthy AI") can support industry and regulators to understand 'how' to operationalise these principles in practice, by providing agreed upon processes, metrics, and frameworks to support and achieve these goals.

For general background on AI assurance, please consult DSIT's Introduction to AI Assurance. For legal compliance, please consult resources from the appropriate regulators including the Information Commissioner's Office (ICO) and the Equality and Human Rights Commission (EHRC). This guidance is **not** to be construed as providing any legal assurance or legal advice. Should you require legal assurance or legal advice, please contact an independent legal adviser.

Department for Science, Innovation & Technology

03

Assurance mechanisms for procurement

03 Procurement

Organisations seeking to procure an AI system from third-party suppliers should consider the potential risks posed by these systems and identify appropriate assurance mechanisms both before and during the procurement process. This will ensure that the system is trustworthy, with a view to being aligned with the UK's AI regulatory principles, as well as their own organisational goals.

3.1 Before procurement

Q What do I need to think about before procuring an AI system?

Before your organisation goes out to tender, you should identify what kind of AI system you are looking to procure and why, and consider how this system will sit within existing organisational processes and structure.

Considerations

Purpose

Prior to procuring an AI system, your organisation should develop a clear vision that **outlines the desired purpose of the AI system**.

Defining the purpose of system will help to determine how the system will offer value for money and benefits for your organisation. For example, implementing a chatbot for potential job applicants may improve efficiency and free up employee time to perform other tasks.

Understanding the intended purpose of the system will also help to ensure that it is used to perform the task for which it was designed and avoid potential misuse.

Suggested questions

What problem is my organisation trying to solve?

How can using an AI system help to address this problem?

What is the task that I want the AI system to perform, i.e., what is the purpose of the system?

Is AI appropriate for the problem I am looking to solve?

How will my organisation effectively communicate the use of AI to potential applicants?

Relevant assurance mechanisms

AI governance framework

Impact assessment

Functionality

It is important to clarify the **desired outputs of the AI system** your organisation is seeking to procure.

Clarifying the desired functionality of the system will help your organisation to develop a set of **functionality requirements for suppliers**. Functionality considerations may be informed by early consultation with suppliers to understand what is feasible/ reasonable to expect from an AI system..

Suggested questions

What systems/processes do I expect the AI system to undertake?

What outcomes do I want the AI systems to produce (i.e. a report? An interactive chat? A list of suitable candidates?)

Do AI systems on the market have these capabilities? Can they produce these outputs?

Relevant assurance mechanisms

AI governance framework

Impact assessment

Resources and governance

Once your organisation has defined the desired purpose and functionality of the AI system, it is important to consider **how the system will integrate into existing organisational processes and interact with employees**.

Your organisation should consult with employees who will use the system to understand what education, training or skills they may require to use it effectively. Employees must be able to meaningfully engage with the outputs of a system to feel confident in acting on the AI-enabled prediction, decision or recommendation.

Suggested questions

How will my employees interact with the system?

How will my organisation maintain effective human oversight of the system and its outputs?

What additional educational resources, training or skills may be needed to support my employees?

How will my organisation address feedback from users of the system?

Relevant assurance mechanisms

AI governance framework

Impact assessment

Applicant accessibility requirements

In any recruitment process, applicants with disabilities, conditions, or impairments **may require reasonable adjustments** to the recruitment and hiring process to ensure that they are not disadvantaged. **This is a legal obligation** pursuant to section 20 of the Equality Act 2010. If/when AI systems are introduced into the recruitment process, these may bring about novel risks of disadvantage.

It is essential to ensure that any system your organisation procures is compliant with the Equality Act 2010 which governs anti-discrimination law in the UK. This is particularly important if the system you are procuring will process data containing protected characteristics (e.g. race, ethnicity, religion, etc.) or proxy indicators of those characteristics.

If you have questions about compliance with the Equality Act 2010, access guidance from the EHRC here. Private organisations seeking to provide services to, and on behalf of, the public sector must also adhere to the Public Sector Equality Duty which carries further obligations.

Suggested questions

Does introducing technology into the recruitment process **create new barriers** to applicants with protected characteristics?

Does introducing technology into the recruitment process **amplify existing risks at scale**, for example, entrenching and/or perpetuating human biases?

Does introducing technology into the recruitment process **create novel risks at scale**?

Relevant assurance mechanisms

AI governance framework

Impact assessment

Data protection

In the UK, the ICO regulates data protection according to the UK General Data Protection Regulation (UK GDPR) and Data Protection Act 2018 (DPA). Before deploying an AI system, organisations should consult the ICO's AI guidance portfolio and/or seek independent legal advice to determine if your use of AI is legally compliant with UK GDPR.

If your organisation uses AI in recruitment, you must give consideration as to whether this type of automated process falls within scope of Article 22 of the UK GDPR, and whether **you must complete a data protection impact assessment (DPIA)** to ensure your AI system is legally compliant and demonstrate that you have mitigated any high risks. The ICO's template for a DPIA can be found here. If you have not been able to mitigate any high risks, you **must** consult the ICO. Examples of high-risk processing include innovative technology (including AI), large-scale profiling, biometric data and data matching – all of which are present in the use cases identified in this guidance.

Relevant assurance mechanisms

Data Protection Impact Assessment

Assurance mechanisms

To address these considerations, your organisation should consider implementing the assurance mechanisms outlined below:

Impact assessment

A process to anticipate the wider effects of a system/product on environmental, equality, human rights, data protection, or other outcomes. Impact assessments can be completed internally or by third party auditors.

Different types of impact assessment:

- Algorithmic Impact Assessment (AIA) – considers the potential impacts of an AI system (both short and long term) including data protection, accessibility, and bias
- Equality Impact Assessments (EIA) – focus specifically on equalities outcomes

Principles

Safety, security and robustness

Accountability and governance

Appropriate transparency and explainability

Fairness

Example

An organisation seeking to procure an automated job description writing tool completes an AIA that covers DPIA requirements prior to procurement.

Part of the AIA requires consultation with potential applicants to identify preliminary risks, alongside appropriate UK regulators including the ICO and EHRC.

Stakeholders identify several risks, including non-inclusive language used in the job description, that may prevent them from applying for roles that they are qualified for.

To mitigate this risk of marginalisation, the organisation ensures all descriptions are reviewed by an employee trained in creating fair job descriptions before adverts go live.

Resources

- **IFOW Algorithmic Impact Assessment**
- **Microsoft Responsible AI Impact Assessment Template**
- **ODI Data Ethics Canvas**

Data protection impact assessment (DPIA)

A process for identifying and minimising data protection risks. Completing a DPIA is required for all development and deployment of AI systems that involve personal data.

Principles

Safety, security and robustness

Appropriate transparency and explainability

Fairness

Accountability and governance

Example

An organisation looking to procure a computer vision system for emotion inferences during job interviews completes a DPIA.

As part of the DPIA, the organisation consults the system operators (interviewers) and the data subjects (applicants).

Data subjects express that they would not expect, or necessarily want, their facial data to be used for the purpose of engagement detection and emotion inferencing.

The organisation takes actions to ensure the use and purpose of the AI system is properly communicated to applicants and allows applicants to opt-out of using the system.

Resources

- **DPIA Template | ICO**
- **Ada Lovelace Institute Participatory Data Stewardship**

AI governance framework

Your organisation should create an AI governance framework to set out how AI will be embedded in and complement your existing business functions. For some organisations this may be driven by a set of guiding principles, goals or objectives.

Your AI governance framework should assign who is accountable for the AI system and create methods for escalation. The framework should also set out commitments to transparency and establish how your organisation will communicate to applicants when/how AI systems will be used in the recruitment process. This includes a risk management framework which sets out how your organisation will address feedback from users.

At this early stage of procurement your AI governance framework is likely to be an evolving document that should be informed by consultation with stakeholders, organisation board members and staff.

Principles

Safety, security and robustness

Appropriate transparency and explainability

Fairness

Accountability and governance

Contestability and redress

Example

An organisation seeking to implement AI, develops an AI governance framework to ensure responsible development.

The organisation analyses existing organisational policies, including frameworks for legal compliance, quality assurance and risk management. These policies are adapted to address unique risks and challenges posed by the introduction of AI systems, alongside consultation with the organisation's board and teams using the system.

This framework identifies a knowledge gap requiring upskilling staff once a system has been procured. Specific training requirements cannot be clearly defined until the AI system has been procured, leading the organisation to update the framework post-procurement.

Resources

- **DSIT Introduction to AI Assurance**
- **CIPD Technology use in recruitment and workforce planning**
- **International standards such as those being developed by ISO/IEC (e.g. 42001), IEEE & ETSI**

3.2 During procurement

Once your organisation has a clear idea of what kind of AI system you want to procure, when going out to tender and speaking with potential suppliers you should consider the following:

Considerations

Accuracy and scientific validity

During the procurement process, suppliers may make claims about the performance, return on investment (ROI), efficiency, fairness and capabilities of the system they are selling.

A supplier should be able to provide evidence of these claims to your team, regardless of your level of technical literacy. We suggest asking your supplier for documentation of impact assessments, risk assessments, model cards and/or a DPIA.

Suggested questions

What type of model has been trained? For example, natural language processing (NLP), computer/machine vision or voice recognition.

What data has been used to train the model? This includes data type, breakdown of protected characteristics of data subjects and where the data was sourced from to inform data protection and equality considerations.

What is the intended purpose and scope of use of the model? How will your organisation ensure the system aligns with your intended goals?

What is the demonstrated accuracy and performance of the model across different groups/protected characteristics with specific reference to understanding the UK context, including UK equality law and demographic makeup?

What are the identified risks and system limitations?

Relevant assurance mechanisms

Bias audit

Performance testing

Model cards

Risk identification and communication

With transparency about the system's performance, accuracy and capabilities, your organisation can work with suppliers to build on impact assessment considerations identified in the pre-purchasing phase.

It is essential to understand the actual functionality of the system being procured, and how it may interact with your organisational processes and employees.

Suppliers may have completed a DPIA and an impact or risk assessment for their model. These assessments should be shared during any procurement exercise to offer transparency around potential risks posed by the AI system, in conjunction with the DPIA completed by your organisation prior to procurement.

Suggested questions

Does the supplier's AI system contain additional functionalities or different output formats to those identified in your organisation's original impact assessment?

Has your supplier provided copies of their own DPIA and impact or risk assessments?

Does your supplier's DPIA and impact or risk assessment identify any additional risks or limitations of the system that you did not identify prior to procurement?

Have you developed mitigation plans for any newly identified risks and communicated these to your team?

Relevant assurance mechanisms

Bias audit

Risk Assessment

Model cards

Assurance mechanisms

To address these considerations, your organisation may consider implementing some or all of the assurance mechanisms outlined below:

Bias audit

A process for assessing the inputs and outputs of algorithmic systems to determine whether there is bias in input data, or in the outcome of a decision or classification made by an AI system.

Bias audits should be repeated at regular intervals after the system is in operation to ensure consistent performance.

Principles

Appropriate transparency and explainability

Fairness

Example

An organisation seeking to procure a screening system that scores applicants based on CVs and cover letters asks the supplier for evidence of a bias audit performed on the system.

The audit evaluates model performance based on sex, ethnicity, age and disability. The results show that applicants with disabilities or who are neurodivergent may be disadvantaged by the system.

The supplier clearly communicates this to the buyer as a system limitation. The buyer chooses to proceed with procurement but determines that applicants who declare a disability will not be scored by the screening system, instead going through a manual review process. To ensure applicants with disabilities declare them, the organisation clearly signposts the use of AI to potential applicants and outlines this alternative path. The supplier also commits to repeated bias audits every six months to ensure consistent performance.

Resources

- **How do we ensure fairness in AI? | ICO**
- **Holistic AI: Audits**
- **BABL AI: Third-party audits**
- **IBM Fairness 360**
- **Fairlearn**

Performance testing

A process for assessing the performance of a system with respect to predetermined quantitative requirements or benchmarks.

Typically, performance testing assesses a model against metrics on accuracy (how often a model is correct) and precision (how often a model is correct against a specific category of a dataset).

Deciding what metric is most beneficial for performance testing will depend on the AI model being assessed. For example, in the context of a transcription tool, an appropriate metric may be the successful identification of words.

Principles

Safety, security and robustness

Appropriate transparency and explainability

Fairness

Example

An organisation procuring an interview transcription tool is told by the supplier that the system is 97% accurate in transcribing words of native English speakers.

The supplier agrees to test the model's accuracy on a sample of audio recordings of a pre-recorded interview featuring both native and non-native speakers, to test model performance against the buyer's human transcription benchmark of 90% accuracy.

The model is shown to achieve 95% accuracy at three times the speed of a human reviewer. While the accuracy on this dataset is slightly lower than the supplier's claims, the buyer proceeds with the procurement as the system exceeds the human benchmark for both speed and accuracy.

Resources

- **DSIT Portfolio of AI Assurance Techniques**
- **TensorFlow Responsible AI Guidance**
- **Portfolio of Artificial Intelligence Guidance | ICO**

Risk assessment

A process for identifying and planning mitigations for a range of potential risks that may arise from the deployment of an AI system.

Risk assessments can be completed as part of, or in parallel with impact assessments and should include a list:

- potential risks
- specific mitigations to reduce the likelihood and impact of identified risks
- identification of acceptable and unacceptable risks

Principles

Safety, security and robustness

Appropriate transparency and explainability

Fairness

Example

An organisation procuring a targeted job advertising tool completes a risk assessment to identify potential risks.

One of the risks identified is that the system may create barriers to applicants over the age of 40 with a lower level of digital engagement. The assessment also recognises the longer-term impact of a subsequently non-diverse applicant pool.

A mitigation plan is created to ensure that potential applicants with these characteristics are not disadvantaged. As a result of the strong risk of discrimination, the organisation opts to advertise the role via traditional means alongside the targeted advertising system.

Resources

- **AI and data protection risk toolkit | ICO**
- **NIST Risk Management Framework**
- **International Standards such as ISO/IEC 23894**
- **Fairlearn**

Model cards

A standardised reporting tool for capturing key facts about AI models, including details on:

- model goals and intended use
- limitations of the model
- training data
- model performance
- identified risks

Model cards should be produced by the developer of an AI system and available to the purchasing organisation.

Principles

Safety, security and robustness

Appropriate transparency and explainability

Accountability and governance

Example

An organisation procuring an interview transcription tool asks their supplier for a model card to assess claims made about the system's performance.

The supplier provides a model card that outlines the model's intended uses, limitations, performance and training data.

One of the limitations identified is that the tool performs worse for applicants who speak Spanish or Portuguese as a first language.

The organisation choses to proceed with the procurement, however, ensures that applicants who are native Spanish or Portuguese speakers have their interviews manually transcribed.

Resources

- **Google Model Cards**
- **Hugging Face Model Card Template**

Department for Science, Innovation & Technology

04

Assurance mechanisms for deployment

4.1 Before deployment

Once your organisation has procured a third-party AI system, prior to the deployment of this system, it is recommended that your organisation pilots the technology with potential users. Pilots should engage a **diverse range of users, including employers as well as affected communities** including jobseekers from different backgrounds and experiences. Inclusive pilots will help to gather information on how the system works in practise, and to ensure that assumptions made prior to procurement reflect the real outputs of procured AI system.

Considerations

Avoiding incorrect usage

Ensuring that employees have sufficient clarity on the purpose, functionality and outputs of an AI system is essential to prevent incorrect usage. During the pilot your organisation should seek to understand how employees are using the tool to identify if and where misuse could occur.

During the pilot, your organisation should assess if there is a risk of employees incorrectly using the system.

Suggested questions

Do employees feel empowered to engage with and act on the outputs of the system?.

Are employees clear on the intended purpose of the AI system – do they know which tasks it is suitable to perform?

Are employees using the system to perform any additional tasks?

Do employees understand the outputs of the system?

Do employees have clarity on why a decision or recommendation was made?

Relevant assurance mechanisms

Training/upskilling

Impact assessment

Assess model performance against equalities outcomes

Ensuring that employees have sufficient training to use an AI system can help to prevent incorrect usage . However, there is still a risk that the system you procure may cause harm or discriminatory outcomes, even when it is used as intended. This is because **AI models can perform differently depending on the environment in which they are deployed**.

Discrepancies in model performance may lead to harms like discrimination or perpetuating biases, particularly if/when the model demonstrates a lower level of performance for individuals with protected characteristics. There are two primary sources of bias in AI systems:

- **Learnt bias:** Data-driven tools “learn” how to make hiring decisions based on training data. If there are historical biases in this data, then the tool is likely to perpetuate these biases at scale.
- **Inaccuracy:** Tools that analyse face, speech, or voice data may be less accurate for groups with certain protected characteristics and introduce discrimination by simply working less well for those groups.

For more information on bias in AI system’s consult the Review into bias in algorithmic decision making.

These issues may not become apparent until running a pilot, where discrepancies between supplier data and your organisation’s data will become more apparent. Consistent monitoring of how the AI system performs in your organisation’s environment is essential for identifying these discrepancies

Suggested questions

Have you assessed the AI system for bias?

Have you set up regular testing to identify whether model performance degrades over time?

Relevant assurance mechanisms

Performance testing

A/B testing

Impact assessment

Plan reasonable adjustments

Employers have a legal obligation to make reasonable adjustments to an interview process for applicants who have disabilities if they would be put at a substantial disadvantage by the way the interview process is carried out, by any physical features of the recruitment process, or if they require extra equipment or support (Equality Act 2010 section 20).

Reasonable adjustments should be considered and planned, before the deployment of the technology, in case the system puts an applicant with a protected characteristic at a disadvantage because of that characteristic (consult the EHRC's examples of reasonable adjustments in practice for examples). If a reasonable adjustment cannot be made, this may require the system's removal from the interview process.

In some cases, it is possible to provide reasonable adjustments that will enable a person with protected characteristics to participate in the recruitment process using the AI tool, without them being disadvantaged. For example, deploying text-to-speech software to enable a candidate with a visual impairment to use a chatbot. However, sometimes the substantial disadvantage a person with protected characteristics experiences can only be avoided if the AI system/ technology is removed from the recruitment process.

In such cases the employer will need to consider what format the recruitment process will take for the person with protected characteristics, so they are able to participate on an equal basis with others.

Suggested questions

Is your organisation aware of the key areas where applicants may face disadvantages due to your use of AI?

Has your organisation planned options for reasonable adjustments that would remove this disadvantage?

If reasonable adjustments to the technology cannot be developed, how will your organisation adapt the interview format to ensure a fair recruitment process?

Has your organisation communicated the limitations of the system that would require reasonable adjustments for applicants with disabilities?

Relevant assurance mechanisms

Ensuring transparency

Impact assessment

Assurance mechanisms

Prior to deployment of an AI system, it is recommended organisations run a pilot where the following assurance mechanisms may be used:

Performance testing

A process for assessing performance of a system with respect to predetermined quantitative requirements or benchmarks.

While evidence of performance testing should have been provided prior to procurement and may have used a sample of your organisations' data for testing, understanding how the model performs in your organisation's real-world environment and protected characteristic makeup is essential before deploying any model at scale.

Unless your organisation has in-house technical expertise, a supplier should set up performance testing for your organisation. Should these tests show results that are below organisational benchmarks, your organisation should work with the supplier to understand and improve model performance

Principles

Safety, security and robustness | Fairness

Example

An organisation pilots a ranking algorithm that assesses candidate qualifications prior to interview. This organisation asks the supplier to test model performance during this pilot.

Performance testing reveals a 20% drop in model performance during the pilot, compared to the results of performance testing demonstrated during procurement. This is communicated to the supplier who investigates causes for the dip in performance.

The supplier hypothesises that the organisation's data labels lead the model to incorrect conclusions. The supplier works with the buyer to update their data governance framework, changing the way data is categorised, which results in improved performance.

Resources

- **DSIT Portfolio of AI Assurance Techniques**
- **TensorFlow Responsible AI Guidance**
- **AI Fairness | ICO**
- **Statistical Accuracy of AI Systems | ICO**

Training/upskilling employees

A process to develop documentation, resources, and/or training for employees that is informed by your pilot of the AI system. During this pilot, your organisation should assess:

- team experience using the tool – do employees find the system easy to use?
- perceived model performance – is the system performing well? Do the outputs seem to be accurate? Fair?
- How often model recommendations were followed – are employees confident to act on the system's recommendations, predictions etc.

Developing and improving training documentation should be treated as an iterative process, with the goal of consistent upskilling of employees. If suppliers provide documentation or tutorials these should be shared with employees.

Principles

Contestability and redress

Appropriate transparency and explainability

Example

An organisation pilots a headhunting software to support the search for qualified applicants. Before piloting, the organisation runs a training session for employees using the system.

During the pilot, employees flag confusion around how to understand why the system has recommend a particular candidate, as the system does not provide rationale for its recommendations.

To rectify this issue, the organisation works with the supplier to understand how candidate skills and/or experience are weighted and incorporates the weighting structure into employee training to provide clarity on the rationale for the system's recommendations.

Resources

- **UK AI Standards Hub Training modules**
- **Tensorflow Responsible AI Toolkit**
- **UK AI Standards Hub events and resources**

Impact assessment

A process to anticipate the wider effects of a system/product on environmental, equality, human rights, data protection, or other outcomes.

Your organisation should complete an impact assessment:

(1) **before procurement**, to identify potential risks and impacts of the **desired system**

(2) **post procurement and pre-deployment**, to identify potential risks and the impacts of **the actual system that you have procured**

Principles

Appropriate transparency and explainability

Fairness

Accountability and governance

Example

An organisation is procuring a chatbot to support applicants in the recruitment process. Before procurement, they complete an impact assessment that identifies that visually impaired applicants may have difficulties using a text-based chatbot. As a result, they ensure the chatbot they procure includes a text-to-speech feature.

During the pilot, the organisation updates the impact assessment, seeking feedback from stakeholders with visual impairments.

The updated impact assessment shows the text-to-speech tool mitigates potential disadvantages faced by applicants who are visually impaired and is positively received by all applicants.

Resources

- **IFOW Algorithmic Impact Assessment**
- **Microsoft Responsible AI Impact Assessment Template**

Ensuring transparency

The use of AI systems in recruitment should be clearly signposted to applicants and potential applicants prior to the launch of the system.

This is essential to allow for contestability and redress, as it is not possible for an applicant to contest decisions, recommendations, or predictions that have been made or enabled by an AI system, without knowing that this system is being deployed.

Where possible, signposting should identify specific limitations of the AI system and how they might apply to individual applicants. Without signposting whether an AI model is used and/or the specific limitations of the system, applicants who could be disadvantaged may not report a need for reasonable adjustments.

Principles

Appropriate transparency and explainability

Accountability and governance

Contestability and redress

Example

An organisation procures a CV scraper to streamline sourcing of candidates.

To ensure transparency around the use of AI, the organisation adds a description of the CV scraper and its purpose in the job description. The organisation then surveys applicants to understand if it is clear that an AI-enabled tool will be used to support the recruitment process.

The consultation shows that 40% of applicants were unaware that AI was being used. To account for this, the organisation places a disclaimer on the use of AI next to the CV upload button in the application form..

Resources

- **How do we ensure transparency in AI | ICO**

4.2 Live operation

Once your organisation has deployed the AI system, you should set up regular monitoring and evaluation to ensure the system continues to perform as expected over time.

Considerations

Ongoing monitoring

An AI system in live operation requires continuous monitoring to ensure that it performs as intended. AI systems may have errors or bugs that reduce the effectiveness of the system. AI systems are also subject to model drift, a phenomenon that sees model performance decay over time due to changes in real world environments, input data, or underlying model goals.

Regular monitoring can help your organisation to identify these and other issues if or when they arise, to prevent a decay in model performance. A failure to proactively monitor for these risks can result in the emergence of harms and a reduction in system efficacy.

Suggested questions

Has your supplier set up a regular testing schedule to measure AI system performance over time?

Relevant assurance mechanisms

Iterative performance testing

Iterative bias audits

Contestability and redress

Ongoing monitoring of the tool that you procure helps to proactively identify potential issues in the AI system. However, it is not realistic to expect that performance testing will be comprehensive enough to identify every possible harm or unintended consequence that may occur.

To supplement performance testing, unintended harms can also be identified by those using and affected by the system, including applicants impacted by the decisions made by the AI system. Providing routes for contestability, where these groups can feedback on issues they have faced, from bugs to bias, can help to flag novel issues and improve the system's effectiveness. Decisions made with or by AI should be contestable if they have the potential to cause harms or violate individual/group rights. Where harms are identified through routes to contestability, appropriate redress should be made.

Suggested questions

Is the fact that AI is being used in the recruitment process clear to potential applicants?

Has your organisation established channels for feedback on the recruitment process?

Are routes for feedback clearly signposted to applicants?

Is human review appropriate if any aspects of the AI's involvement are questioned and/or contested?

Relevant assurance mechanisms

User feedback system

Ensuring transparency

Assurance mechanisms

Iterative performance testing

Prior to deployment, your organisation should have received documentation from suppliers that describe the results of performance testing with respect to model accuracy, effectiveness and fairness.

These tests should be repeated to assess model performance over time. The frequency of testing will depend on your organisation's preferences and capacity.

Principles

Safety, security and robustness | Fairness | Accountability and governance

Example

An organisation deploys a targeted advertising system to improve the quality of applicants for advertised positions. The organisation runs a suite of performance tests every three months to assess the number of qualified applicants that see the advert.

A year of testing reveals that model performance has worsened over time when searching for applicants for software engineering positions.

This issue is communicated to the supplier, who identifies an emerging skill that many qualified applicants have, but was not included in the initial training data. This discrepancy in skillsets on CVs and desired criteria has led to the drift. The supplier retrains the model on data that includes the new skill to rectify the issue.

Resources

- **DSIT Portfolio of AI Assurance Techniques**
- **TensorFlow Responsible AI Guidance**
- **AI Fairness | ICO**
- **What do we need to know about accuracy and statistical accuracy | ICO**

Iterative bias audits

A process for assessing the inputs and outputs of algorithmic systems to determine whether there is bias in input data, or in the outcome of a decision, or classification made by an AI system.

Bias audits should be regularly repeated once the model is in live operation to ensure the system continues to deliver fair outcomes.

Principles

Safety, security and robustness | Fairness

Example

An organisation deploys a job description review tool which has been subject to a bias audit during procurement.

The organisation asks their supplier to run repeated bias audits every six months, and/or when major updates to the system are released.

After a new update to the system, the supplier conducts the bias audit which identifies a drop in the number of female applicants for positions. The supplier informs the organisation and rolls back the update until the issue can be identified and resolved.

Resources

- **How do we ensure fairness in AI? | ICO**
- **Holistic AI: Audits**
- **BABL AI: Third-party audits**
- **IBM Fairness 360**

User feedback system

A process where stakeholders can report issues with an AI system. This includes both employees and applicants who interact with the tool. Feedback systems may include chatbots, surveys or a contact email. At a minimum, feedback systems should:

- Include options for providing a detailed description of the issue (bug, bias, etc.).
- Report the severity of the issue
- Report whether the issue prevented an applicant from progressing
- Be clearly signposted to users (tutorial, website button or verbal communication)

Principles

Fairness | Accountability and governance- | Contestability and redress

Example

An organisation deploys a chatbot to provide answers to FAQs about the hiring process and assist with interview scheduling. The organisation implements feedback routes.

An applicant reports that a bug in the system prevents them from selecting an interview slot

Upon receipt of this feedback, the organisation manually schedules an interview with the applicant and notifies the supplier of the issues with the system. The supplier replicates the issue and resolves the cause of the bug

Resources

- **Google People + AI Guidebook**
- **What is the impact of Article 22 of the UK GDPR on fairness | ICO**
- **Automated decision-making and profiling | ICO**

Department for Science, Innovation & Technology

05

Conclusion

Conclusion

There is no one size fits all approach to AI assurance, and no single assurance mechanism is enough to deem an AI system ‘assured’. The considerations outlined in this guidance should be considered by all organisations seeking to procure and deploy AI regardless of context. AI assurance is an iterative process that should be embedded throughout your businesses practices, to ensure your systems are set up responsibly and for long term success. Your organisation may have the resources to undertake a broad spectrum of assurance mechanisms or only enough to select the key areas of highest risk. It is likely that your organisation’s approach to assurance will draw on a number of these assurance mechanisms, depending on where you are in the procurement lifecycle and the type of system you are seeking to deploy.

Organisations seeking to use AI in recruitment should embed all of these considerations into their procurement and deployment strategy. To address issues discovered during the considerations, organisations should incorporate relevant assurance mechanisms into their governance processes.

DSIT welcomes organisations with ideas or opportunities for future collaboration, insights or resources to share. To get in touch, email us at rtau@dsit.gov.uk

Department for Science, Innovation & Technology

06

Annexes

Annex A: Example use cases and risks

The following is a non-exhaustive list of some of AI tools deployed for recruitment, alongside some risks of the technologies.

Sourcing

AI tool	Description	Risks
Job description review software	Used to help organisations construct, manage, and store job description information. Features include text analysers, keyword optimisation, and regulation compliance.	General text generation tools that are not specialised for creating job descriptions **may result in outputs that are not legally compliant** or may use vague or dissuasive language that discourages prospective applicants from applying. Job description tools must comply with legal requirements on being non-discriminatory. Following EHRC's guidance on what constitutes discrimination in the UK may be helpful in identifying how job description tools may be discriminatory.
Targeted advertising	Used at the sourcing stage of the recruitment process to reach the most suitable pool of candidates and broaden talent pools for recruitment. These tools can also support potential applicants by showing them the job adverts most relevant to their experience and skill set.	Targeted advertising is based on candidate profiling – job adverts are shown to specified profiles that are likely to be a good match for the position. However, profiling can often be based on protected characteristics, which creates a **risk of perpetuating historical biases**. For example, showing administrative roles to women and senior roles to men. Additionally, these tools may target advertisements based on geographic location, which may **impact the mix of racial groups that see an advert**. Many targeted advertising platforms also use automated bidding processes that have been shown to **bias against women** when adverts are optimised for cost, as advertising to women can be more expensive.

AI tool	Description	Risks
Recruiting chatbots	Used to engage with candidates and guide applicants through the recruitment process, which can improve efficiency.	General-purpose chatbots that are not specialised for recruitment and hiring tasks may be inappropriate, as they **may not be trained on relevant or sufficient data**, risking answers being wrong, **potentially illegal**, or even pushing candidates towards other **positions with competitors**.
Headhunting software	Used to help recruiters search for relevant candidates either through a keyword search or more complex identification of active and passive candidates.	Headhunting software may amplify biases if there are pre-conceived notions of an 'ideal' candidate for a position.

Screening

AI tool	Description	Risks
Qualifying screening tools	Automated screening, sifting and ranking tools are a set of data-driven technologies which often use natural language processing (NLP) to evaluate CVs and personal statements and assign them a score. In most cases, the tool will score candidates using keyword search results, based on criteria defined by the employer.	As highlighted our Review into bias in algorithmic decision-making, screening tools are often trained using historical data. This creates a risk that the system might **inherit bias** from past recruitment practices or use proxy indicators for success that are **not relevant to a position**. A sifting tool may also consider gaps in employment in a way that **disproportionately affects parents, people with care-giving responsibilities, people with disabilities or long-term health conditions and neurodivergent candidates**.

AI tool	Description	Risks
CV matching	Also known as resume parsing or CV scraping, these systems are a type of screening tool that extract key data from CVs to find semantic similarities between applicants and pre-determined ideal CVs.	These tools **risk perpetuating existing biases** – for example reinforcing patterns of employment where certain groups are underrepresented (ethnic minorities, women, disabled people) – particularly in sectors that are less diverse, such as engineering, policing, and construction.
Psychometric tests and games	AI-powered games and assessments used to measure cognitive skills including problem solving and working memory, as well as job-relevant personality traits. Usually, these tools produce a report that claims to provide insight on the candidate's personality, thinking style and behaviour.	AI-powered psychometric tests are often argued to perform better than psychometric tests delivered by humans. However, many of these tests **lack scientific validity** whether AI-driven or not, and the outputs of these **systems lack reliability, replicability or objectivity**. Without evidence of scientific validity, psychometric tests risk making **arbitrary recommendations** about applicants. Psychometric tests **may also be inaccessible** for neurodivergent candidates or those using assistive technology.

Interview

AI tool	Description	Risks
Facial recognition in video interviewing	These systems use voice and image recognition combined with inferential biometric technology. They claim to be able to detect emotion, engagement, and desirable candidate qualities and behaviours through expression, posture, and tone during interviews.	Facial recognition systems have been shown to have **divergent error rates** across demographic groups – with the **poorest accuracy consistently** found in subjects who are female, Black, 18-30 years old or have facial differences/paralysis. There is little to **no scientific consensus around the validity** of inferences of emotion.
Asynchronous video interview tools	Interviews where applicants are asked to submit a video recording of themselves answering a set of predetermined questions, where answers are subsequently analysed using automated technology (NLP), human oversight, or a combination of the two.	Many systems use eye detection as a proxy for engagement which may produce **discriminatory outcomes** if, for example, an applicant is: • neurodivergent, as they may find it more difficult to maintain eye contact over a sustained period • a parent/care giver who may be distracted by children or other dependents • Of a cultural/ethnic background where eye contact is seen as inappropriate/ disrespectful

AI tool	Description	Risks
Transcription tools	Transcription tools translate voice recordings from interviews into text data. These tools are often deployed in conjunction with insights systems to perform sentiment and quality analysis on interview transcripts.	Transcription tools can be **biased against regional and non-native English speakers**, and/or individuals with a **speech impediment**.

While AI tools for selection exist, at present there is limited understanding of how industry is using these tools, and what potential risks may be. As such, we are unable to provide an overview and analysis of tools used at this stage of the recruitment process.

Annex B: Glossary of acronyms

Acronym	Name
AI	Artificial Intelligence
DSIT	Department for Science, Innovation and Technology
ICO	Information Commissioner's Office
EHRC	Equality and Human Rights Commission
RTA	Responsible Technology Adoption Unit
DPIA	Data Protection Impact Assessment
AIA	Algorithmic Impact Assessment
EIA	Equality Impact Assessment
GDPR	General Data Protection Regulation
DPA	Data Protection Act 2018
ROI	Return on Investment
NLP	Natural Language Processing
CV/MV	Computer Vision/Machine Vision
FAQ	Frequently Asked Questions

Department for
Science, Innovation
& Technology

For more information contact us at:

rtau@dsit.gov.uk

Department for Science, Innovation & Technology