

Human-Centred Artificial Intelligence for Human Resources: A Toolkit for Human Resources Professionals

TOOLKIT

DECEMBER 2021



Contents

| | |
|---|----|
| Introduction | 3 |
| 1 The big picture | 4 |
| 1.1 The many uses of AI in HR | 5 |
| 1.2 What AI is and how it works | 7 |
| 2 Getting started | 12 |
| 2.1 Forming an assessment team and planning for the long term | 13 |
| 2.2 Determining the purpose of adopting the AI-based tool | 14 |
| 2.3 Delving into the core elements of the tool | 15 |
| 2.4 Assessing the risk level of a tool | 16 |
| 3 Key considerations | 17 |
| 3.1 Bias | 18 |
| 3.2 Data privacy and security | 21 |
| 3.3 Transparency and explainability | 22 |
| 4 Implementation and buy-in | 25 |
| 5 Ongoing maintenance and monitoring | 27 |
| Tool Assessment Checklist | 29 |
| Planning Checklist | 49 |
| Acknowledgements | 55 |
| Endnotes | 58 |

Disclaimer

This document is published by the World Economic Forum as a contribution to a project, insight area or interaction. The findings, interpretations and conclusions expressed herein are a result of a collaborative process facilitated and endorsed by the World Economic Forum but whose results do not necessarily represent the views of the World Economic Forum, nor the entirety of its Members, Partners or other stakeholders.

© 2021 World Economic Forum. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, or by any information storage and retrieval system.

Introduction

Artificial intelligence (AI) has gained considerable attention and excitement in recent years. Broadly defined as the effort to program computers to take on human-like cognitive processes, the recent prominence of AI is closely tied to the success of machine learning (ML), an approach to developing AI systems using real-world examples. The ML approach is applicable to a surprisingly wide variety of use cases; therefore, there is a proliferation of AI-based tools in every sector of the economy and of life.

The field of human resources (HR) is no exception. Indeed, by one count there are over 250 different commercial AI-based HR tools available.¹ These tools offer a lot of promise and excitement. Beyond their ability to process information quickly, these tools have the potential to improve HR processes, leading to better decisions and outcomes. Their variety reflects the creativity and innovation spurred by recent advancements in AI, as their creators seek to both tackle long-standing challenges in HR and expand capacities into new realms.

At the same time, this proliferation and variety of tools creates a confusing landscape to navigate, especially because most HR professionals do not feel that they have the technical expertise required to evaluate these tools. The first goal of this toolkit, therefore, is to equip HR professionals with a basic

understanding of AI to assist them in their efforts to assess AI-based tools.

The second goal of the toolkit is to provide guidance on the responsible and ethical use of AI in HR. Awareness of the ethical challenges that AI systems can pose has been growing in recent years, concerns that are particularly heightened in the HR context. There is increasing consensus globally about broad principles for the ethical use of AI, including privacy, fairness, transparency and explainability, but only limited guidance on how to operationalize these principles. This toolkit is part of a broader effort by the Centre for the Fourth Industrial Revolution to help organizations put responsible AI principles into practice.

The final goal of the toolkit is to help organizations use AI-based HR tools effectively. Many organizations find their investments in AI fall short of their expectations because the tools are adopted for the wrong reasons, they do not anticipate the work necessary to integrate the tool, or because they did not gain sufficient buy-in from the people who were supposed to use it or are affected by it. The toolkit and especially the accompanying checklists, therefore, focus on assessing AI-based products and on the organizational practices needed to support their use.

Balancing perspectives on AI in HR

This toolkit is a collaborative effort bringing together HR professionals and professional associations, start-ups, larger companies, employment lawyers, AI ethicists, data scientists, and academics from a range of disciplines. They share a common desire to promote the responsible use of AI in HR, yet vary in their views and concerns. At one end of the spectrum are individuals who are very concerned about the potential downsides of using AI in HR. At the other end are individuals who, while recognizing the need to implement AI responsibly, strongly believe in the potential of AI-based tools to improve HR outcomes. A tension in using AI in HR is the

need to acknowledge the shortcomings of HR practices as they are currently performed, whether by humans or by non-AI systems such as keyword filters and assessment tests. AI systems tend to face greater scrutiny than these other methods. While some community members felt that this scrutiny was necessary, others felt that it ignored similar or potentially larger problems with current practices. The toolkit aims to present these different perspectives, combating the misperception that AI algorithms are intrinsically objective and fair, while at the same time highlighting the need to recognize the flaws in current practices.

The structure of the toolkit

The toolkit consists of three components. The guide provides an overview of AI in HR, how AI works, and key considerations for the responsible adoption and monitoring of AI systems. Two questionnaires that parallel each section of the guide accompany it. The [Tool Assessment Checklist](#) focuses on the decision to adopt a specific AI-based HR tool. It includes both questions to ask

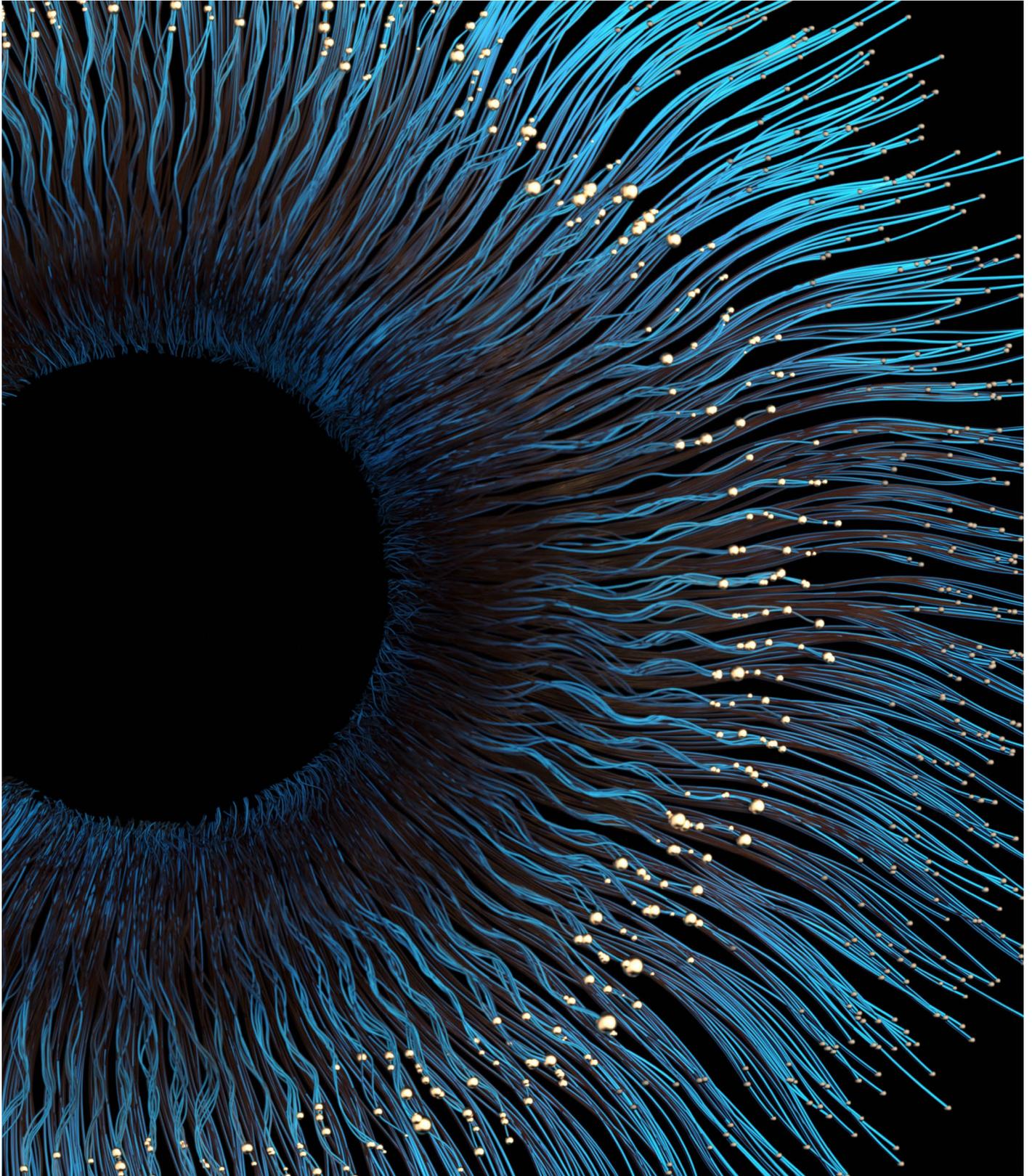
the vendor (or internal creator of the tool) as well as questions for the organization to consider for the successful use of the tool. The [Planning Checklist](#) focuses on organizational priorities, policies and procedures. Its aim is to assist organizations in thinking strategically about how they want to use AI in HR and to put into place systems to support its responsible and effective use.

“ At one end of the spectrum are individuals who are very concerned about the potential downsides of using AI in HR. At the other end are individuals who, while recognizing the need to implement AI responsibly, strongly believe in the potential of AI-based tools to improve HR outcomes.

1

The big picture

AI-based HR tools vary in their tasks and goals, but share a common method that is important to understand.



This section provides an overview of artificial intelligence (AI) in human resources (HR). First is a discussion of the broad categories of the different AI-based HR tools available. Second is an explanation of the basic principles of how AI works, which are important to understanding both the power and the limitations of these systems.

1.1 The many uses of AI in HR

Key takeaways

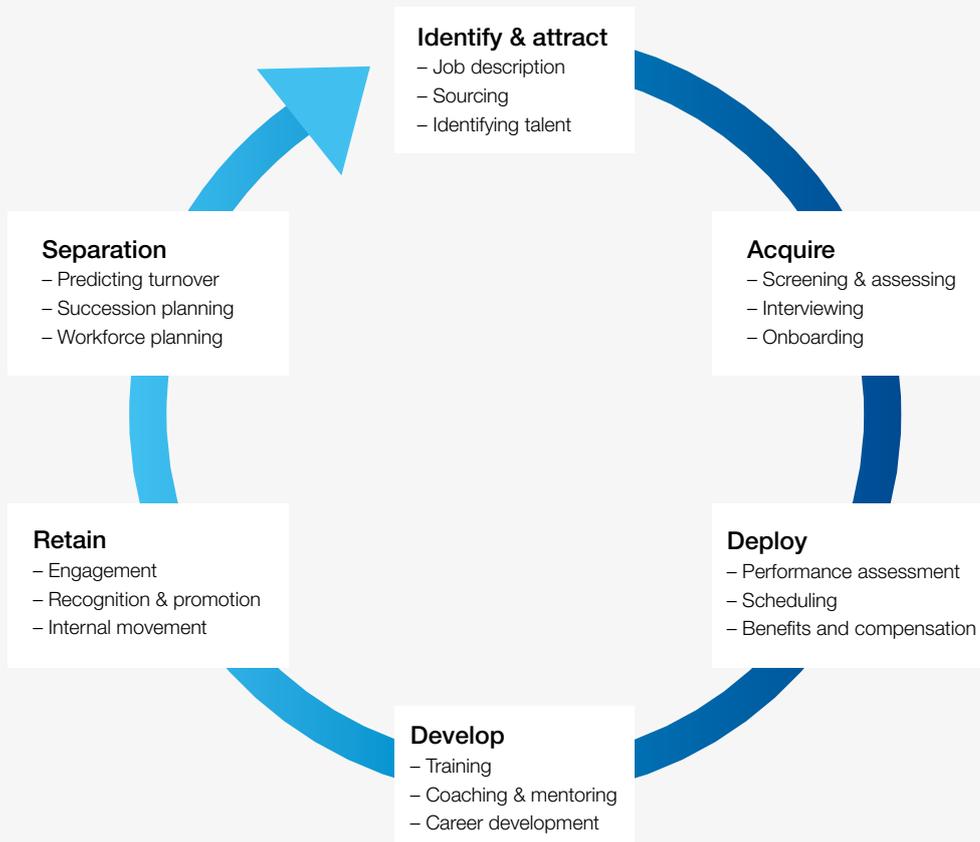
- Developers are creating AI-based tools for almost all aspects of human resources.
- Tools seek to improve organizational processes through a few different approaches.
- Human resources departments should periodically check for AI-based tools that the organization may have already implemented without full review.

AI is a general-purpose technology, which means that it has a wide range of uses in HR. Each AI-based tool is slightly different but it is possible to broadly categorize them by answering two questions: 1) What task is it doing? and 2) How is the tool promising to change current processes?

What task is the tool doing?

While the most attention has been paid to AI-based tools for hiring, AI has the potential to be used in almost every facet of HR across the full spectrum of the employee life cycle.

FIGURE 1 Developers are creating AI tools for almost every stage of the HR life cycle



Despite this wide range of possibilities, though, AI will not be taking over HR completely any time soon. Current AI systems, while powerful in some ways, are still limited in many other ways. Developers therefore design most tools to take over a specific part of an HR task rather than replace the human completely.

How is the tool promising to change current processes?

Organizations may adopt AI-based HR tools for a number of reasons; and different tools make different promises for how they can help an organization. Below is a list of the claims that the creators may make about their product. Note, however, that fulfilling these promises requires thoughtful design and therefore careful scrutiny before adoption. The remainder of the guide aims to help you better assess the tool's claims.

- **Speeding up or taking over basic tasks (automation).** Creators design some tools to take over tasks that are particularly tedious and repetitive, replicating the human process but doing it faster and saving human effort. Automation should only be considered, though, for tasks that are relatively objective and low-risk, and some human monitoring is likely still necessary.
 - *Examples:* resume parsing system, interview scheduling chatbot
- **Improving or changing processes (augmentation).** Quite a few tools go beyond automating a task and look to change processes in one or more of the following ways.
 - **Maximizing/predicting a specified outcome.** AI systems are usually designed to maximize or predict a particular outcome (employee performance, worker engagement, etc. See the chapter on What is AI and how does it work?). If there is a good, measurable outcome then an AI tool may identify ways to improve/better predict that outcome.
 - *Example:* recommending sales job candidates based upon predicted sales
 - **Bringing new information to an existing task.** Some tools use AI to bring in new forms of information, often information that

would be hard or too time consuming for humans to collect or analyse.

- *Examples:* new ways of assessing job candidates, summarizing key insights from free-form answers in an employee survey
- **New tasks.** AI is also being used to take on new tasks or do an existing task in a fundamentally different way. This can also include more personalized and tailored services.
 - *Examples:* predicting turnover, personalized career path recommendations
- **Pursuing specific goals.** Creators may design AI-based HR tools to address specific goals, one of the most common of which is improving diversity and inclusion. Some tools promise to make the task fairer, with the view that this fairness will itself increase diversity and inclusion. Other tools may target the specified goal more directly, in some cases solely focusing on that goal rather than taking on an HR task more broadly.
 - *Examples:* recommending job posting language that attracts more diverse applicants, automatically masking the gender and race of an applicant
- **Complete systems (autonomous).** An autonomous system would go beyond the automation of simple tasks to have an AI system that takes responsibility for more complex or higher-stakes decisions. Few AI in HR tools seek to be fully autonomous and **it is strongly recommended not to delegate high-stakes decisions to autonomous AI systems.**

“ Organizations may adopt AI-based HR tools for a number of reasons; and different tools make different promises for how they can help an organization.

You may already be using AI

Many HR information systems are adding AI-based tools to their offerings, while AI vendors for other functions such as marketing may offer HR tools as add-on packages. In addition, some companies initially implement an AI-based tool on a trial basis and never review it. It is important, therefore, to occasionally scan for AI tools that might already be in use and ensure that you have also properly reviewed them.



Keep an eye out for these other terms that are synonymous with AI: machine learning, predictive analytics, decision algorithms, recommendation engines. In contrast, people analytics does not necessarily involve AI. AI-based tools use

data to automatically generate predictions, recommendations or decisions. Non-AI analytics may include dashboards to track key metrics or the use of data to identify and diagnose HR issues within the organization.

[Tool Assessment Checklist](#)

[Planning Checklist](#)

1.2 What AI is and how it works

Key takeaways

- Most AI-based HR tools use some form of machine learning (ML), an approach that creates an algorithm based upon patterns identified in real-world examples called training data.
- To understand how the system will work, it is important to know the source of the training data, as well as the inputs and outcome (if applicable) the ML system is using.

You do not need to be a genius to understand the basic nature of current AI systems. While the inner workings can be complex, almost all current AI systems are a type of machine learning (ML), which

share a basic underlying principle. Knowing a bit about this underlying principle will help you understand the strengths and limitations of AI-based tools.

Identifying patterns in data

Both the incredible power and the greatest limitation of machine learning is its reliance on training data.

Traditional programming vs machine learning

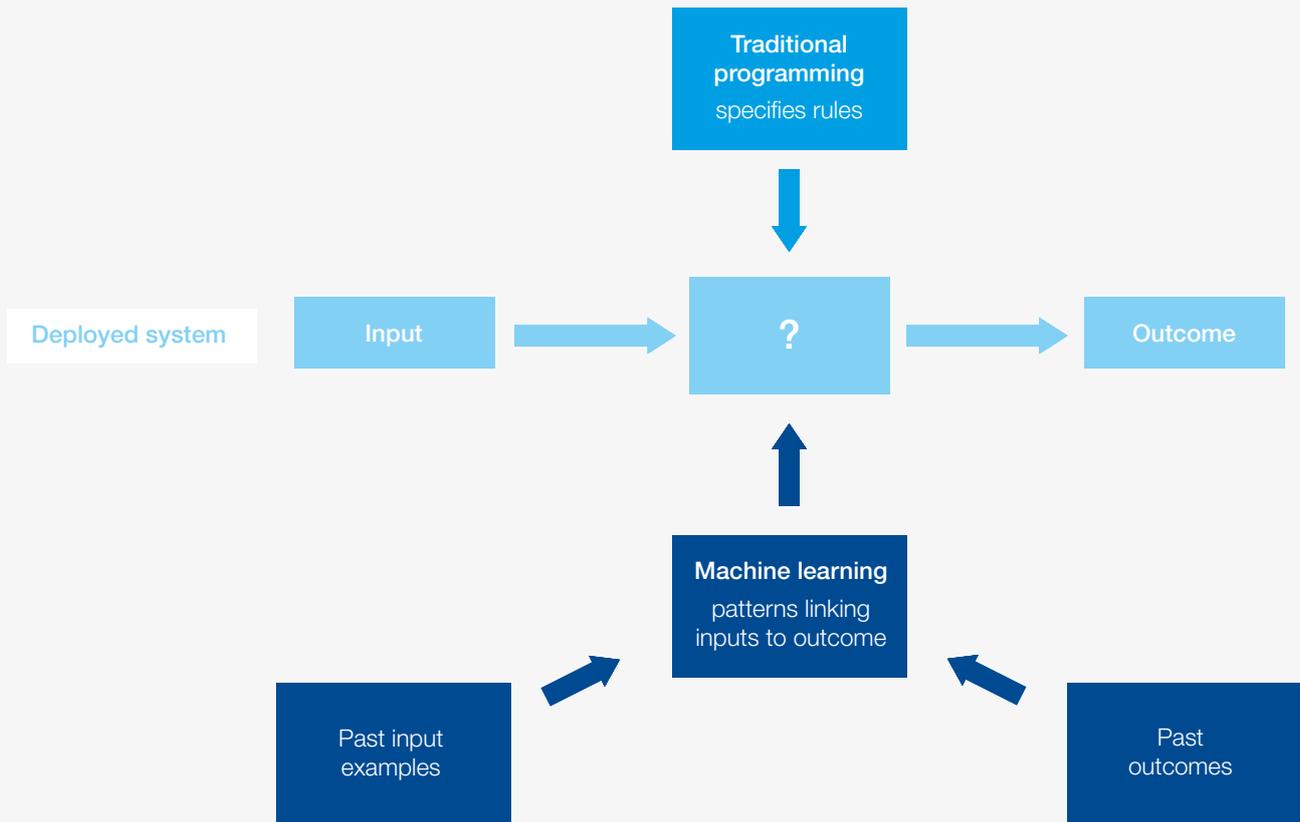
To understand the power of machine learning, it is necessary to contrast it with a traditional computer program. When writing a computer program, programmers specify every step and rules for each decision. The problem with the programming approach is that it will usually lack nuance. Just like a bureaucracy, it will be too simplistic and rigid.

Machine learning is an umbrella term for many different methods that share a common approach. Machine learning methods use **training data** consisting of **many examples of a task**, usually past examples. It then **looks for patterns in the**

training data. Following that, it **assumes that these patterns will hold when applied to new cases** in a deployed system.

For example, a machine learning algorithm designed to predict turnover might look at historical records to identify patterns in the types of people who have left the organization in the past. Or an algorithm seeking to identify high-potential job candidates might look at past hires and their subsequent performance, identifying combinations of candidate characteristics that correlate with higher or lower performance.

FIGURE 2 | Traditional programming vs machine learning



ML systems lack a key element of human intelligence: the ability to learn lessons and apply them to new situations. In this sense, **ML systems are not truly intelligent and will only work for the specific task for which they are trained.** However, unlike traditional programming, it can be a surprisingly complex task. An AI system that can truly learn (sometimes called

artificial general intelligence) does not yet exist. For the foreseeable future, therefore, it will be necessary to train AI algorithms for a given task.

The training data determines how an ML system will operate; thus, it is critical to consider the **source** and **content** of the training data.

Training data source

In order to build a successful ML algorithm, it is necessary to identify a source of training that is both **relevant** and **large enough** to train the system. In HR settings there may be trade-offs between these two goals.

Relevant training data

The first requirement for a successful ML algorithm is training data that matches well with the context where the algorithm will be deployed. If the training data does not match the deployment context, then the patterns the ML systems identify may not be applicable. In the HR setting, it is necessary to consider several dimensions of the context, including differences in jobs, individual factors (e.g. age, gender, etc.), location, organization and cultural contexts, and differences over time.

The type of task may determine the level of specialized training data required to be successful. For instance, it may be necessary

to customize an algorithm seeking to identify the best job candidate for the specific job, while a tool predicting turnover might be relevant to multiple jobs but perhaps specific to the organization and cultural context. The key takeaway is that it is **important to know the source of the training data and consider whether the patterns that the ML system has identified in that data are likely to be relevant for the context in which you will deploy the tool.** It is also important to consider whether the training data might capture biased or other problematic processes, as discussed in the chapter on Bias.

Enough training data

ML algorithms generally require a lot of data. A large training dataset allows the ML system to identify complex and subtle patterns, while a small dataset risks algorithms that are too simple and therefore not accurate when deployed. In an HR setting, however, few organizations have hundreds of

thousands of employees and the available data may be even smaller if focusing on a specific job within an organization.

Some AI-based HR tools address this issue by **pooling training data** spanning jobs or organizations. Pooling the data allows for a more complex algorithm. It might also provide a greater variety of examples to consider, including providing information about social groups that are underrepresented in your own

organization. Pooling, however, may pose problems if the data is very different from the context where you will be using the tool.

Some vendors instead **rely on just a few selected examples**, for instance a few top performers in a position, rather than a large training dataset. This approach may allow customization but may find people who are similar to those top performers without discerning which traits actually matter.

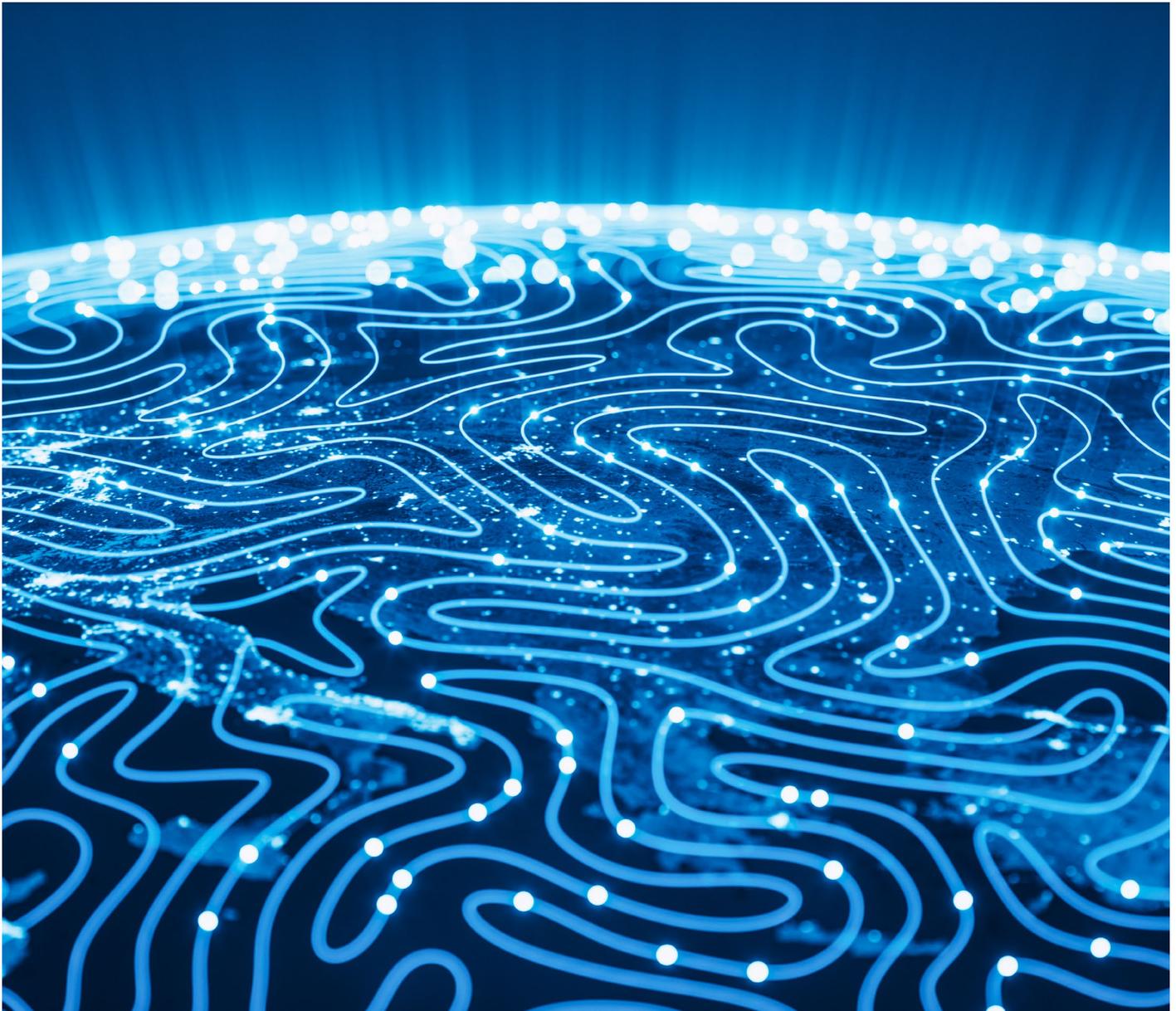


Specialized AI systems

Natural language processing (NLP) refers to ML systems designed to extract meaning from text or speech. They are often created by a third party specializing in NLP and then used by the tool creator. Many NLP systems are quite advanced, but designed for general use so it is necessary to pay attention to how the context or users in your case might use language differently.

Chatbots are a common tool, including in HR.

They generally involve a front-end that tries to figure out what the person is asking, and a back-end that decides a response. Most chatbots will use NLP for the front-end. The back-end is often rules-based programming but can also use ML, with the training data being past examples of questions and answers from human interactions. The ML system will then identify patterns that link the texts of different questions to their corresponding answers.



Training data content

Building an AI system requires translating an abstract idea into concrete training data. Programmers use their judgement on what factors are relevant; they are limited to what is measurable and available.

It is important, therefore, to understand what measures are included in the training data, why they are relevant to the task, and what is and is not captured in that information. The HR context can be challenging because it often involves factors that are difficult to measure. While this challenge is not unique to AI systems, using AI does not automatically solve these problems. If you know that there are limitations or problems with the data, you should clearly convey these limitations to the user of the AI output.

The training data will typically include two parts: **inputs** and an **outcome**.

Inputs

The training data will include a number of input measures, and the system will collect these same inputs for new cases when deployed. In the example of job applicants, the input measures would be characteristics of the applicants when they apply. For predicting turnover, the inputs would be various characteristics of employees believed to influence the likelihood of leaving. **Knowing which inputs are included in the algorithm and which are not will give you an idea of how the algorithm will operate.** In some cases the ML algorithm may not work very well because key inputs are not available, perhaps not even measurable. Humans may be able to take into account some of these unmeasured factors but there is also evidence that such “gut” assessments can be incorrect or subject to biases.² In other cases, ML systems may be able to bring in new measures, process large amounts

of data, or identify patterns overlooked by humans. In this way they might offer improvements over human decisions. In the HR setting, though, neither ML systems nor humans are likely to be perfectly accurate, especially if trying to predict into the future, because it is usually impossible to account for all of the factors and random events that can affect social processes.

Outcome

Most ML systems are designed to predict or maximize an outcome. For instance, an applicant screening system might be set to predict future performance; a turnover prediction program would predict leaving the organization. The ML system will then use the examples in the training data to try to identify patterns that link the input factors to that outcome: what combinations of characteristics of past job applicants tend to occur among high versus low performers? What characteristics of past employees are associated with them leaving? These patterns will then form the basis for the algorithm’s predictions of future cases once deployed.

An ML system requires a measurable outcome; so it is necessary to turn abstract ideas such as “a high-potential employee” into something measurable in the training data, such as an employee’s performance rating. The ML system will be trained solely on the specified outcome; therefore, it is important to identify that outcome and consider what it does and does not capture.

Not all ML systems will focus on an outcome: Some ML systems seek to simply group similar items together. For instance, a recommendation algorithm might identify people who are similar to you and use their choices as recommendations.

Comparing human and machine decisions

The project community brainstormed the strengths and weaknesses of both human and machine learning decision-making processes, as summarized in the table below.

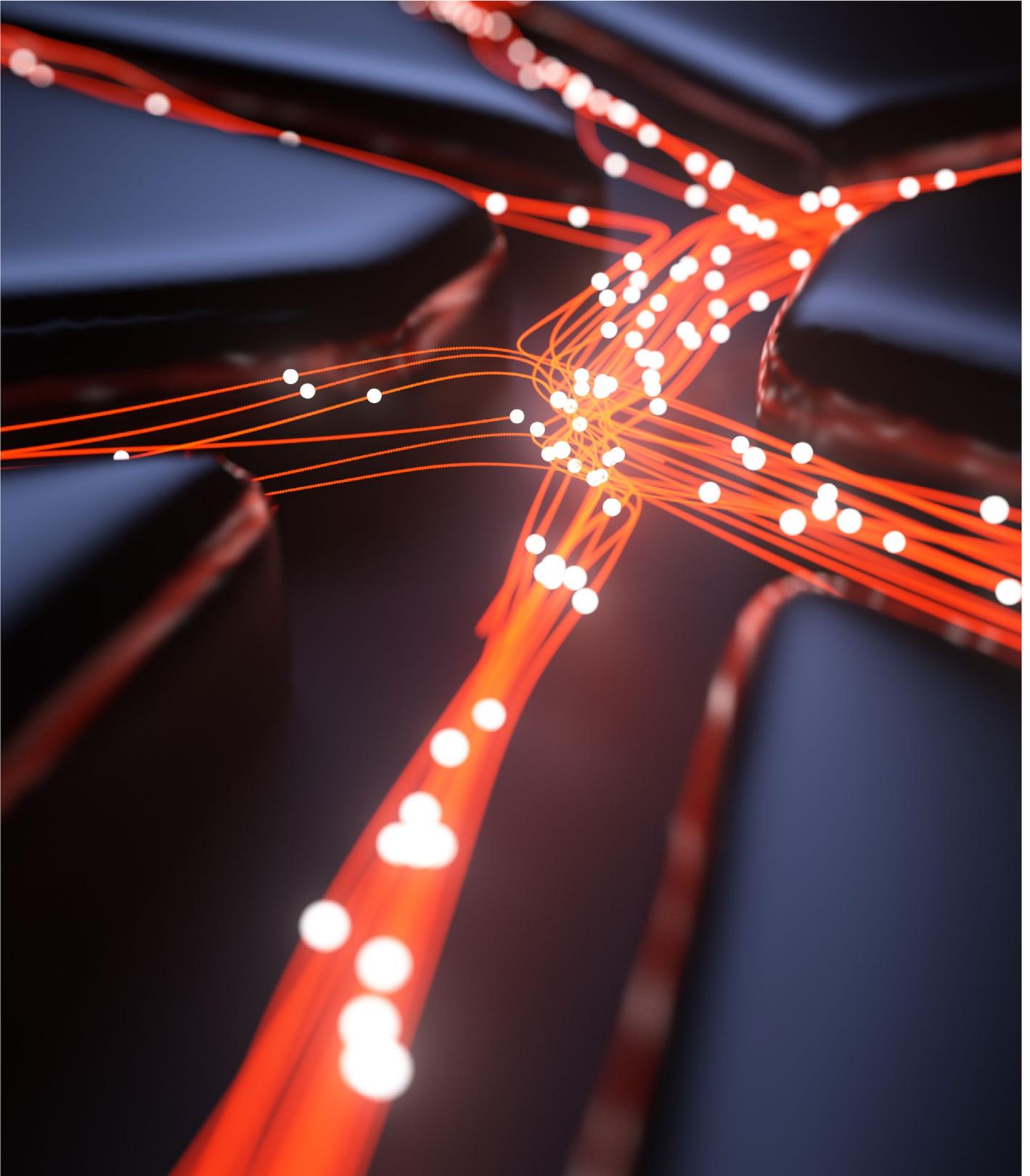
TABLE 1 | Strengths and weaknesses of human and ML decision-making processes

| Human | Machine learning |
|--|---|
| Basis for understanding the world | |
| <i>Life experiences</i> | <i>Training data</i> |
| <ul style="list-style-type: none"> ✓ Can make educated guesses in novel situations | <ul style="list-style-type: none"> ✓ Makes decisions based upon training data rather than assumptions |
| <ul style="list-style-type: none"> ✓ Has “common sense” | <ul style="list-style-type: none"> ✓ Can identify patterns overlooked by humans |
| <ul style="list-style-type: none"> ✓ Better at perceiving causal relationships | <ul style="list-style-type: none"> ○ Won’t work as well in contexts that differ from training data |
| <ul style="list-style-type: none"> ✓ Ability to recognize when conditions have changed or when cases are special | <ul style="list-style-type: none"> ○ May struggle with unusual cases not represented in training data |
| <ul style="list-style-type: none"> ✓ Can envision fundamental changes to processes | <ul style="list-style-type: none"> ○ Can’t check that a decision “makes sense” |
| <ul style="list-style-type: none"> ○ Overreliance on “gut” feelings and assumptions | <ul style="list-style-type: none"> ○ Limited to considering inputs specified in the algorithm |
| <ul style="list-style-type: none"> ○ Can bring in irrelevant information | <ul style="list-style-type: none"> ○ Can learn and amplify human bias reflected in data |
| <ul style="list-style-type: none"> ○ Unconscious and conscious bias can lead to discriminatory or harmful decisions | — |
| Speed/capacity | |
| <ul style="list-style-type: none"> ○ Limited speed, limited capacity compared to AI | <ul style="list-style-type: none"> ✓ Once trained, large capacity for speed and scale |
| <ul style="list-style-type: none"> ○ Limited in number of factors can consider | <ul style="list-style-type: none"> ✓ Can potentially consider a large number of factors (once specified by humans and with enough training data) |
| — | <ul style="list-style-type: none"> ○ Ability to scale means the flaws may have widespread impacts |
| The human element | |
| <ul style="list-style-type: none"> ✓ Capable of empathy and emotional intelligence | <ul style="list-style-type: none"> ✓ Could provide feedback that humans would find difficult to say |
| <ul style="list-style-type: none"> ✓ Can communicate with sensitivity | <ul style="list-style-type: none"> ○ Humans may not trust, may view process as dehumanizing |
| <ul style="list-style-type: none"> ○ Unconscious bias and preference for people similar to oneself | <ul style="list-style-type: none"> ○ Can give the illusion of objectivity while in fact amplifying human bias |
| Responsibility, liability and transparency of decisions | |
| <ul style="list-style-type: none"> ✓ Can be held responsible for decisions | <ul style="list-style-type: none"> ✓ Can provide consistency to process and decisions |
| <ul style="list-style-type: none"> ○ Decisions are often not well-explained | <ul style="list-style-type: none"> ✓ Creates documentation/paper trail of decision (even if difficult to understand) |
| <ul style="list-style-type: none"> ○ Susceptible to corruption or personal gain | <ul style="list-style-type: none"> ○ Who is responsible/liable? |
| <ul style="list-style-type: none"> ○ Can involve bias and favouritism | <ul style="list-style-type: none"> ○ Can be difficult to provide an understandable reason for decision |

2

Getting started

Gathering the right people and identifying the key opportunities and risks of an AI-based HR tool.



2.1 Forming an assessment team and planning for the long term

Key takeaways

- Involving individuals from multiple areas of the organization in the decision to adopt an AI for HR tool will head off problems early and ensure successful implementation.
- Consider developing an evaluation protocol for AI in HR tools to guide both current and future decisions.

Evaluating an AI for HR tool starts with having the right people involved and a plan of action.

Buy-in from the chief human resources officer (CHRO) and organizational decision-makers will be critical to the successful adoption of a tool. The proponents for using AI in HR will need to develop a pitch for the adoption of the tool, including information about expected costs, risks and benefits.

Who should be part of the assessment team

The decision to adopt an AI-based HR tool should involve multiple parties from an early stage. This multistakeholder approach will help you anticipate major problems, better evaluate the match between the tool and your organization's capacities, and gain buy-in to ensure that staff actually use any tool adopted. Below is a list of individuals who should be part of the decision-making process.

- **People with knowledge and experience with the task.** These “subject matter experts” are the people most familiar with the task that the AI-based tool will be undertaking, such as the people currently doing the task or outside experts. These individuals can assess whether the tool's overall concept is sound, if the data that it is using is relevant and well-measured, and how to use the tool effectively and efficiently.
- **HR team members.** In addition to subject matter experts, individuals from HR should include:
 - Those who will use the tool and act on its output
 - Someone with a high-level perspective on how this tool might impact other areas of HR
 - HR business partners.
- **Employee representatives.** It is important to bring in the perspectives of the individuals who will be the subject of AI-based tools. While the adoption of any new practice raises risks, an AI-based tool has the potential to receive particularly strong negative reactions if it is seen as being imposed or “taking the human out of human resources”. It is critical, therefore, to gain insight into how stakeholders would perceive the tool and involve them, including trade unions

or worker representatives, in the tool's selection and implementation.

- **Information technology.** The IT department will be familiar with current data management, security and storage systems, and can anticipate problems integrating the new tool. AI/ML is a specialized area of computer science, however, and many IT professionals will not have expertise in this area.
- **Legal.** While people often wait until the last minute to bring in legal counsel, involving them early in the process can avoid roadblocks later.
- Also consider **other relevant departments**, depending upon your organization's configuration, such as compliance and procurement.

In addition, the individuals listed below with specific expertise will provide valuable input. For organizations that do not have in-house expertise in these areas, you may wish to look for external help, especially when considering the adoption of high-risk tools (see chapter on Assessing the risk level of a tool).

- **Data scientists.** Data scientists with knowledge of machine learning can be a valuable resource in evaluating AI-based tools. HR data analysts may be particularly valuable because they also understand HR practices and needs; but they may not have expertise with AI/ML.
- **Diversity and inclusion experts.** They can help identify opportunities for positive impacts and anticipate possible issues of bias.
- **Accessibility experts.** They can help to ensure that the tool is accessible for everyone.

“ This multistakeholder approach will help you anticipate major problems, better evaluate the match between the tool and your organization's capacities, and gain buy-in

- **AI ethicists.** AI ethics is a growing area of expertise, with a number of companies now employing in-house AI ethicists or hiring AI ethics consultants.
- **Data privacy expert.** You may want to involve someone with expertise in data privacy to ensure proper handling and compliance.
- **Localization team.** If your organization operates in multiple locations or if the tool was developed

in a different cultural context, you will want to include individuals with knowledge of each local context to evaluate the extent to which the tool is appropriate or may need adapting to each setting.

The assessment process will only be effective if all members feel empowered to raise concerns and ideas. You will need to assign a project manager who is responsible for guiding the process, documenting decisions, and following through on action points.

Developing AI principles and documenting processes

Check to see whether your organization already has a set of AI principles or policies. If not, consider developing such policies, including ones specific to HR.³ AI principles will describe your priorities and goals for the use of artificial intelligence and will outline what you won't do with AI. For instance, one principle that is particularly relevant to HR is the commitment to use AI to augment rather than replace workers.

In addition, scan the organization for existing governance structures that might impact the process, for instance data policies, HR policies, procurement guidelines, and information and consultation agreements with worker representatives that you need to follow.

Finally, it is recommended that you create a log of your decision-making process, to both keep a record of what you have done and to begin to develop best practices.

[Tool Assessment Checklist](#)

[Planning Checklist](#)

2.2 Determining the purpose of adopting the AI-based tool

Key takeaways

- Be clear on why you are considering an AI-based tool and how you anticipate it will compare to the process as it is currently done.
- Ensure that the tool is appropriate for your organization's capacities and AI journey.

As you begin to consider an AI-based HR tool, you will want to clarify your purpose and priorities. Consider **the problem you hope to solve** through the use of this tool. Before focusing on the tool itself, discuss the process as it currently operates, involving both those who typically do this task as well as those who are affected by this task. Consider how the proposed tool might change that process, how it might both improve the system and raise new challenges.

Identify **how you anticipate this tool will bring value to your organization**, for instance by affecting a specific outcome or key performance indicator. This information will be useful in assessing the value of the tool and should also be integrated into your implementation and monitoring plans.

The value of adopting a tool may be that it will help you build experience using AI, considering **where the organization is in its AI journey**. The successful implementation of AI requires multiple capacities, including data infrastructure, buy-in from leadership, and trust among the users and subjects of the tool (workers or job candidates) to ensure effective use. Consider the direct outcomes of the tool and how it fits in to a larger strategy of building AI capabilities and credibility. If you are early in the process, you might need an initial project that serves as a proof of concept. You might look for quick wins where the impact might be less dramatic but the resources required are smaller and the likelihood of success is high.



Will you be able to use the results of the AI-based tool effectively? For example, a turnover prediction tool looks to identify individuals who have a higher likelihood of leaving the organization.

This might be interesting information but what will you do with it? Who will be responsible for it? Perhaps you will look to convince key individuals to stay but how would you do this?

[Tool Assessment Checklist](#)

[Planning Checklist](#)

2.3 Delving into the core elements of the tool

Key takeaways

– Each AI-based tool is different, even when tackling seemingly similar tasks. You need to examine the details of the tool under consideration.

– The accompanying questionnaire provides a list of questions to help you delve into the key components of the tool.

Just like a snowflake, no two AI-based HR tools are the same, even when they are taking on similar HR tasks. The specific design of the tool will reflect the ideas and decisions of their creators, so it is important to understand the specific tool and its key components.

Tools often involve multiple components and steps, and AI may play a role in only a few of these steps. The accompanying questionnaire provides a list questions to consider when examining an AI for HR tool. These questions aim to help you grasp the overall design of the tool as well as the key details of the AI components.

After exploring the details of the system, revisit the tool's big picture. Consider whether the tool seems reasonable and well-designed, and how it compares to how your organization currently undertakes the task. Since creators strongly influence their AI tools, consider the backgrounds and expertise of the developers. While AI can be a powerful tool, be cautious of products that promise extraordinary results but do not fit with existing evidence, your own understanding of HR systems, or the knowledge of experts.

[Tool Assessment Checklist](#)

[Planning Checklist](#)

2.4 Assessing the risk level of a tool

Key takeaways

- Different AI-based tools pose different levels of risk to the organization and its users. The risk worksheet provides a starting point in assessing the risk level of a given tool.
- Higher risk tools will require greater scrutiny before adoption, greater care in implementation, and closer and more frequent monitoring.

While it is important to evaluate whether and how a tool will help your organization, it is also important to consider the nature and scale of the consequences if something goes wrong, including legal, organizational, ethical and reputational risks. **The presence of risk does not by itself mean that you need to abandon the tool. The larger the consequences, though, the more intense the scrutiny should be, along with the care that will be required to ensure that proper safeguards and monitoring are put in place** by both the tool creators and the organization deploying the tool.

In assessing these risks, you should consider:

- The tool's intended use
- Its potential for misuse (intentionally or unintentionally) or misunderstanding
- How it might change surrounding behaviours, including attempts to game the system

- The consequences for both the organization and individuals if the tool fails to perform as expected or provides false predictions
- The consequences if people start over-relying on this tool
- The risks involved with how the organization is currently completing the task and how adopting the tool might change this.

The accompanying risk worksheet provides a series of points to consider in assessing the risk level of an AI for HR tool. After assessing these and other risks, revisit the question of why you are adopting the tool. If you feel the benefits are justified, consider the overall risk level and its implications for the level of scrutiny the tool should undergo, who needs to be involved in the approval process, what safeguards and other preparation needs to be in place, and the intensity of monitoring that is necessary. Read the chapter on Key considerations and then reassess these questions once more.

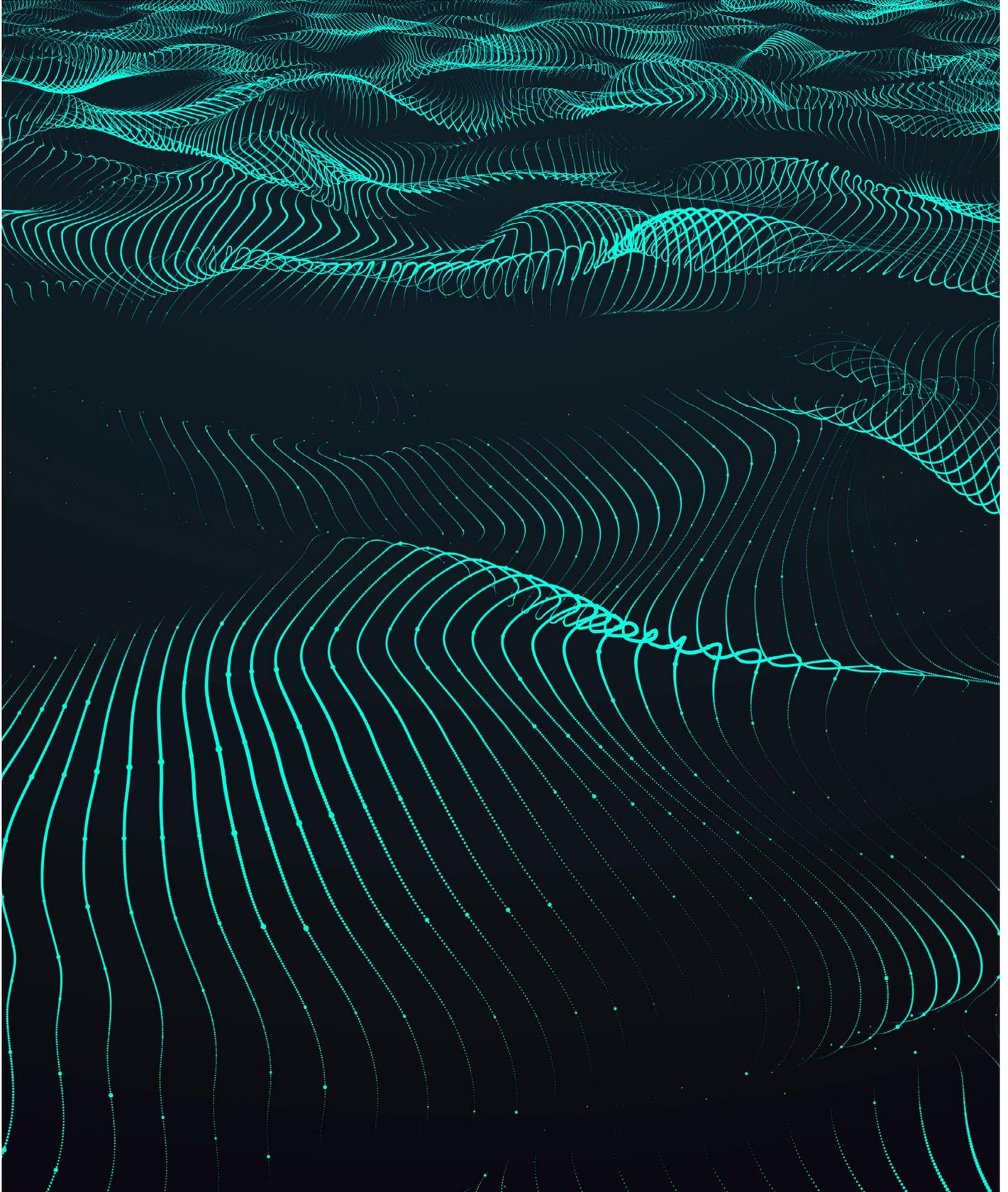
[Tool Assessment Checklist](#)

[Planning Checklist](#)

3

Key considerations

An overview of the key principles for the responsible use of AI in HR.



Concern about several key aspects of AI systems has been growing in recent years. The following sections provide a brief background on each of these areas and how they may apply to AI-based tools in HR. These key considerations are central pillars in the responsible use of artificial intelligence. Careful consideration of these issues will help steer you toward AI in HR that is effective and reduces operational, reputational, ethical and legal risks.

3.1 Bias

Key takeaways

- The potential of AI to reduce bias stems primarily from the innovative and careful designs of its creators, while poorly conceived AI tools risk exacerbating bias and misuse.
- When examining an AI-based tool, consider the representativeness of the training data, what the data measures, and which measures might incorporate human biases.
- The tool should have a system to test for biases both before deployment and as part of ongoing monitoring.

The issue of bias is one of the biggest concerns regarding the use of AI for HR tasks. **While algorithms are often viewed as objective and impartial, they in fact have the potential to encode and amplify existing biases.** This is a heightened concern in the HR context, since HR decisions affect critical life outcomes. Biased algorithms also pose considerable reputational and legal risks for organizations.

A biased algorithm is one that results in outcomes for people from certain social groups that would be considered unfair or unjust.⁴ Bias is not unique to AI systems; in fact, humans are the source of bias in these systems through the assumptions that they make when designing them as well as through the training data. Similar biases can and do affect other approaches to doing HR tasks, including both simpler systems (for example keyword filters

for job applicants) as well as human decisions. Indeed, in many cases organizations are adopting AI-based tools to help move HR processes away from human “gut decisions” that may rely on stereotypes and biases. AI systems tend to face greater scrutiny, though, because of mistrust of complex algorithms, concerns that they create the illusion of objectivity, and concerns that they might have more uniform impacts on a larger scale.⁵

The following sections focus on the different ways that it is possible to introduce bias into AI algorithms. The aim is to help organizations minimize such biases. When considering an AI-based tool, however, it is also necessary to consider biases that likely exist in how humans currently perform the task, how an AI-based tool might reduce and/or exacerbate these issues, and what improvements you could make to the task regardless of whether you adopt AI.

Overall conception and the algorithm’s purpose

A well-designed AI system has the potential to reduce bias

Machine learning-based AI tools do not automatically reduce bias. There are ways, though, that AI can improve on existing processes.

Focusing on a well-defined outcome. Sometimes an AI system can reduce biases by training it on a well-defined outcome, as long as that outcome does not itself incorporate major biases. The algorithm may reduce biases compared to humans if the humans are not as focused as they should be on that outcome (e.g. favouring people they like instead) or if the algorithm identifies patterns

in the training data overlooked by humans (e.g. employees without elite university degrees perform just as well as those with elite degrees).

Rethinking the way things are done. Some AI for HR systems go further by reimagining how a task can be done, for instance by bringing in new information or performing the task in a fundamentally different way. Here AI is serving as a tool to realize this idea; whether the system does in fact reduce bias will ultimately depend on the effectiveness of the innovation.

Problematic purposes, designs and uses

In other cases, however, creators may poorly conceive AI-based HR tools, they may take for granted unjust social processes and encode them into the tool, or they may fail to anticipate how others could misuse

their tool. Similarly, tool creators may only have one target population in mind, resulting in a tool that is not equally effective for people from different gender, race, ethnic, cultural, economic or disability groups.

How bias can enter algorithms

In addition to the overall purpose and design of a tool, several avenues can introduce bias into AI algorithms.

Non-representative training data

Problems can arise in AI systems when the training data used to create them are not representative of the full population of potential users. For instance, Amazon famously shut down an AI hiring tool because it favoured men, partly because a large majority of Amazon's past employees, used for the training data, were men.⁶ This problem can also affect non-AI systems as well, for instance human judgements may be based

upon past experiences dominated by a particular social group.

Non-representative training data is a common issue with AI systems and therefore it is important to scrutinize the training data. On the other hand, well-represented training data may offer an opportunity to improve decisions over individuals who base their judgements on a narrow set of past experiences.

“ These imperfect measures are problematic for an organization whether they are used by an algorithm or a human. The key takeaway is that an AI algorithm will not fix these problems.

Imperfect measures

Sometimes measures do not perfectly capture underlying intentions. In the HR context, a number of abstract concepts may not be measured perfectly, including key measures such as performance. Most organizations recognize that performance measures do not capture all dimensions of actual performance. These measures may also be subjective and

therefore incorporate human biases. Imperfect measures will be particularly influential if they are used as the outcome because the tool is designed to maximize that outcome. These imperfect measures are problematic for an organization whether they are used by an algorithm or a human. The key takeaway is that an AI algorithm will not fix these problems.

Real-world biases in the data

The training data can also reflect real-world biases. Many of these biases come from society at large; for instance, social groups will likely differ in their levels of educational attainment and work histories. Even a seemingly objective measure such as who leaves the company could

be the result of a hostile work environment. Again, these are not issues unique to AI systems. To minimize bias in the algorithm, though, it is necessary to examine each factor included in the training data and consider both its relevance to the task and its potential to incorporate bias.

The importance of diverse perspectives

While not a guaranteed fix, involving a diversity of individuals in the design of an AI for HR tool increases the likelihood that potential problems of bias will be identified and addressed.

Your organization should also draw on a diversity of perspectives and voices when conducting its own assessment AI-based tools. If you do not have the internal capacity to do such an assessment, consider external sources to provide these perspectives.

Measuring and mitigating bias

What is fair?

In order to measure bias, it is first necessary to decide what is fair, which can be hard to do whether machine or human decisions are concerned. There are many ways to define fairness; in most cases it is impossible to satisfy all types of fairness. In the HR context, there is a general consensus, spanning multiple countries, on one definition: different social groups

should have the same or similar average outcomes from an algorithm. For instance, in the example of hiring and promotions, applicants from different social groups should have similar rates of selection. If one group has much lower rates of selection than others, this is what is often called adverse impact, disparate impact or indirect discrimination.



The “4/5ths” rule in the USA. In the United States, several federal agencies use the “4/5ths rule” (as well as tests of statistical significance) for initial assessment of discrimination in employment decisions. Under this rule, if a protected group is selected at a rate that is less than 80% of any other group, that gap is considered initial evidence of adverse impact.

Many creators of AI-based HR tools that serve the US market, therefore, will seek to ensure compliance with this rule and may use it as their primary definition of fairness. Satisfying the 4/5ths rule does not guarantee protection from legal claims of discrimination in the US, but tools that do not satisfy the 4/5ths rule may face scrutiny and require additional justification.

Which social groups?

Measuring bias also requires identifying which social groups to assess. Many countries have legally specified protected groups; organizations may want to consider additional groups of concern and underrepresented groups beyond those specified by law.

The treatment of individuals with disabilities can pose particular challenges to AI algorithms. Addressing these issues can be important to ensuring that people with disabilities are treated fairly as this also often results in systems that work better for all individuals.

Reducing bias by omitting variables

Unfortunately, it is not possible to solve the problem of bias by simply not including data on social group memberships, such as gender and race. It would also be necessary to leave out any other measures that are correlated with gender and race (called

proxy variables). In some cases, removing these variables can reduce the accuracy of the algorithm. However, the accuracy gained by these variables may be accuracy that you don’t want to use if they primarily capture the biased aspects of the task.

Testing and monitoring outcomes

There is no guaranteed, easy fix to eliminating bias in AI algorithms. AI-based HR tools, therefore, should have established systems that test for

potential biases both before deployment and as part of ongoing monitoring.

Limits and opportunities in the debiasing process

Machine learning takes the conditions represented in the training data as a given, predicting outcomes with the assumption that these conditions will persist (e.g. which job candidates are likely to perform well under the conditions represented in the training data). AI algorithms cannot consider ways in which the context itself might be changed.

Organizations can go further by using the process of debiasing an AI algorithm as an opportunity for broader organizational improvement, working outside of the AI tool to change the processes that create biased data in the first place. Ask the AI algorithm creators about the bias mitigation steps that were needed and which measures posed particular

“ There is no guaranteed, easy fix to eliminating bias in AI algorithms. AI-based HR tools, therefore, should have established systems that test for potential biases

problems because they were correlated with social groups. Use this information to inform and change your own organizational practices. For a more detailed

discussion of this process, read “[Using Algorithms to Understand the Biases in Your Organization](#)” by Jennifer M. Logg in *Harvard Business Review*.

[Tool Assessment Checklist](#)

[Planning Checklist](#)

3.2 Data privacy and security

Key takeaways

- AI requires both training data and data to be used as inputs once deployed, raising both reputational and legal risks for privacy and security.
- Organizations should carefully consider the data being collected, the possible implications of its use, and what controls are necessary to ensure privacy and security.
- The European Union’s General Data Protection Regulation (GDPR) regulates the use of data in these countries and is increasingly used as the reference for data protection and regulation around the world.

AI’s heavy reliance on data means that it is necessary to balance the potential benefits of a tool with the impact of collecting and using such data, especially data considered private, sensitive or ethically questionable.

The use of data by AI-based tools raises two risks. The first is **reputational risk**. AI-based tools that appear to violate privacy, have a “creepy” or “stalking” element, or that result in data breaches risk permanent damage to the trust of your employees and serious loss of reputation. In addition, the use of data poses considerable

legal risks. Organizations operating in the European Union need to pay particular attention to its **General Data Protection Regulation (GDPR)**. Even organizations operating outside of the EU should look closely at the GDPR because many jurisdictions are in the process of adopting legislation modelled after it. The GDPR also requires organizations to conduct a [Data Privacy Impact Assessment](#) (DPIA) for high-risk activities (which will include many HR activities). Even outside of the EU, completing a DPIA can be useful in assessing data concerns and mitigation strategies.

Awareness

To mitigate reputational and legal risks, organizations should notify data subjects and their representatives of the collection, sharing and use of personal data

ahead of time. The level of detail provided should be guided by legal requirements and provide adequate levels of transparency and explainability.

The challenge of consent

The GDPR requires businesses to obtain consent from data subjects or to specify another legal basis for the processing of data. Given the power imbalance between employers and data subjects in the HR context, however, it can be challenging to gain true consent because the subjects may fear

negative consequences. Organizations should still seek to gain such consent and should reconsider uses where data subjects would be unlikely to consent; they should also be prepared to further justify how their use of the data constitutes a legitimate interest.

Data minimization

Look to collect the minimum data that is required and store this data only for as long as is necessary. Also consider aggregating or anonymizing the data

when possible. In addition, organizations should carefully consider and minimize who has access to personal data.

Data security

Organizations should implement appropriate technical and organizational measures to ensure data security, especially for highly sensitive data. Steps include:

- **Anonymization** of data by removing information that can identify individuals from the data or **pseudo-anonymization**, which involves removing identifying information from the main dataset and storing it separately, with a coding system that allows an authorized person to link back the data if necessary.
- **Data encryption and protection.**
- **Permissions.** Who has access to and the ability to alter data?

- **Sharing and storage of data, third parties.** Will the vendor have access to personally identifiable data? How can the vendor use the organization's data and how will it ensure privacy and security?
- **Data transfer.** Pay particular attention to cross-border data transfers and their legal ramifications.
- **Data retention.** How and for how long will the data be stored?
- **End of life and erasure.** What is the plan for moving older data out of the system? When and how will data be erased when no longer used?

[Tool Assessment Checklist](#)

[Planning Checklist](#)

3.3 Transparency and explainability

Key takeaways

- Transparency and explainability are critical to maintaining trust in AI systems and ensuring the effective and responsible use of an AI-based tool.
- Consider three different groups: the organization, the users of the tool, and the people the tool impacts.
- These three groups should have information about the overall design of the tool, the key factors driving the decisions, and explanations for individual decisions.

Transparency and explainability play a critical role in the successful adoption and implementation of AI-based systems. Without them, the risks are misused tools, poor decisions, the undermining of trust, and reputational and legal consequences.

There should be **transparency** between all parties involved in the use of an AI for HR tool. Transparency refers to the general principle that all parties involved in the use of an AI for HR tool should be **aware that an AI algorithm is being used and be provided with information about the data collected, how the tool works, and how the output will be used.** This is a legal

requirement for organizations in Europe under the GDPR. Also consider whether and how individuals will have the **ability to opt out** of the use of the AI-based tool and what alternatives will be provided for those individuals to ensure that they are not punished for opting out.

A closely related topic is **explainability** – the need for an AI system to provide an explanation for its decision that is comprehensible to humans. There is tension here because the real world is complicated and so a complex AI system may sometimes be necessary to perform well; but humans are very uncomfortable with an algorithm making

“ The level of explainability necessary for an AI for HR tool may vary. In situations where the consequences of the wrong decision are high, explanations will be more important.

decisions that they cannot understand. In addition, complex AI systems can sometimes make “weird” decisions from a human perspective, potentially undermining trust. At the same time, the same level of explainability is often not demanded for human decisions. Humans are often allowed to make decisions without fully documenting their reasons and there is little guarantee that any explanation that they do provide is in fact the true explanation.

The level of explainability necessary for an AI for HR tool may vary. In situations where the consequences of the wrong decision are high, explanations will be more important. In other cases, such as a chatbot that schedules interviews, it may be less important to understand the inner workings of the algorithm.

Multiple audiences

There are several levels of audiences to consider in transparency and explainability.

The organization. Organizations are adopting tools that will affect their operations and could likely be held liable for the consequences of their decisions; therefore it is critical to have a clear idea of how the system works. Organizations should be wary of vendors who do not wish to provide details about their systems.

Current efforts to ensure AI explainability take two different approaches. One approach is to favour simpler algorithms where it is easier to identify the factors driving a decision. The second approach is to gain insight into a complex AI algorithm by identifying the most influential factors that drive the system. This approach will not provide a full explanation for a complex algorithm but will identify the key aspects of the system.

Finally, **third-party auditing is becoming increasingly common** and can play a valuable role in ensuring the veracity of the claims made by an AI vendor. It is important to find out details, though, about what they actually audited, since many different types of audits are possible (e.g. accuracy of the system, fairness outcomes, data handling procedures, etc.).

Tool users and their managers. The individuals who will be using the systems day-to-day should receive clear information about how the system works, explanations for decisions, and guidance on how to use the output.

Those impacted by the system (employees/applicants). Most AI for HR tools make decisions about employees or job applicants. These individuals should be aware that the organization is using an AI-based system and should receive information about how it works.

What information should you share

Information about the AI-based tool should occur at several levels.

Overall process and basic design of the algorithm

- The fact that an AI-based system is being used
- How the tool will be used
- Whether and how a human will be involved in the decision/process
- Appropriate and inappropriate uses of the tool
- The context in which it was designed to work (and possibly inappropriate contexts)
- The source of the training data
- What inputs are used, what data will be collected
- What the algorithm is trying to predict (outcome)

“Global” explanation of algorithm. This is a description of the trained algorithm, what patterns it has learned from the training

data. Often this involves a report on which factors (inputs) are the most influential.

“Local” explanation of individual decisions.

An explanation for each decision that the algorithm makes once deployed. A local explanation of a job candidate screening tool, for instance, would provide information about what characteristics of a given candidate led them to be recommended or what the candidate was missing that led that individual to not be recommended.

A few features of local explanations may further improve how the results are used. First, when possible the system should indicate a level of confidence in a result. Certain individuals, particularly those quite different from the training data, may result in higher levels of uncertainty. This is valuable information for the algorithm to report. Similarly, be cautious with reports that provide precise scores for each individual, which might lead the user to make decisions based upon small differences in scores that may not actually be meaningful.

The creation of local explanation reports also raises the possibility of providing such reports to the

individuals affected. A job candidate, for instance, could be given information on what skills they were missing and therefore why they did not make the cut. Even if this information is not shared automatically

with employees or candidates, individuals should be able to request this information (this is legally required under the GDPR). Such information will also be important in documenting the organization's practices.

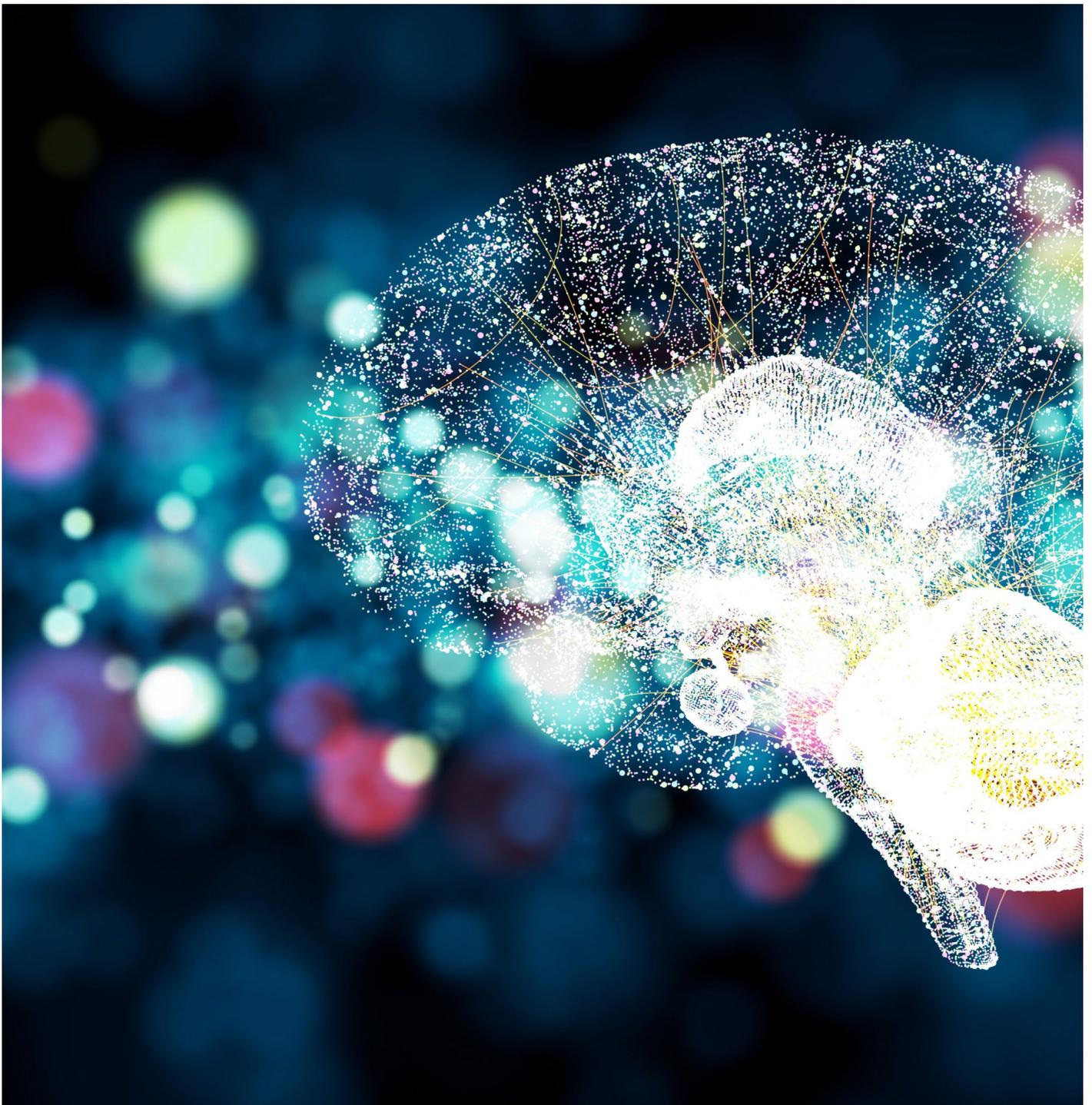


Illinois Artificial Intelligence Video Act. In 2020, the state of Illinois in the United States [passed an act requiring transparency and explainability when AI is used to analyse job interview videos](#). The law requires employers to notify applicants that the

system is being used, provide an explanation for how the system works, obtain consent, maintain confidentiality, and destroy the data if requested by the applicant.

[Tool Assessment Checklist](#)

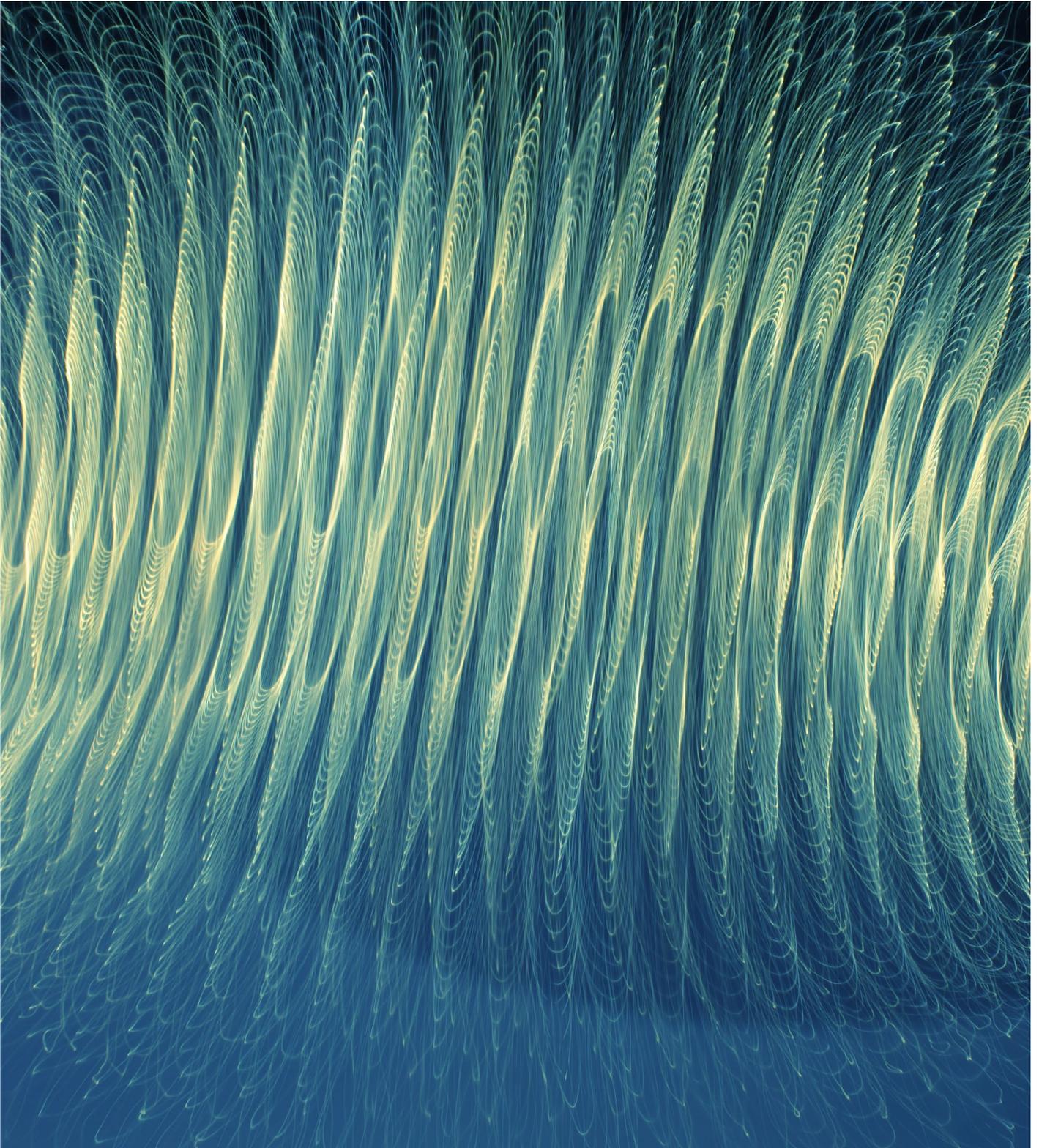
[Planning Checklist](#)



4

Implementation and buy-in

Planning for the successful adoption and
use of AI-based HR tools



Key takeaways

- Much like any organizational change, the successful adoption of AI-based HR tools requires considerable planning, training and communication.
- The human users of an AI-based tool will be most effective if they have a clear understanding of its strengths and

weaknesses, as well as guidance on how to incorporate the results into their decisions.

- Widespread trust in and buy-in to use the tool are also crucial.
- Consider how to best roll out its adoption and assess its impact.

One of the biggest mistakes that organizations make in adopting AI-based systems is underestimating the amount of communication, education and training that is necessary to ensure success.

It is necessary to dedicate particular thought and time to **providing education and training to tool users and to specifying how to incorporate the output of the tool into their decisions**. While it is often said that it is better for a human to make the final decision and not a machine, without proper guidance there is a risk of one of two extremes: people who overestimate the power of AI and always accept the recommendation of the algorithm without question; or people who are distrustful of the system and ignore its recommendations. The users will be much more effective if they understand how the system works, its strengths and its limitations, and have access to the global and local explanations described in the section on Transparency and explainability. The organization should also **provide specific guidelines** on how users should use the system, **highlight areas where human input is particularly important, and create procedures to document the human aspect of the decision** as well as the algorithm.

The use of an AI-based HR tool has the potential to either improve or undermine trust within an organization. Poorly implemented AI tools in HR

may raise concerns about violating privacy or “taking the human out of human resources”, while thoughtfully adopted and implemented tools will improve outcomes and the employee experience. To avoid undermining trust, **employees should be involved in the decision-making process from the start** so that the adoption of the tool reflects their needs and concerns and is not perceived as something that is imposed on them. In addition, successful implementation requires **a strong internal communication and consultation plan with employees and other affected individuals** (e.g. job applicants) that provides clear information on the reasons for adopting the tool and its benefits and an opportunity for feedback. It is also necessary to continue to communicate about the tool's impacts after deployment to build further trust in the tool and for future projects.

Finally, **plan how you will roll out the use of the tool and assess its impact**. You may want to start with a pilot, which will help you refine processes, learn key lessons and create tangible evidence of why the tool is useful. Consider running the pilot as a randomized experiment, randomly selecting one or more test and control groups to concretely measure its effects. Have a plan for **measuring the impact of the tool**, tracking the key outcomes and benefits that you expect from the tool.



Going against the algorithm's recommendations can sometimes be valuable. Machine learning systems work best when given a variety of examples in the training data, allowing it to consider a wide range of possibilities.⁷ This could be a reason to pool training data from other organizations. An additional challenge is that once a tool is deployed, the deployment itself may limit the variety of examples that might be used to update the algorithm in the future: if you always follow the algorithm's recommendations, you won't have examples of what happens when you don't follow its recommendations to ensure that you are not missing an even better option. One way to address this problem in low-stakes situations is to introduce occasional random choices. For instance, in a system recommending training programmes, occasionally include a

randomly selected programme as part of the set of recommendations. If people consistently choose the random recommendation over the ones predicted by the algorithm, this information can be used to update and improve the algorithm. In high-stakes situations where random decisions are less feasible, an alternative is to allow decisions that do not follow the algorithm, especially if there is reason to believe that the outcome might be positive. Such an approach encourages the user to continue to look for new possibilities rather than blindly following the algorithm. It is necessary to develop a system to track and assess such experimentation, both to ensure that it does not simply become an opportunity for users to return to unjustified gut decisions and to truly use this experimentation to learn and improve both the algorithm and the organization.

[Tool Assessment Checklist](#)

[Planning Checklist](#)

5

Ongoing maintenance and monitoring

AI in HR requires ongoing monitoring, with special attention paid to changes in context.



Key takeaways

- The creator of an AI-based HR tool should provide ongoing reports, monitoring key outcomes and concerns.
- Organizations should develop a plan for monitoring the performance of the tool, how it is being used, and its impact on the organization.
- Organizations should monitor the context within which the tool is deployed for changes that might threaten its performance.

The AI tool's creator or vendor should have established systems for documenting and monitoring its performance. You will also need to monitor changes in performance either due to updates to the algorithm or changes in the deployment context.

In addition, your organization should develop its own monitoring plan. This plan should involve ongoing assessment of the tool as well as monitoring the organizational practices and impacts.

Monitoring the algorithm

Algorithms should be subjected to ongoing monitoring. A monitoring plan should be agreed upon with the vendor before implementation, with responsibilities clearly laid out. The monitoring should include:

- Compliance with necessary regulations (e.g. GDPR)
- Ongoing assessment for bias and fairness
- The tool's accuracy, whether it is successful in predicting the eventual outcomes of the cases it has assessed
- Changes in the training data and algorithm, when and how it is updated, as well as tracking the age of the training data and considering the possible retirement of older data
- Ensuring that the system is subjected to periodic security penetration tests.

Organizational monitoring plan

Before deploying an AI-based HR tool develop a clear plan for monitoring the organization's use of the tool and its outcomes.

Specify **how you expect the tool to improve organizational and individual outcomes** and how you will measure and monitor that impact. Document the baseline values for those outcomes so that you can measure the change later.

Create a system for regularly **auditing a sample of decisions**. This audit should include **the inputs and**

outputs of the algorithm and the actions of the user and final decision-maker to ensure that they are using the tool appropriately and as expected.

Assess the **impacts on and sentiments of the users and subjects** of the system to identify and address concerns. Track how **user and subject behaviour may change as they use the tool**, paying attention to possible misuses or gaming of the system.

Monitoring the context

AI-based tools are designed to function in contexts that match well with their training data. If the context changes significantly, the AI-based tool is at risk of failure. The COVID-19 pandemic, for instance, posed challenges to many AI systems⁸ but not necessarily all

of them. The key question is whether the patterns that the algorithm has identified in the training data would still be applicable in the new context. In addition to major events, it is necessary to monitor for changes in the local context where the tool is deployed.

[Tool Assessment Checklist](#)

[Planning Checklist](#)

Tool Assessment Checklist

This document provides a list of questions to answer when considering adopting a specific artificial intelligence (AI)-based human resources (HR) tool. The questions are tied to each section of the guide,

which provides a background for why each of these questions is important. The questions aim to assess both the tool and your own organization's goals in adopting the tool and your implementation plans.

The many uses of AI in HR

This tool aims to help in which area of HR?

What task(s) does it focus on?

How is this tool promising to help?

Automate tasks:

Maximizing/predicting an outcome:

Bringing in new information:

Enabling a new task:

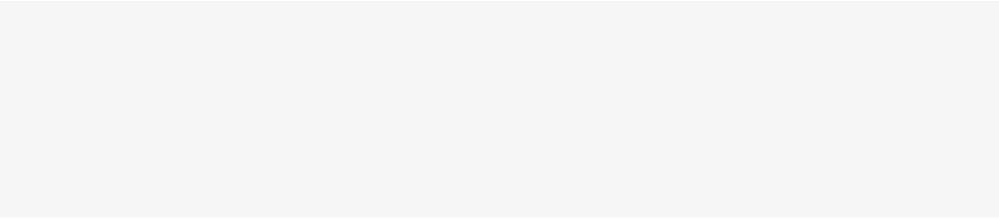
Promising to help in a specific goal

[Return to section](#)

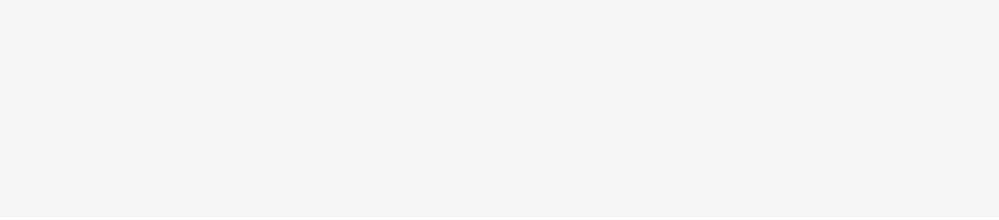
Forming an assessment team and planning for the long term

Which departments/individuals in your organization should be brought in on this decision?

What types of external expertise will be necessary and where will you find them?



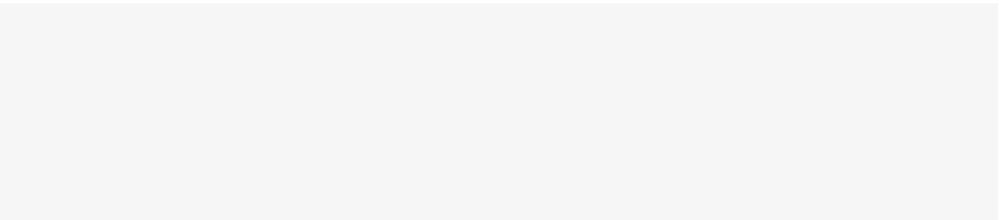
Which decision-makers and individuals at the top of the organization need to be informed or involved? When?



Who will be the project manager?



Scan the organization for existing governance structures that need to be followed



[Return to section](#)

What is the purpose of adopting the AI-based tool?

What problem do you hope to solve by adopting this tool?

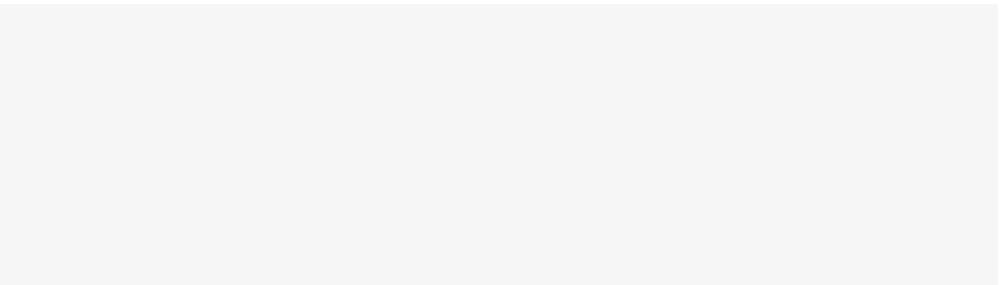
How is the task currently done and what are the problems?

How do you hope the AI-based tool will improve the process?

How will it change processes?

Will you be able to use the results effectively?

Notes on these points:



What organizational and employee outcomes do you hope that the tool will improve?

Why are these outcomes important/valuable?

Will you be able to assess/document this improvement?

Notes on these points:

How does this tool fit into the organization's AI journey?

What is the organization's level of experience and knowledge of AI?

What is the current state of the data infrastructure?

Is there buy-in from top leadership?

What is the attitude of employees about technology and AI?

Given the answers to these questions, how will the adoption of this tool fit into and possibly advance this AI journey?

Notes on these points:

[Return to section](#)

What are the core elements of the tool?

What are the main steps in the process from start to finish?

Where is artificial intelligence (AI)/machine learning (ML) used? Only a subset of steps in most tools will use AI/ML. Sometimes AI/ML will be used just one place; in other cases multiple AI/ML algorithms will be used at different steps.

For each place where AI/ML is used, consider:

Details about the training data.

What is the source of the training data? Is it relevant to your context?

*What is the time horizon of the training data? How quickly will it become obsolete?
Will new data be added and old data retired?*

Will the system pool training data from other job types or from other organizations?

Notes on these points:

What inputs does the algorithm consider? Do these inputs make sense?

What inputs might be relevant to the task but are not included? (see Implementation on communicating to users the limitations of an AI-based tool and establishing processes for combining machine and human decisions)

Is the algorithm predicting/maximizing an outcome?

Does this outcome make sense?

How is it measured?

What does this outcome measure capture, what might it miss?

Notes on these points:

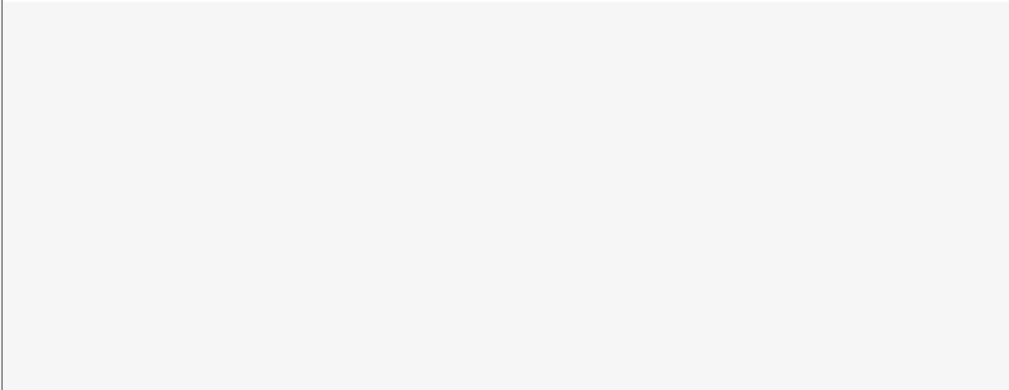
What data will the system require from your organization (either as training data or once deployed)?

Consider the availability and quality of the data in your organization. What work will be necessary to clean and prepare the data? Does the vendor provide help for this?

Are there consequences (e.g. legal or reputational) to collecting and using this information?

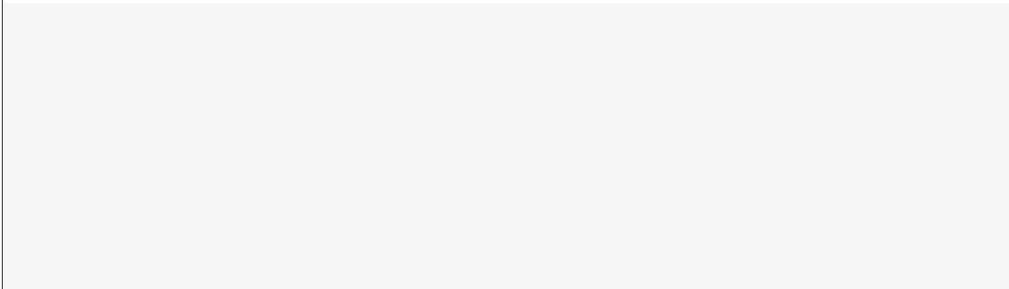
Will the tool integrate into your existing HR information systems?

Notes on these points:



How will the system be updated? Some ML systems are trained initially and then only occasionally updated. Other systems are designed to have new training data fed into the system for constant updating.

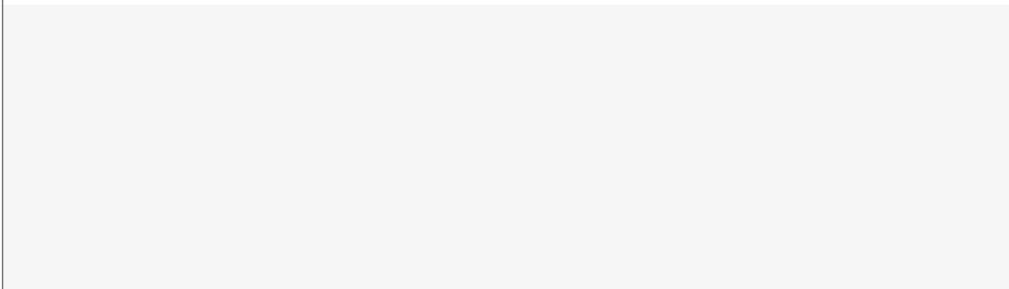
Algorithm accuracy/validity.* How have the developers tested the system to ensure that it is effective and accurate? What is its accuracy?



**Beware: Accuracy measures for predicting an event or a yes/no question can be misleading if most people fall into one category. For instance, consider a turnover algorithm predicting who will leave in the next year. If 85% of employees on average stay, I can achieve 85% accuracy by just predicting that that everyone will stay! A useful tool would need to have even higher accuracy.*

Has a third party audited the tool?

If they have, what did the audit assess?



How are users going to interact with the system?

What does the user interface look like?

What training will be necessary for users and does the vendor provide this training?

Review these components and discuss:

Whether the tool as a whole seems reasonable and well-designed

What the strengths and limitations of the system are

What the strengths and limitations of how you currently do this task are

Notes on these points:

[Return to section](#)

Assessing the risk level of a tool

Consider the following factors in assessing the risk level of the tool. See the following page for a risk scoresheet.

- **How the output will be used.** Example uses, ranging from lower to higher risk:
 - Provide information (e.g. number of vacation days remaining)
 - Suggest options for the user to act on voluntarily (e.g. training recommendations, career suggestions, etc.)
 - Inform/influence decision-maker
 - Make the decision
- **Objective versus subjective tasks.** Tools that primarily deal with objective information are less likely to have major consequences. Tools that make decisions that involve more subjective data bring higher risks.

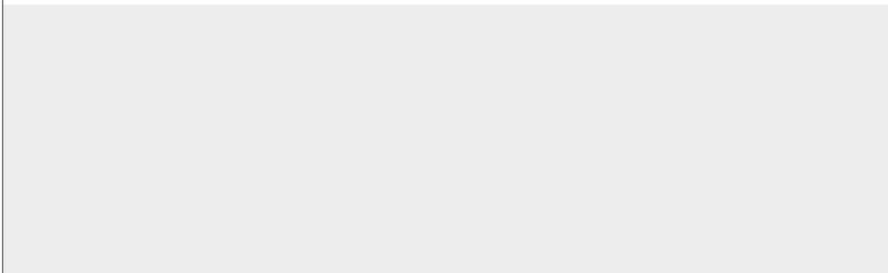
- **Consequences for affected individuals.** What would be the consequences for individuals (employees, job candidates, etc) if the algorithm gets it wrong? Tools that may have significant impacts on people's lives, for instance affecting their economic situation, career outcomes, health, or happiness will require particular attention and scrutiny. Also consider the variability of the consequences. Will people benefit equally or will some people benefit a lot while others not at all or negatively?
- **How central will the tool be to the organization?** Would the tool affect core aspects of how the organization operates or have potentially large operational or financial consequences? What are the organizational impacts if the tool fails to perform or has unintended consequences?
- **Societal impacts.** Does this tool risk exacerbating societal problems or inequalities? Will it impact local communities, the environment, etc?
- **Scale of the consequences.** How many people will be directly impacted?
- **Does the tool use sensitive data?** What information does the tool require and is this data sensitive to individuals and/or the organization?
- **Legal risks.** Is the tool operating in an area that is regulated? What are the relevant laws and regulations and how do they vary across the locations where the tool will be used? How does the vendor ensure compliance and what assurances do they provide?
- **Organizational culture and employee trust.** Might the use of the tool, or its potential misuses or failure, risk disrupting trust in the organization or other aspects of the organizational culture?
 - Does the tool make decisions or use data in a way that employees might be displeased to learn? Will the use seem creepy or violate trust?
 - Does it risk a sense of losing of human touch or empathy?
 - How might it change employee incentives or behaviours?
 - Is the tool intended to automate and replace workers?
- **Reputational impacts.** Might the use of the tool, or its potential misuses or failure, hurt the reputation of the organization?

Ability to mitigate. Consider the safeguards already put into place by the tool's creator as well as additional safeguards that your organization could put into place to mitigate these risks. Consider the time and resources required for this mitigation.

Consider various scenarios for failure and how you would respond.

What is the overall risk level of this tool?

What does this risk assessment indicate for the level of scrutiny that is necessary in the adoption, implementation and monitoring stages?



| Risk level of tool | | After mitigation | |
|--------------------------------------|------|------------------|------|
| Low | High | Low | High |
| How the output will be used | | | |
| | | | |
| Objective versus subjective tasks | | | |
| | | | |
| Consequences for affected individual | | | |
| | | | |
| How central to the organization | | | |
| | | | |
| Societal impacts | | | |
| | | | |
| Scale of the consequences | | | |
| | | | |
| Use of sensitive data | | | |
| | | | |
| Legal risks | | | |
| | | | |
| Organizational culture and trust | | | |
| | | | |
| Reputational risks | | | |
| | | | |
| Other risks | | | |
| | | | |
| Summary assessment | | | |
| | | | |

[Return to section](#)

Bias

Does the tool make claims about reducing bias in the task and/or increasing diversity?

How does the tool propose to reduce bias/increase diversity?

By using an “objective” algorithm rather than a human. Remember that AI systems are trained on real-world data and therefore are not automatically objective. An AI system might reduce bias by creating more uniform processes or by removing from consideration factors such as whether an applicant is a friend. However, examine closely the data the model uses and consider how biases likely also remain in the data.

Focusing on a specific outcome. If there is a well-defined outcome, the algorithm has the potential to reduce biases if humans currently doing the task are not paying much attention to that outcome or they are overlooking key patterns that predict the outcome.

Changing how the task is done. Consider how the AI-based tool is proposing to do the task differently. How would it reduce bias?

What evidence do they have that the tool does reduce bias/increase diversity?

Does the tool make assumptions that could perpetuate or encode biases? Could it be misused?

What is the source of the training data?

Does it match the context and population where it will be deployed?

Is it representative? Does it contain enough records for different social groups to be accurate?

Consider the inputs and outcome used in the algorithm, both in the training data and once deployed

How well are they measured? What do they capture and what do they miss? How might they encode biases?

Consult with individuals from diverse backgrounds to help anticipate problems with the tool and its use

What are your organization's priorities and policies regarding bias, diversity and discrimination?

Are there particular protected groups that need to be monitored?

How would your organization define fairness in this context?

Is it representative? Does it contain enough records for different social groups to be accurate?

Ask the vendor/creator how they conceive of fairness and bias

How does the tool ensure compliance with applicable laws?

Ask the vendor/creator how they will test for and mitigate bias, both before and after deployment

If they do need to mitigate bias, how do they go about doing this?

What information will they provide the organization about the mitigation process?

[Return to section](#)

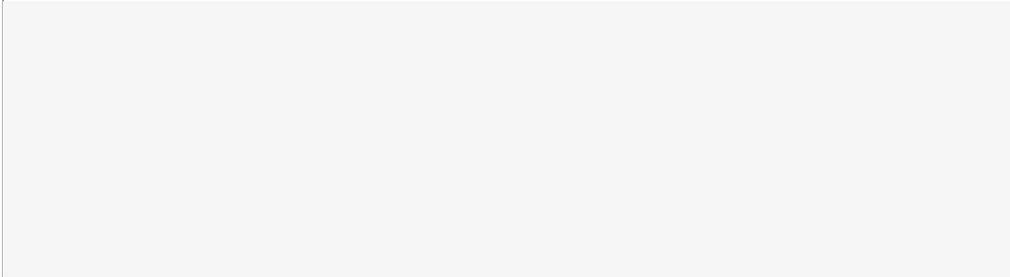
Data privacy

What types of data will the tool use?

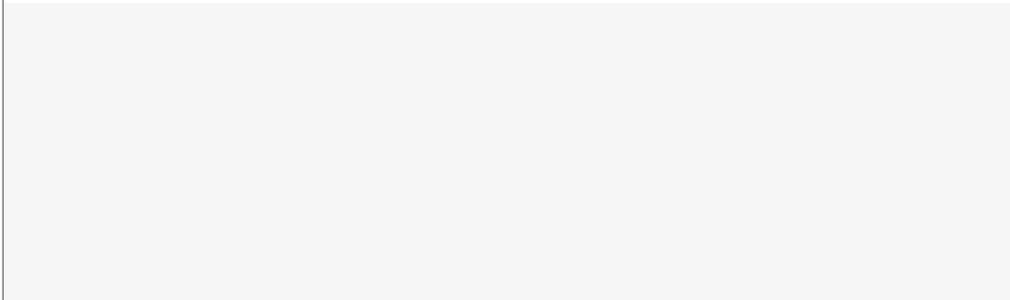
Does it meet the criteria of data minimization (only using data that is necessary)?



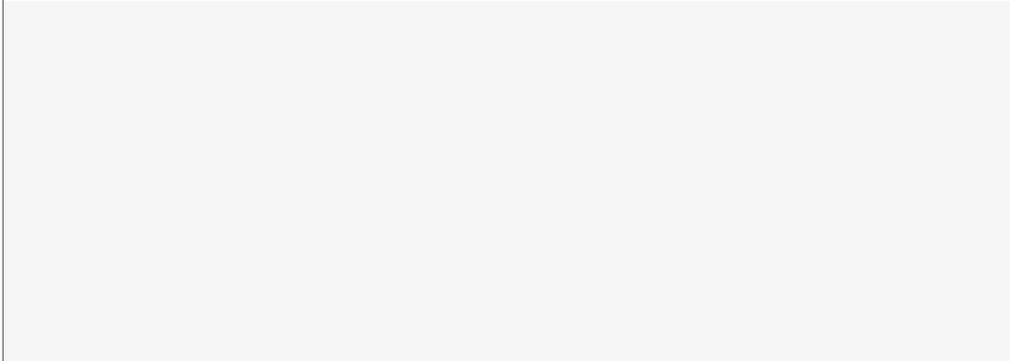
How might the collection or use of that data raise reputational risks or undermine trust in the organization?



Will the subjects be aware that the data is being collected and of how it will be used? Can they give meaningful consent?



Will the subjects be aware that the data is being collected and of how it will be used? Can they give meaningful consent?



Find out details about the following:

Anonymization or pseudo-anonymization of sensitive data

Encryption

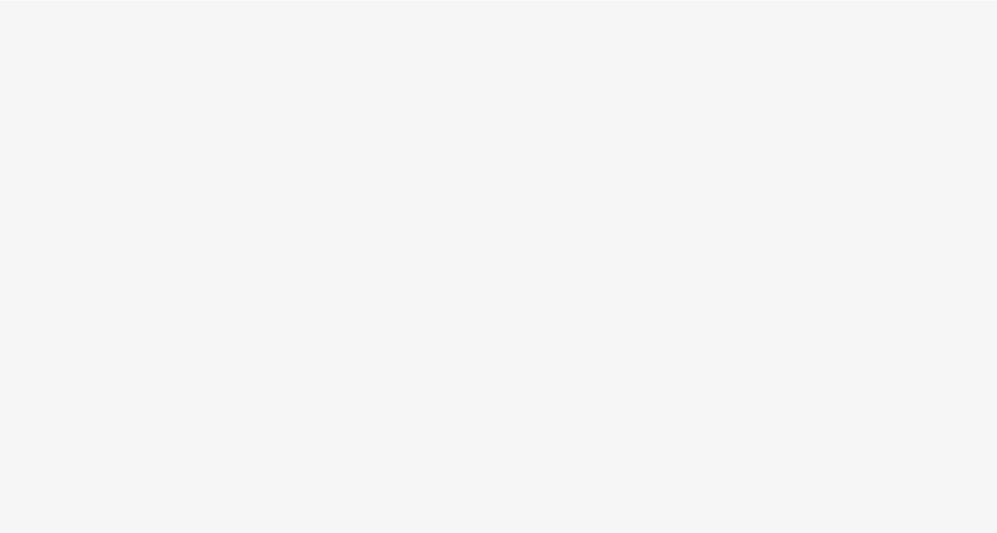
Who has permission to access the data both within the organization and at the vendor

Will the vendor have access to the data? How can the vendor use the data?

Will the data be transferred? How? Will there be cross-border transfers?

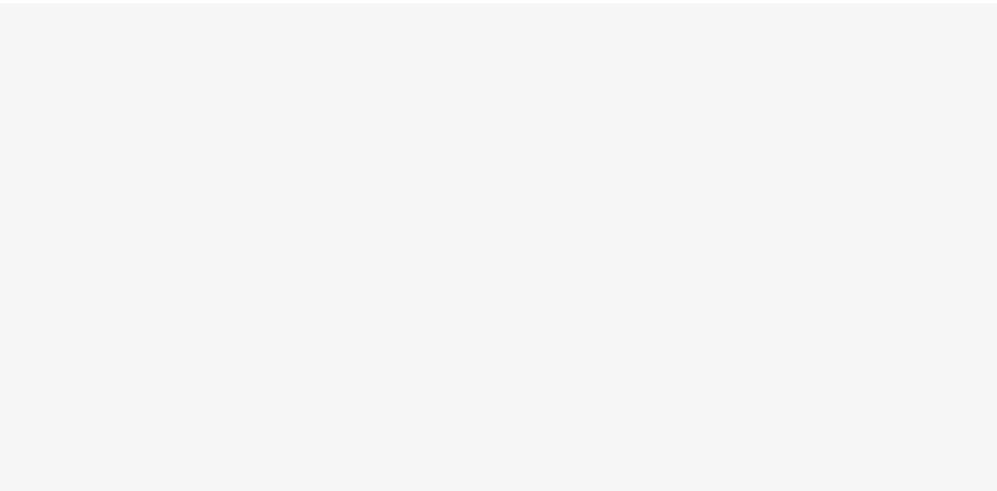
How and for how long will the data be stored?

What is the plan for data that is no longer used or relevant?



What are the data privacy laws in the jurisdictions where you operate? Does the vendor comply with these laws?

What processes will you need to put in place to ensure ongoing compliance with relevant laws, especially General Data Protection Regulation (GDPR) if in Europe?



Consider creating a [Data Protection Impact Assessment](#)

[Return to section](#)

Transparency and explainability

How will you provide transparency to those impacted by the tool (employees, applicants, etc.)?

How will they be informed about the collection of data and the use of the AI-based tool?

Will you provide a meaningful opportunity to opt out?

Has a third party audited the tool?

If they have, what did the audit assess?

What was your determination of the risk level of the tool and what does this mean for the level of explainability that is necessary?

Develop plans for sharing information about the overall design of the tool

Your organization should determine the following in collaboration with the creator

How the tool will be used

Whether and how a human will be involved in the process

Appropriate and inappropriate uses of the tool

The context in which it was designed to work (and possibly inappropriate contexts)

Some aspects of the overall design you should have already collected in previous sections, including:

Source of the training data

What information is collected, what inputs are used

The outcome, what the algorithm is trying to predict

Develop a system for conveying this information or review how this information is already conveyed within the tool, to:

The user of the tool (HR team members or employees)

Those impacted by the system (employees or applicants)

Global explanation (how the algorithm itself works)

Will the creator/vendor generate a global explanation specific to your training data and deployment context?

What does that global explanation report look like? Is it easy to read and understand?

Develop a plan for conveying these global explanations to the users of the tool

Decide what aspects of the global explanation should be shared with those impacted by the tool

Local explanation (individual decisions)

Will the tool create a local explanation for each case?

What does that local explanation look like?

Is it easy to read and understand?

Is it provided automatically to the user?

Will you share local explanations with the affected individual?

[Return to section](#)

Implementation and buy-in

Determine how the AI-based tool should be used in the organization and how it should be combined with human processes and judgements

What are the strengths of the AI-based tool? What are its weaknesses or aspects of the task that it overlooks or does not address?

Are there areas where human input will be particularly important?

Consider developing systems to document the human side of decisions

Create training materials and a training plan for users.

Involve employee representatives in the selection, adoption and implementation of AI-based HR tools to gain their perspective and increase the chances of a positive response to their use

Develop a communication plan to:

Explain why a tool is being adopted and its expected benefits

Gather feedback from employees or impacted individuals

Continue communication after deployment

Going against the algorithm's recommendations

Is it feasible to occasionally include random choices to test the algorithm?

How might you encourage and document the results of cases where you do not follow the algorithm?

[Return to section](#)

Ongoing maintenance and monitoring

What reports will the creator/vendor provide for ongoing monitoring?

What information is provided in the reports? Does it cover the necessary information?

Are the reports easy to understand and user-friendly?

How often will the algorithm and/or training data be updated?

For systems that are periodically updated, pay particular attention to any changes in performance with each update

For systems that are continuously fed new training data, monitor the system for drift over time

Develop a plan to monitor the impact of the tool, focusing on the key outcomes you expect it to influence

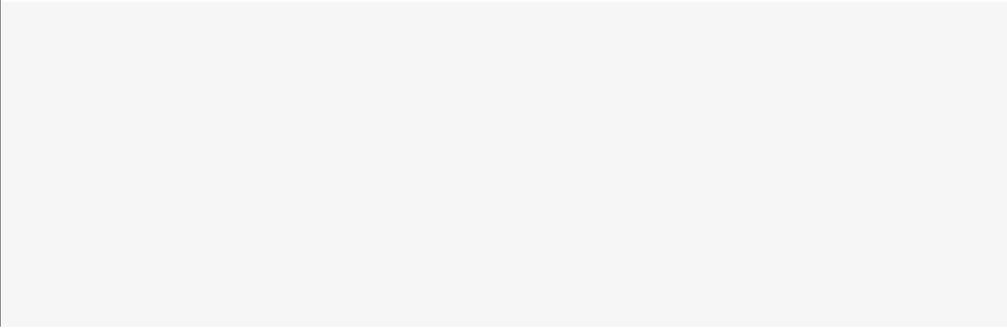
Measure baseline values

Develop a system to assess the outcome and use of the tool, for instance by examining a random selection of decisions, including:

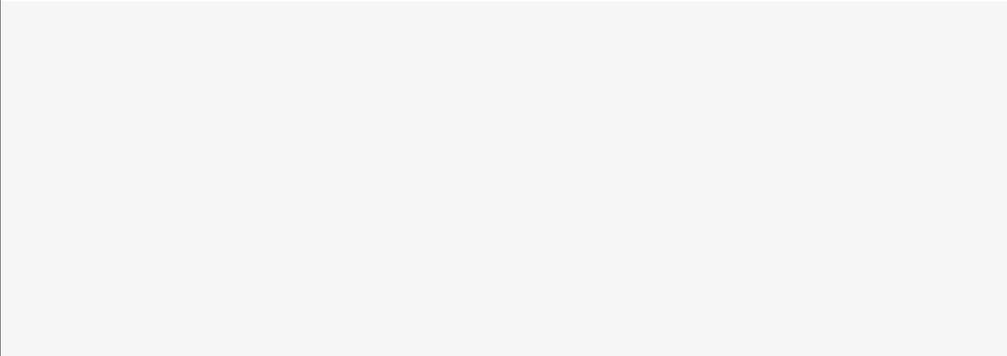
The output of the algorithm, both its assessment as well as the "local explanation" if provided

The user's and/or decision-maker's actions, whether they used the tool appropriately and whether they are under- or over-relying on the output of the tool

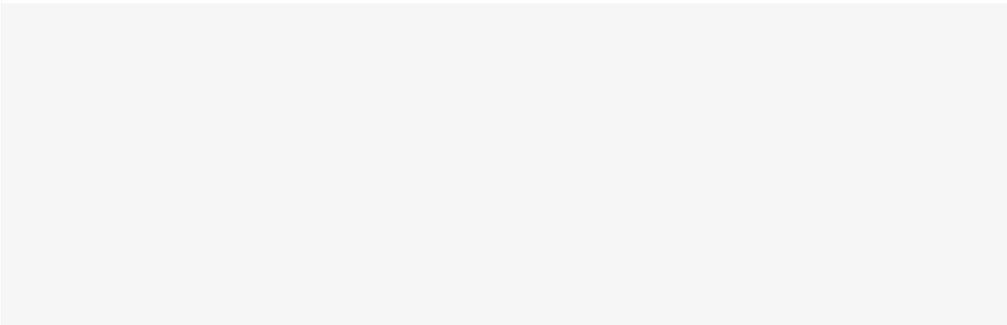
Develop a plan for monitoring users' and subjects' attitudes toward the tool



Periodically assess whether and how users and/or subjects are changing their behaviours, including unexpected uses and well as attempts to game the tool



Monitor the context, looking for changes that might mean that the patterns in the training data may no longer be relevant



[Return to section](#)

Planning Checklist

This document is designed for organizations looking to plan strategically for the use of AI in HR and develop organizational capacity to use AI in HR

responsibly. This checklist accompanies the guide, providing a list of questions to discuss based upon the topics covered in each section of the guide.

The many uses of AI in HR

What type of AI for HR tool would be of the most interest for your organization?

What HR tasks or parts of the HR lifecycle might be higher priority?

Is your priority to automate or augment processes? What types of augmentation?

Are there specific goals such as diversity and inclusion that are a priority for the organization?

Notes on these points:

Scanning for AI-based tools that are already implemented

Check with your HR information systems provider(s) to determine which AI-based tools are included in their systems and which are enabled for your organization

Develop a notification/reporting system to flag the adoption of AI-based tools

Develop a review policy for tools that are initially adopted as a trial or pilot

Notes on these points:

[Return to section](#)

Forming an assessment team and planning for the long term

Which departments/individuals in your organization should be involved in the assessment and adoption of AI for HR tools?

What types of external expertise might be necessary and where might you find them?

Which decision makers and individuals at the top of the organization need to be informed or involved?

Scan the organization for governance structures that:

Would impact the adoption of AI for HR tools

Specify organizational principles and policies for the use of AI

Could serve as a model for an AI for HR assessment process

[Return to section](#)

What is the purpose of adopting the AI-based tool?

Assess the organization's current capabilities and AI journey

Has the organization used AI-based tools before? In HR?

What were these previous uses and were they successful? What lessons can be learned from these previous experiences?

Are there individuals in the organization with knowledge of AI?

What is the current state of the data infrastructure?

Is there buy-in from top leadership?

What is the attitude of employees toward technology and AI?

Given the answers to these questions, what should be the first/next step on your AI journey?

Notes on these points:

Consider a few types of tools identified as possible priorities in your answers to the [Many uses of AI](#) chapter

What would be the purpose of adopting such a tool and how would it change existing practices?

What organizational outcomes would you hope that the tool would improve?

Would you be able to assess/document this improvement?

How would it fit into the organization's AI journey?

Notes on these points:

[Return to section](#)

What are the core elements of the tool?

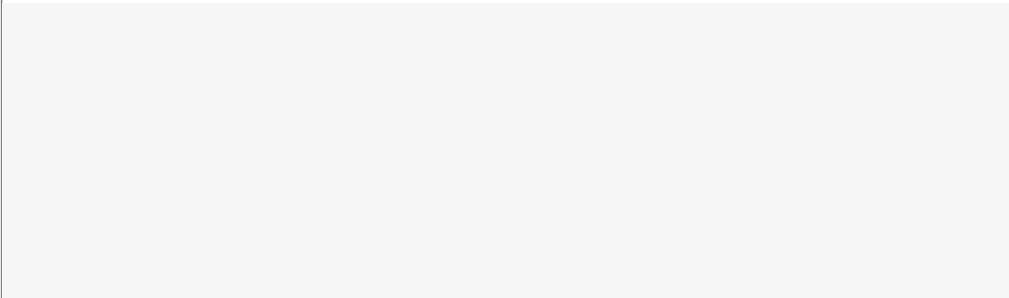
Choose one or two example of tools that you have seen on the market and try to use the public information available on these tools to answer the questions in the tool assessment questionnaire for this section.

[Return to section](#)

Assessing the risk level of a tool

Using the same example tools from the previous section, complete the risk assessment checklist in the assessment questionnaire for this section.

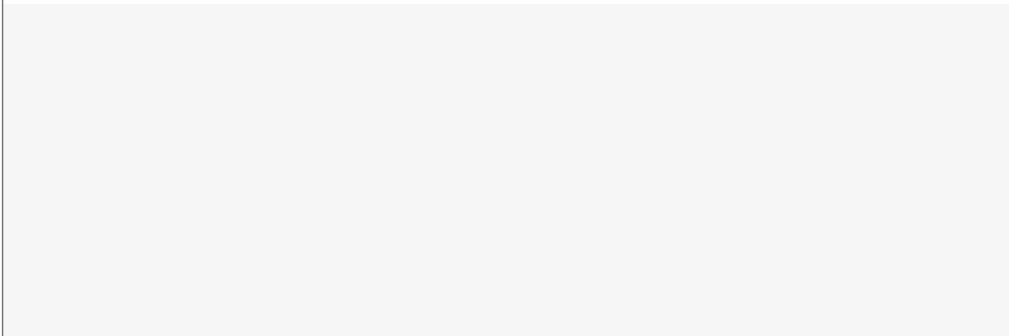
Does this risk checklist work for your organization? Is there anything that should be added or removed?



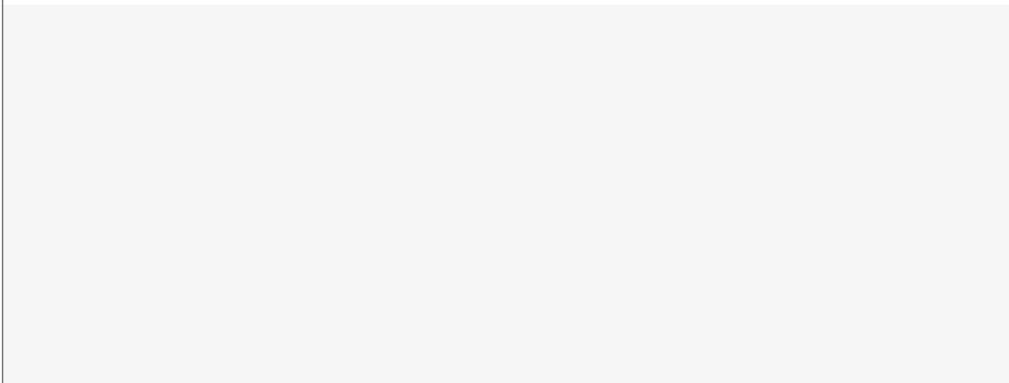
Discuss the items in the risk checklist

Which are your highest priorities?

Are there "no go" points, types of tools or risks that your organization will rule out ahead of time?



Does your organization want to create a formal risk assessment system with scoring and consequences for each assessment level?



[Return to section](#)

Bias

Consider going through an example tool and discuss how this tool might both improve and encode bias using the assessment questionnaire.

What are your organization's priorities and constraints regarding diversity, bias, and discrimination?

What are your current policies and procedures?

What are the legal requirements around discrimination, protected groups, etc in each of the jurisdictions where you operate?

Notes on these points:

What resources do you have in your organization, internal or external, that you could draw on to evaluate and anticipate problems with bias in an AI-based HR tool?

Consider key metrics and outcomes in your organization that might be used in an AI system such as performance, promotion or turnover.

What do they capture and what don't they capture?

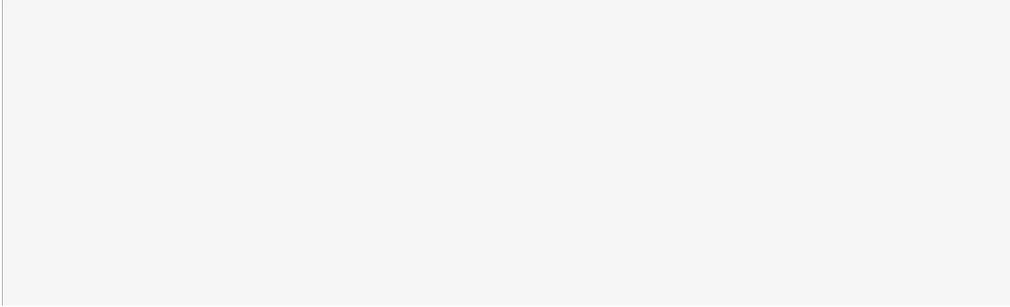
To what extent could bias get incorporated into these measures?

Notes on these points:

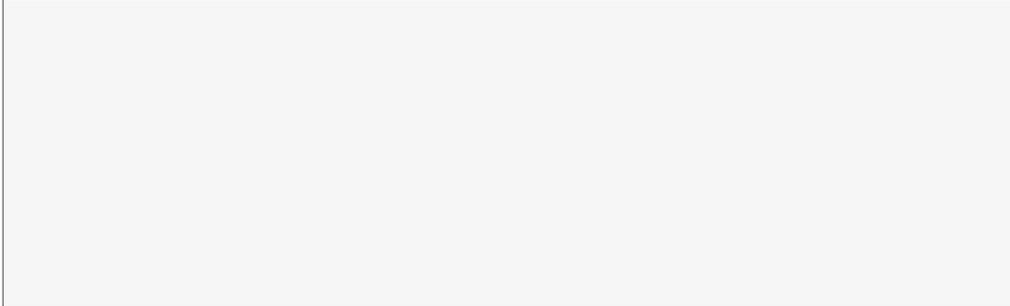
[Return to section](#)

Data privacy

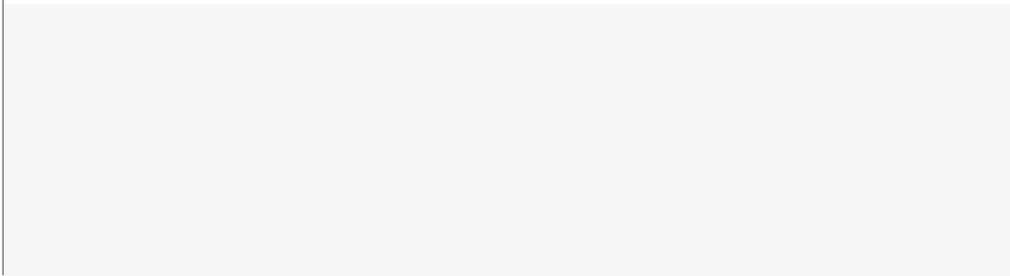
Who will provide expertise on data privacy and security?



What are the data privacy laws in the jurisdictions where you operate?



If you operate or have employees in the European Union, does your organization have a designated data protection officer tasked with ensuring compliance with GDPR?



Consider an example tool

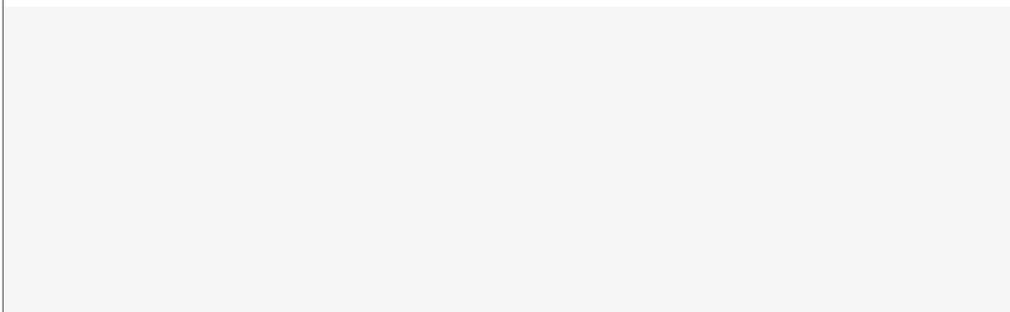
What types of data would it use?

Might the collection or use of that data raise reputational risks or undermine trust in the organization?

Might it raise legal risks?

How might these risks be mitigated?

Notes on these points:



[Return to section](#)

Transparency and explainability

Consider one or two example tools

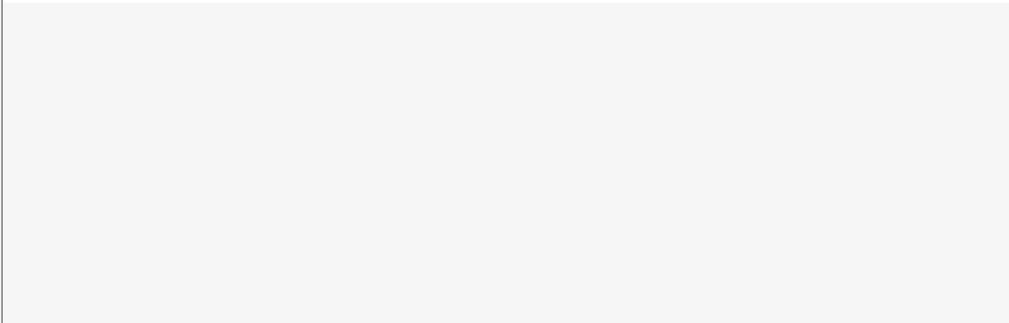
What might transparency look like in this case?

How important would explainability be for this use case?

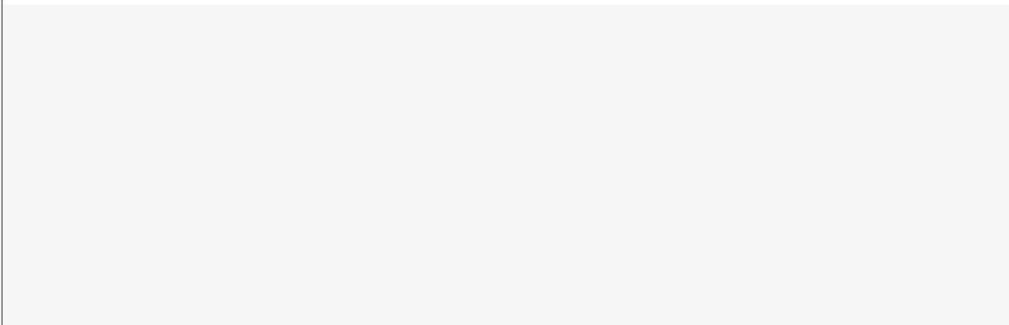
What information would be important in a global and local explanation?

Would you share local explanations with the individuals?

Notes on these points:



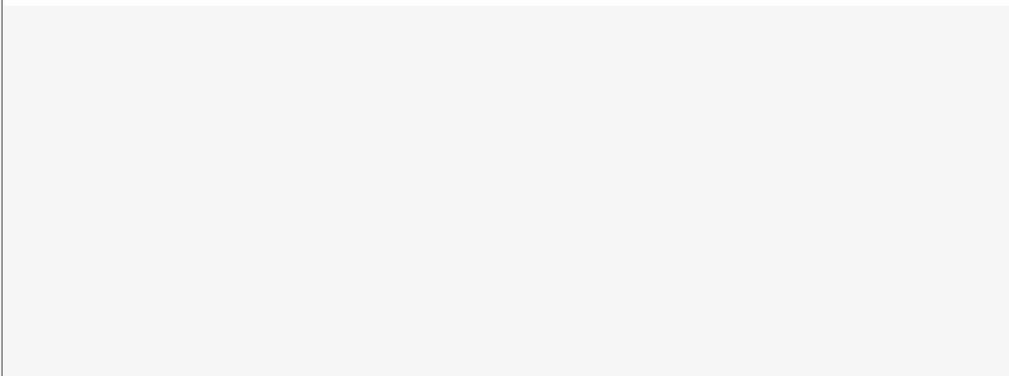
Do you want to set policies on transparency and explainability?



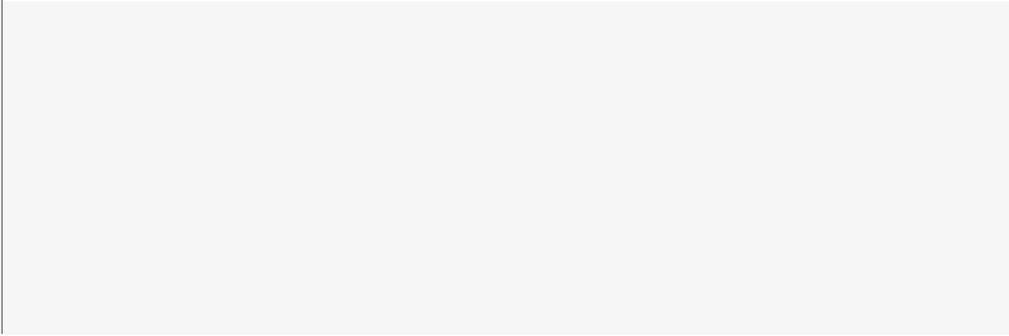
[Return to section](#)

Implementation and buy-in

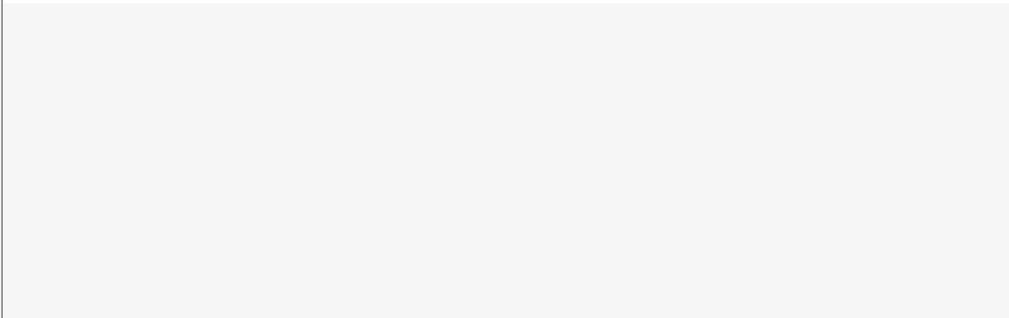
What are your current organizational resources and practices for change management? What additional resources or practices might be needed for a change that involves the adoption of an AI-based technology?



What is your current system for training HR employees? Would training on the use of an AI tool be incorporated into that training system or implemented separately?



What are the current communication channels with employees? How might information about the adoption and use of AI-based HR tools use these channels? How might feedback be gathered from employees or impacted individuals?

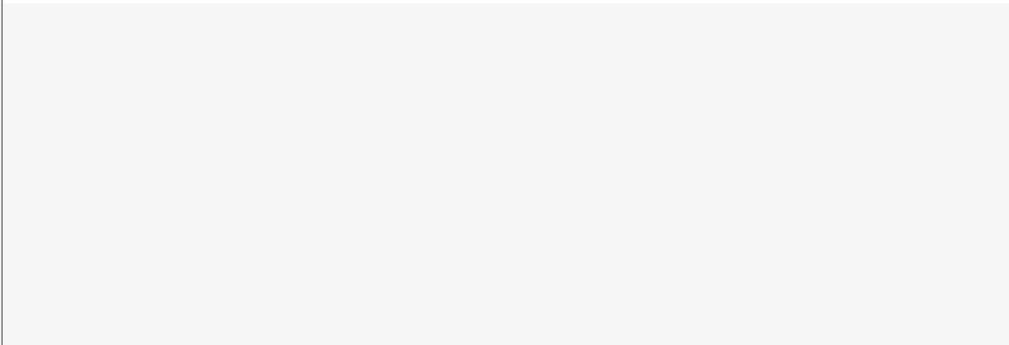


Consider setting broad policies, guidance or principles on:

Humans always having oversight of AI in HR and making final decisions, possibly specifying a system to allow for exceptions

Employee involvement in decisions on the adoption and use of AI-based HR tools

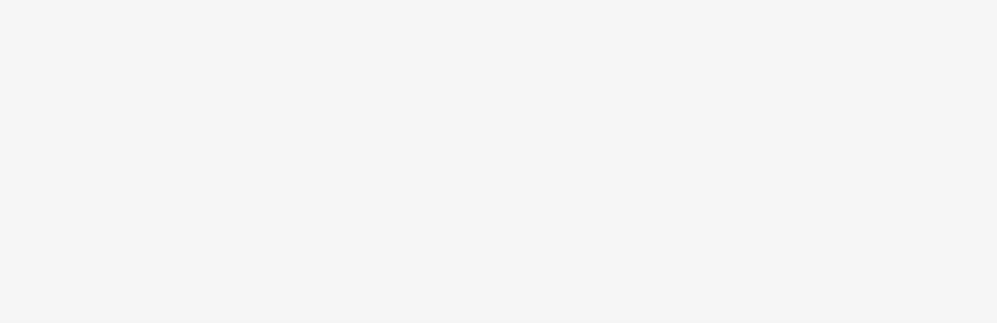
What opportunities might there be to pilot or test AI-based tools within the organization? Would you consider doing a randomized experiment?



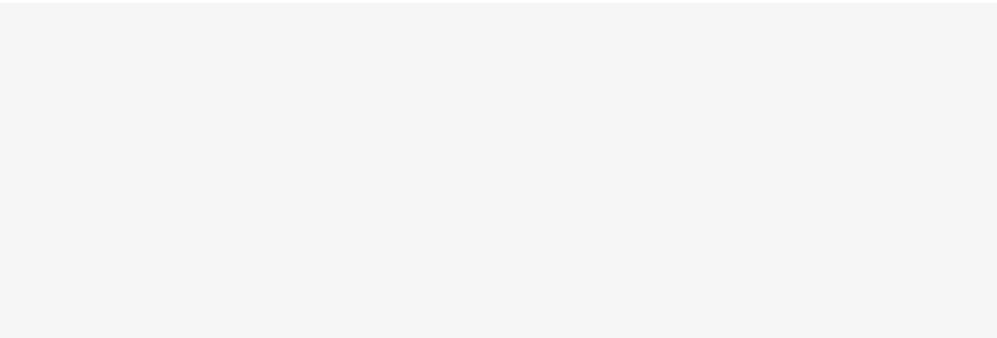
[Return to section](#)

Ongoing maintenance and monitoring

What monitoring systems are currently in place in the organization and how might the monitoring of AI-based tools be incorporated into these systems? For instance: measurement and reporting of key performance indicators (KPIs) that might be affected by the tool, employee engagement surveys that could assess attitudes, HR dashboards, etc.



Take one or two example tools and consider a recent change that has faced your organization (e.g. the COVID-19 pandemic). Would you anticipate this change of context impacting the performance of an example tool? There is no definitive answer here, you will need to use your own judgement on whether you believe the patterns in historical training data would still hold for the specific task that the tool is undertaking. This exercise will be useful for recognizing future changes in context where you will want to step up monitoring for possible problems.



[Return to section](#)

Acknowledgements

The World Economic Forum's Human-Centred Artificial Intelligence for Human Resources project is a multistakeholder initiative that has engaged leaders from private companies, human resources professional associations, governments, civil society

organizations, and academia to identify challenges and develop best practices for the responsible use of AI in HR. The opinions expressed in this toolkit may not correspond with the opinions of all members and organizations involved in the project.

Lead author

Matissa Hollister

Assistant Professor of Organizational Behaviour,
Desautels Faculty of Management, McGill University

Project leaders

Tunç Acarkan

Technology Management Director, Centre for
the Fourth Industrial Revolution Turkey, Turkish
Employers' Association of Metal Industries (MESS)

Roderick Beudeker

Senior Associate, Baker McKenzie

We would also like to thank the following members of the project community and other individuals who contributed their time and insights

Cortnie Abercrombie

Chief Executive Officer and Founder, AI Truth

Alexander Alonso

Chief Knowledge Officer, Society for
Human Resource Management

Şirin Altıok

Big Data Architect, TOFAŞ

Colin Anderson

Managing Director, Accenture

Dani Benreytan

Co-Founder, Sparkus

Eric Bokelberg

Associate Partner, Global Human
Resources Innovation, IBM

Alexander Cohen

Director, Marketing, Eightfold

Brian Crane

Senior Director, Medtronic

Stephanie Du

Integrated Management Student Fellow,
Desautels Faculty of Management, McGill University

Merve Ergin Tosun

Turkish Employers' Association of Metal
Industries (MESS)

Jill Finlayson

Director, Expanding Diversity and Gender
Equity in Tech, University of California, Berkeley

Justin Furlong

Director, Human Resources Compliance and
Employee and Labour Relations, Medtronic

Alya Gabsi

Master's Degree Candidate in Management
Analytics, McGill University

Ashutosh Garg

Chief Executive Officer and Co-Founder,
Eightfold AI

Elif Gedik

Turkish Employers' Association of Metal
Industries (MESS)

Steve Gill

Director, Digital Talent AI, EY Global Services

Ilana Golbin

Director, PwC Emerging Technologies

Elaine Greenway

Director, Workforce Research, Society for Human Resource Management

Tom Hogan

Professor of Practice, Human Resource Management, Penn State University

Ehsan Hoque

Associate Professor of Computer Science, Rochester University

Ani Huang

Senior Vice-President, HR Policy Association

Veena Jadhav

Associate Dean and Associate Professor of Leadership and Human Resources Management, S P Jain School of Global Management

Shalin Jyotishi

Senior Policy Analyst, Education and Labour, New America; Visiting Scholar, American Association for the Advancement of Science (AAAS)

Athena Karp

Chief Executive Officer and Founder, HiredScore

Michelle Keating

Director, Artificial Intelligence Centre of Excellence for HR, Accenture

Alex Ley

Senior Business Analyst, Innovation, Science and Economic Development, Government of Canada

Stela Lupushor

Senior Fellow and Programme Director, The Conference Board

Garry Mathiason

Senior Partner and Co-Chair, Robotics, AI and Automation Industry Group, Littler

Aaron Matos

Chief Executive Officer and Founder, Paradox.ai

Elizabeth Morrison

Learning and Development Specialist, Private Fundraising and Partnerships, UNICEF

Leyla Nascimento

President, World Federation of People Management (2018 - 2020)

Ernest Ng

Vice-President, Global ES Strategy and People Analytics, Salesforce

Selin Nugent

Assistant Director, Institute for Ethical AI, Oxford Brookes University

Ntsibane Ntlatlapa

Head, Centre for the Fourth Industrial Revolution South Africa

Yasemin Oral

Head, Corporate Communications, Turkish Employers' Association of Metal Industries (MESS)

Tania Pagliarello

Public Sector Human Resources Professional

Aleatha Parker-Wood

Principal Privacy Engineer, Amazon

Frida Polli

Chief Executive Officer and Founder, Pymetrics

Manish Raghavan

Postdoctoral Fellow, Harvard Center for Research on Computation and Society

Ozlem Sarioglu

Co-Founder, Sparkus

Matthew Scherer

Senior Policy Counsel, Worker Privacy, Center for Democracy and Technology

Susan Scott-Parker

Founder, Business Disability International

Avi Simon

Co-Founder and Chief Technology Officer, retrain.ai

Zach Solan

Vice-President, Data and Artificial Intelligence, retrain.ai

Xavier St-Denis

Assistant Professor of Social Inequalities, Institut National de la Recherche Scientifique

Kelly Trindel

Head, Public Policy, Pymetrics

Lynette Yarger

Associate Professor of Information Sciences and Technology, Penn State University

Valery Yakubovich

Executive Director, Computational Social Science, Wharton School, University of Pennsylvania

Josh Zywiec

Chief Marketing Officer, Paradox.ai

We thank the following organizations for hosting pilots, workshops and focus groups

- Centre for the Fourth Industrial Revolution Turkey, an Affiliate Centre of the World Economic Forum established by the Turkish Employers' Association of Metal Industries (MESS) and Ministry of Industry and Technology
- Türk Traktör
- Mercedes Benz Türk
- United Nations Fund for Children, Private Fundraising and Partnerships Division; Canadian UNICEF Committee; and United States Fund for UNICEF
- Society for Human Resource Management
- Centre for the Fourth Industrial Revolution South Africa

Endnotes

1. Data shared by CB Insights: <https://www.cbinsights.com/>.
2. Kahneman, D., *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 2011.
3. For guidance on establishing organization-level AI principles and practices, see World Economic Forum, “Empowering AI Leadership”, 2021, <https://www.weforum.org/projects/ai-board-leadership-toolkit>.
4. Confusingly, the term bias is used in several different ways. In statistics, bias often refers to problems in the measurement or representativeness of the data. In machine learning, bias refers to an algorithm that performs poorly. Here, though, we are using the term bias as it is used more widely in society: a system that treats some social groups unfairly.
5. However, humans likely underestimate the uniformity and scale of human biases, as social sciences and neuroscience have shown that bias is widespread and likely fundamental to human nature.
6. Dastin, J., “Amazon scraps secret AI recruiting tool that showed bias against women”, Reuters, 11 October 2018, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.
7. Cowgill, B., “Bias and Productivity in Humans and Machines”, Upjohn Institute Working Paper 19-309, 2019, Columbia Business School Research Paper, 6 August 2019, <https://ssrn.com/abstract=3433737>, <http://dx.doi.org/10.2139/ssrn.3433737>.
8. Hollister, M., “Here’s how to check in on your AI system, as COVID-19 plays havoc”, World Economic Forum, 22 May 2020, <https://www.weforum.org/agenda/2020/05/here-s-how-to-check-in-on-your-ai-system-as-covid-19-plays-havoc/>.



COMMITTED TO
IMPROVING THE STATE
OF THE WORLD

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744
contact@weforum.org
www.weforum.org