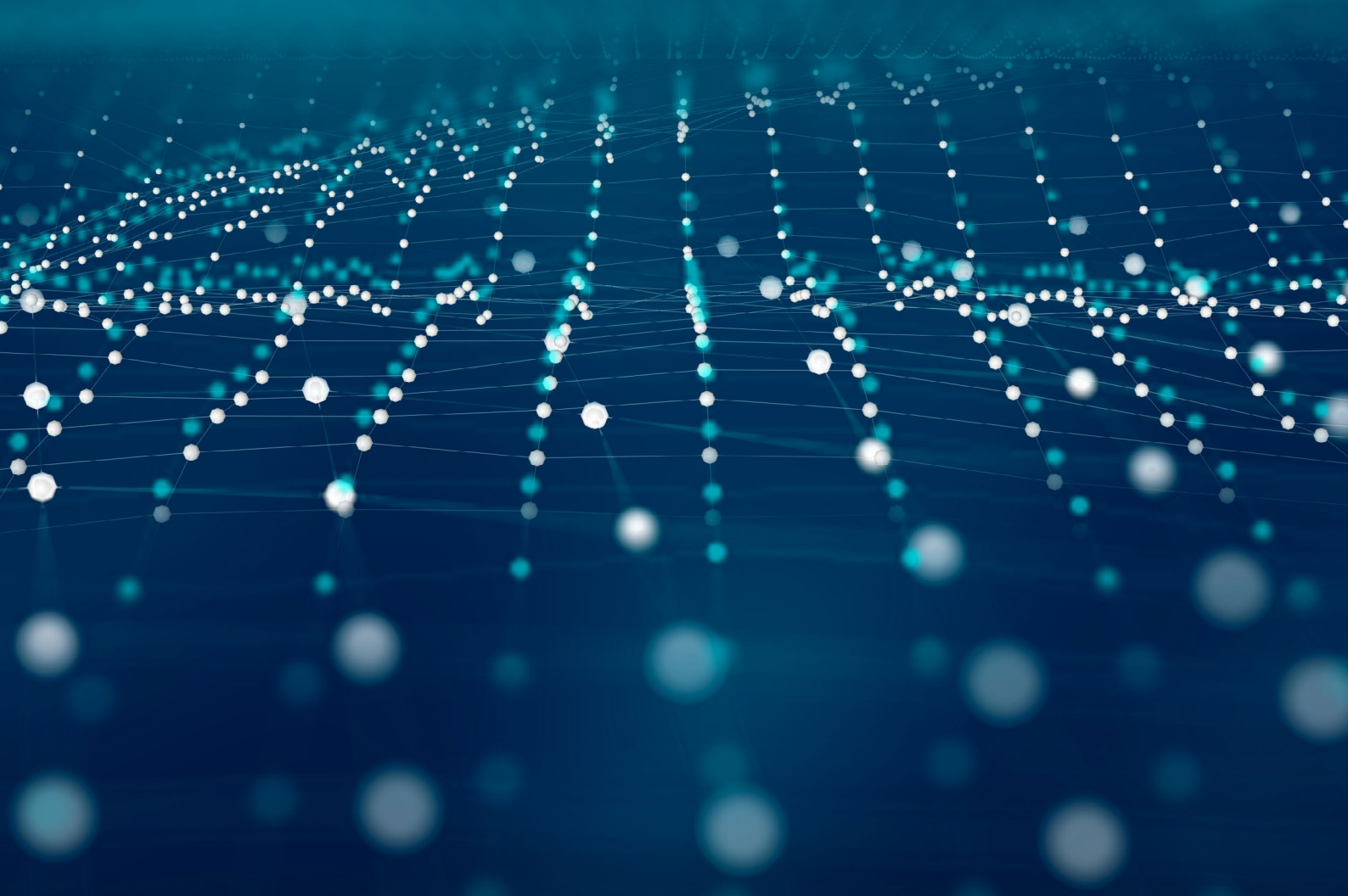


Human-Centred Artificial Intelligence for Human Resources

CHECKLIST

DECEMBER 2021



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

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

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:

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:

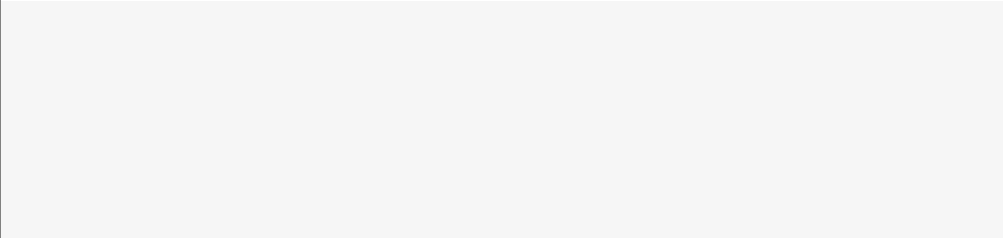
Details about the training data.

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?*

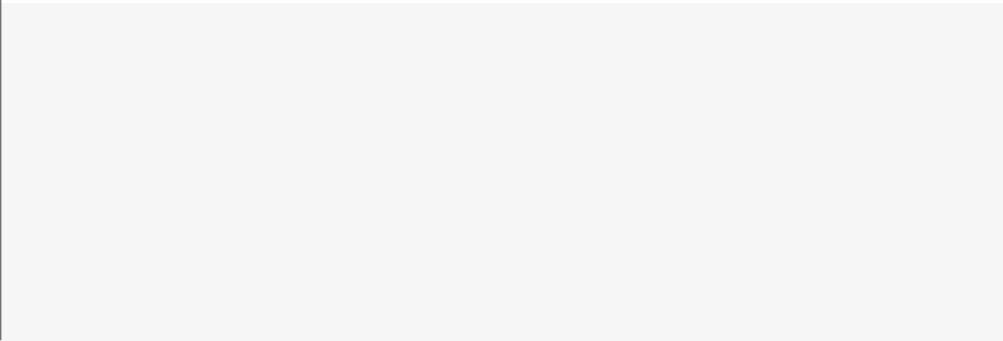
Will the system pool training data from other job types or from other organizations?

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)



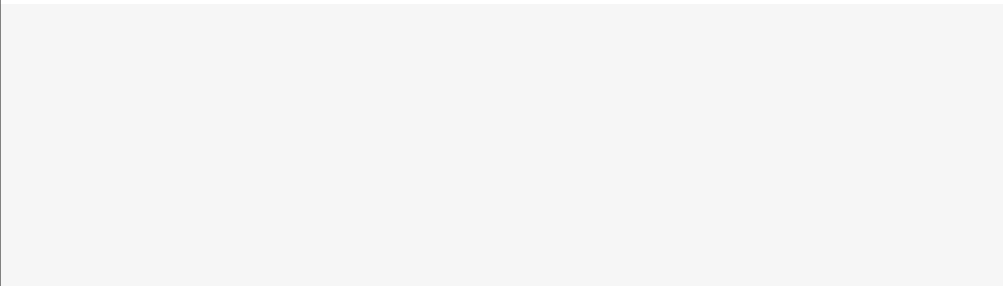
Is the algorithm predicting/maximizing an outcome?

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.*

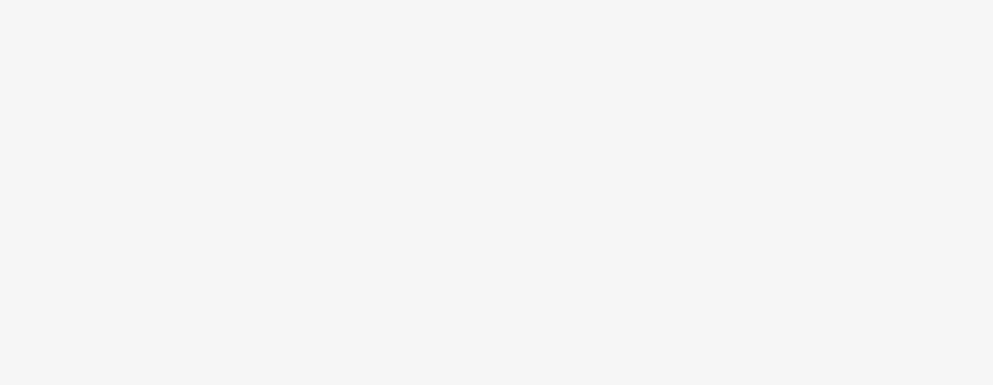
Has a third party audited the tool?

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?



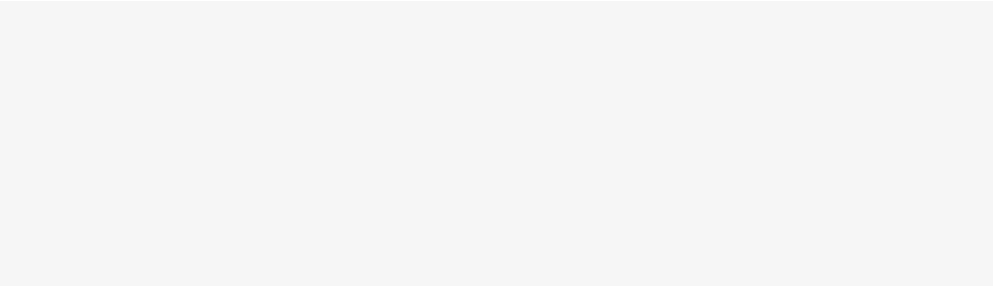
Review these components and discuss:

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:



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			

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?

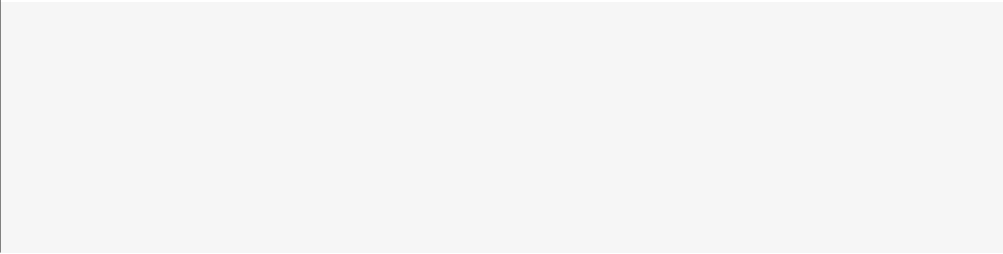
Does it match the context and population where it will be deployed?

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

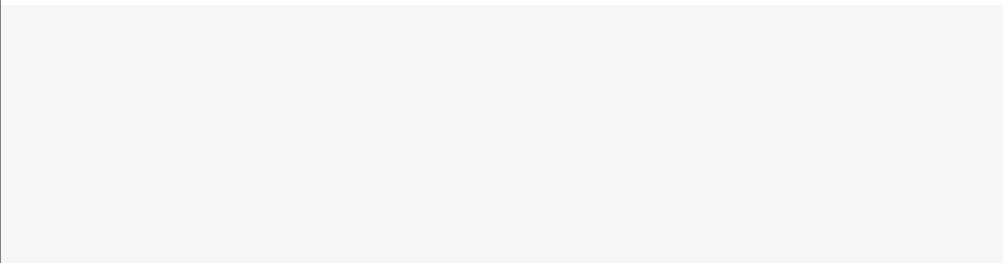


What are your organization's priorities and policies regarding bias, diversity and discrimination?

Are there particular protected groups that need to be monitored?

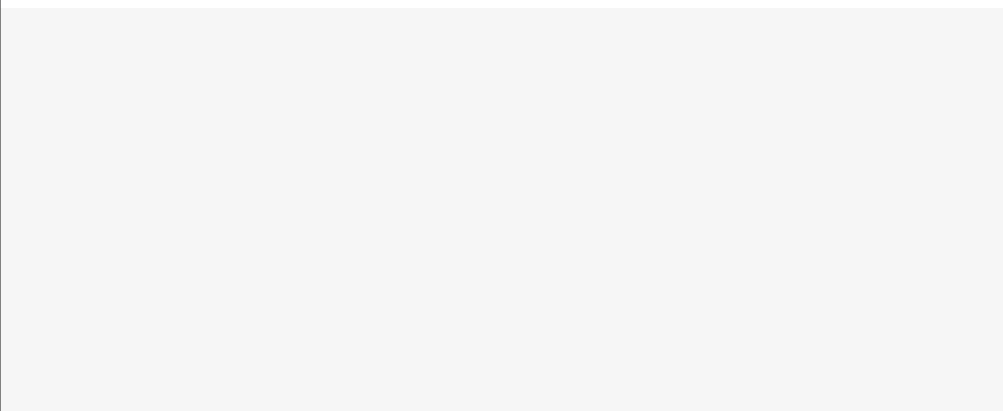
How would your organization define fairness in this context?

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

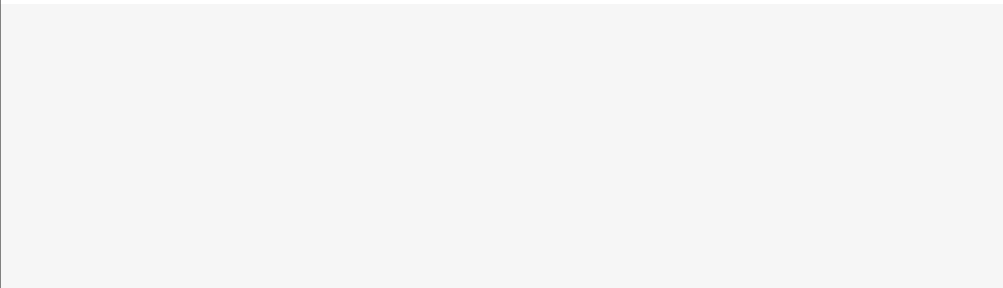
How does the tool ensure compliance with applicable laws?



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?

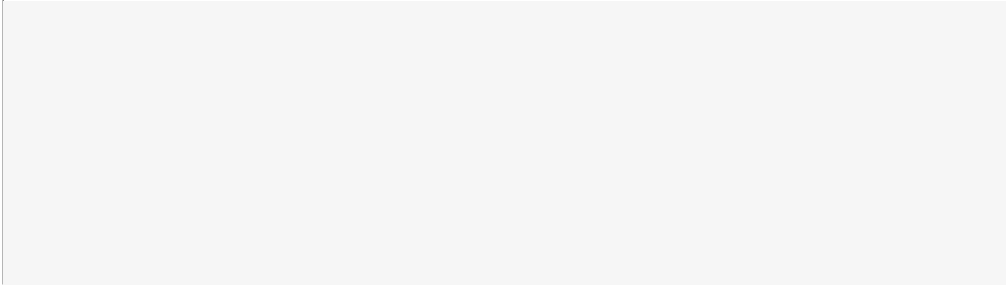
What information will they provide the organization about the mitigation process?



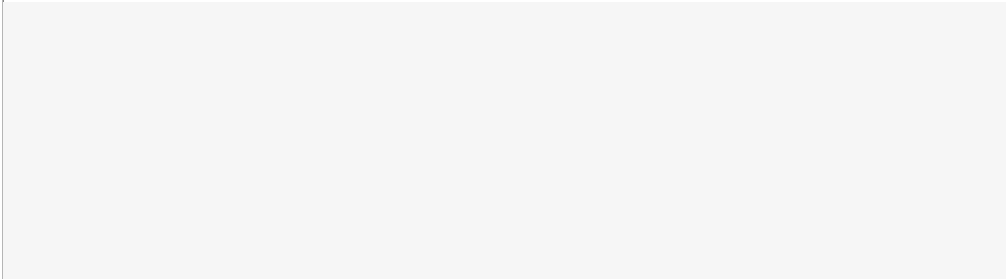
Data privacy

What types of data will the tool use?

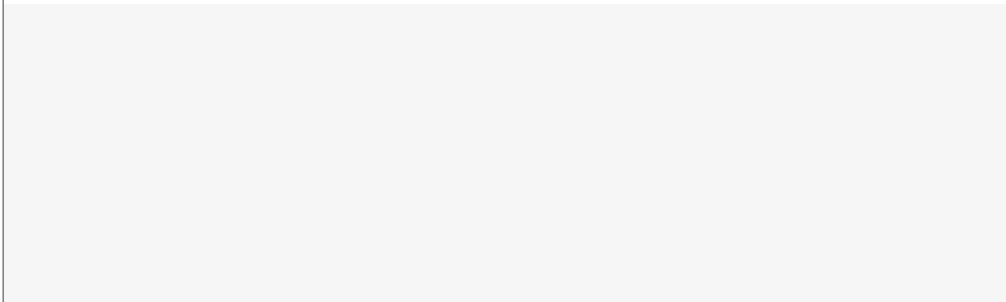
Does it meet the criteria of data minimization (only using data that is necessary)?



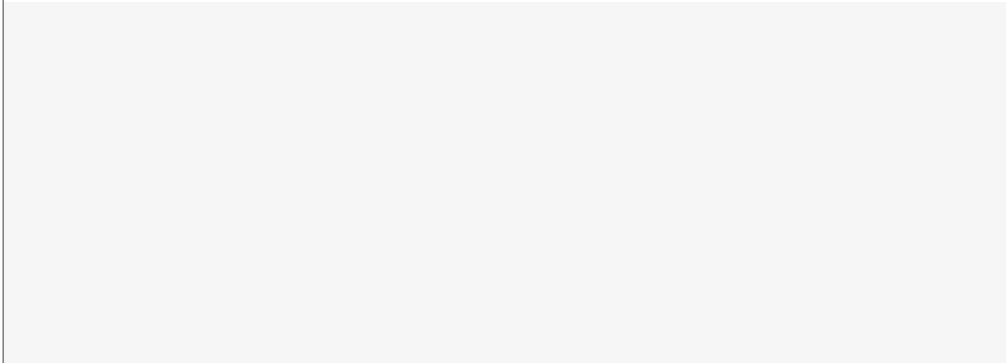
How might the collection or use of that data raise reputational risks or undermine trust in the organization?



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

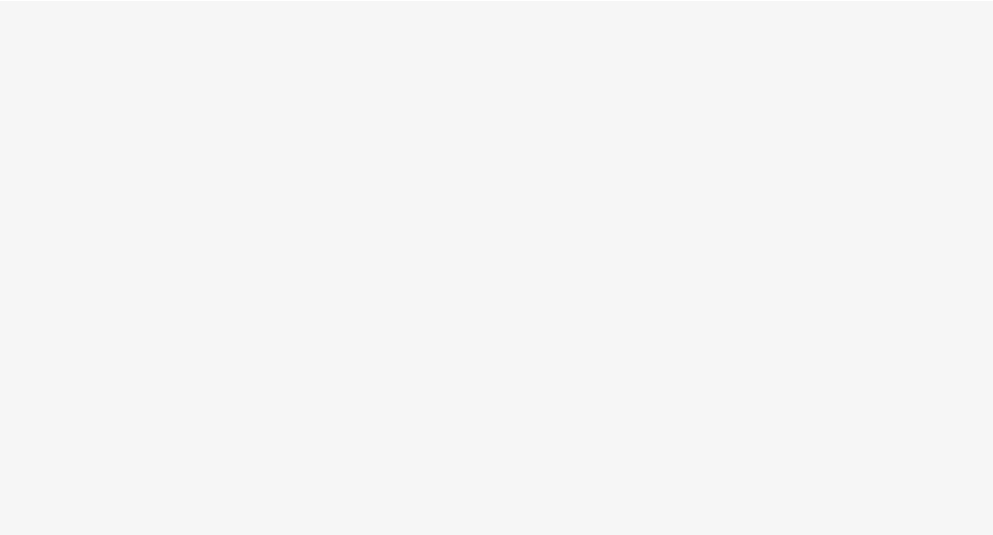
Who has permission to access the data both within the organization and at the vendor

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?

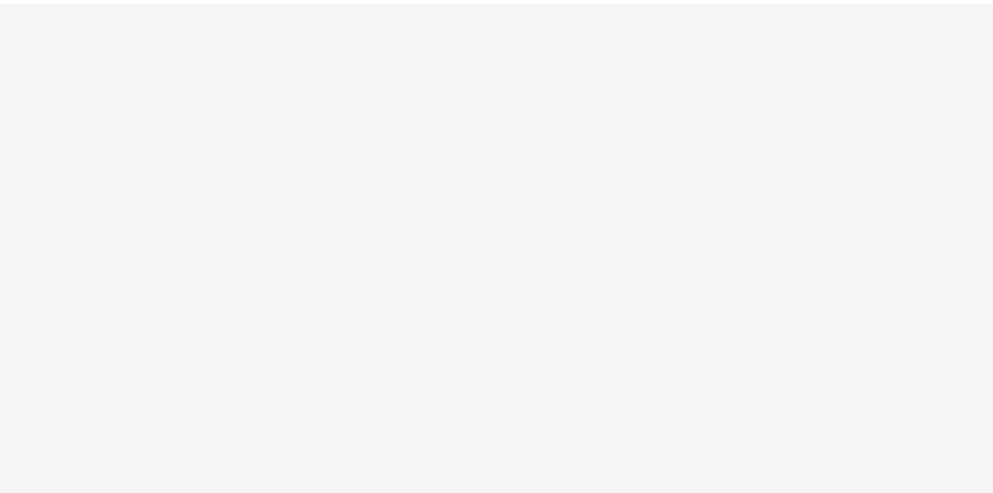
How and for how long will the data be stored?

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?



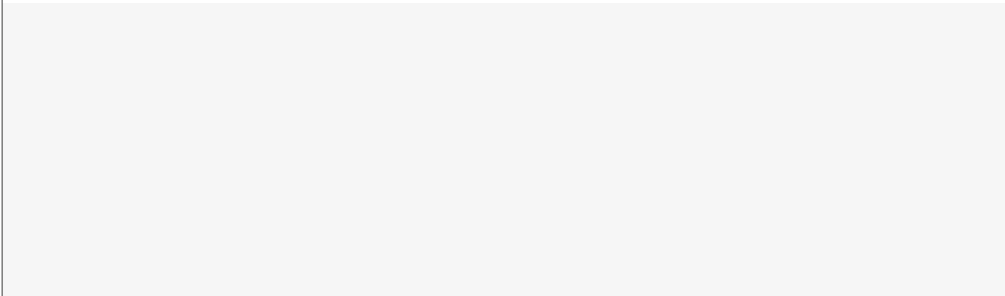
Consider creating a [Data Protection Impact Assessment](#)

Transparency and explainability

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

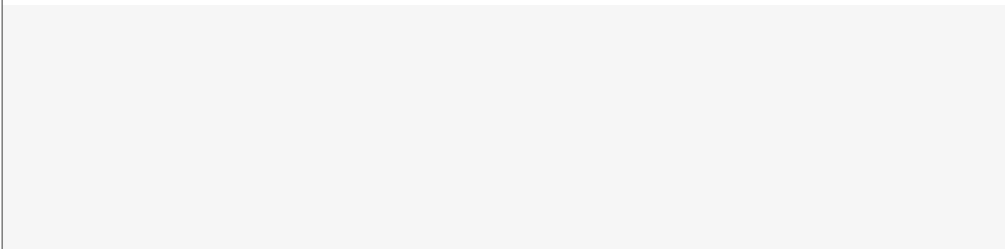
How will they be informed about the collection of data and the use of the AI-based tool?

Will you provide a meaningful opportunity to opt out?

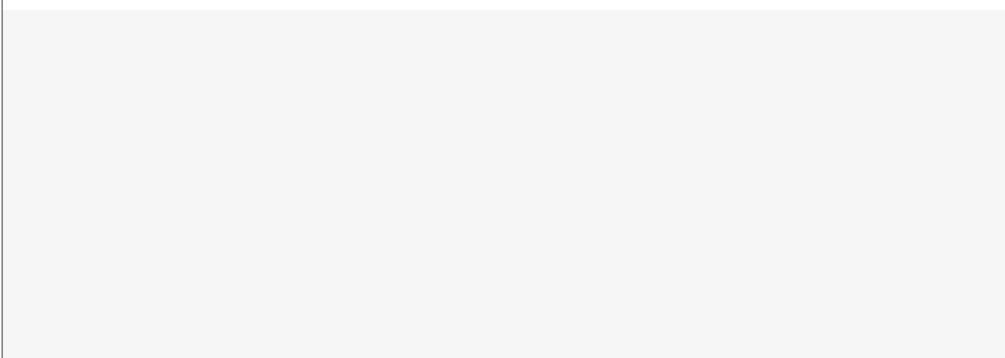


Has a third party audited the tool?

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?

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?

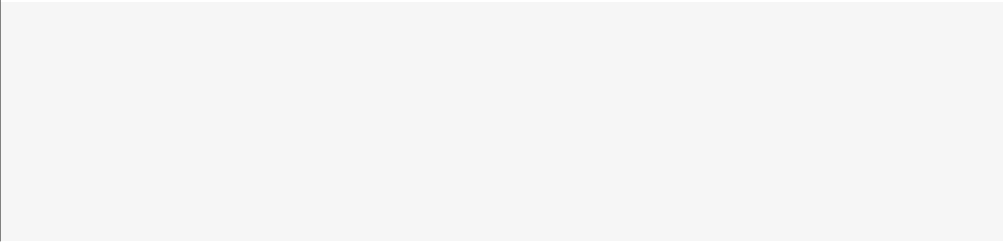
How might you encourage and document the results of cases where you do not follow the algorithm?

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?

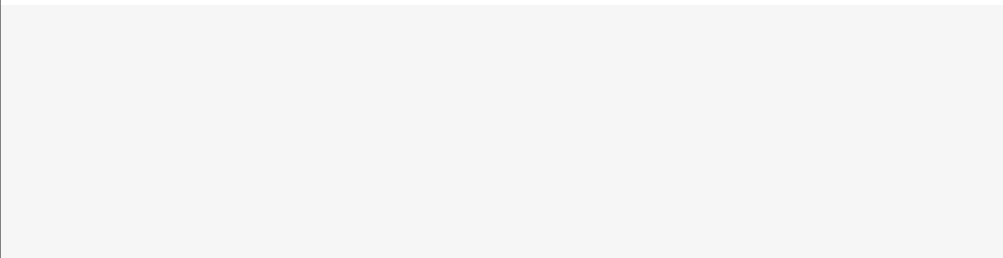
Are the reports easy to understand and user-friendly?



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

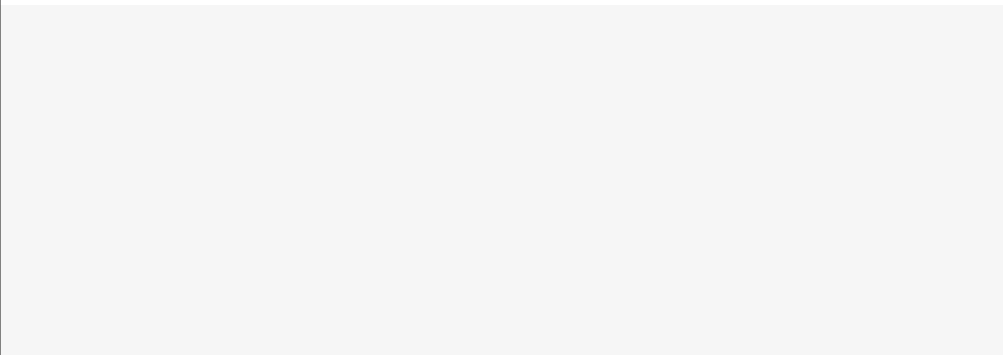
For systems that are periodically updated, pay particular attention to any changes in performance with each update

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

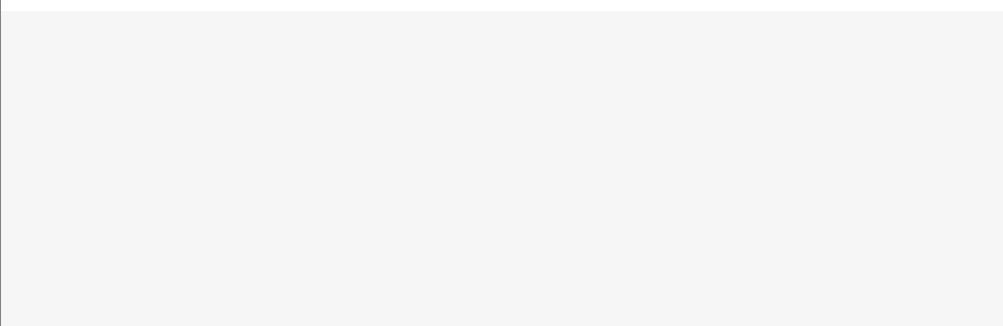
Measure baseline values



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

A large, empty rectangular box with a light gray background, intended for the user to write their 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

A large, empty rectangular box with a light gray background, intended for the user to write their periodic assessment of user and subject behaviors, including unexpected uses and 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

A large, empty rectangular box with a light gray background, intended for the user to write their monitoring of context and changes in training data relevance.

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

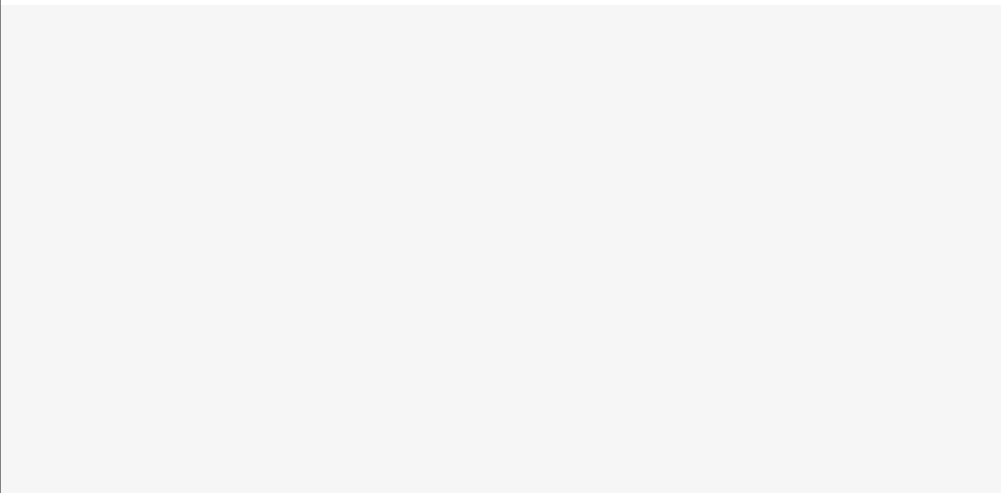
What type of AI for HR tool would be of the most interest for your organization?

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:



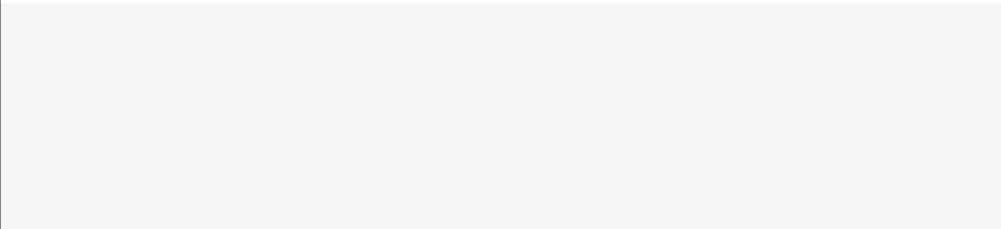
Scanning for AI-based tools that are already implemented

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:



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

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?

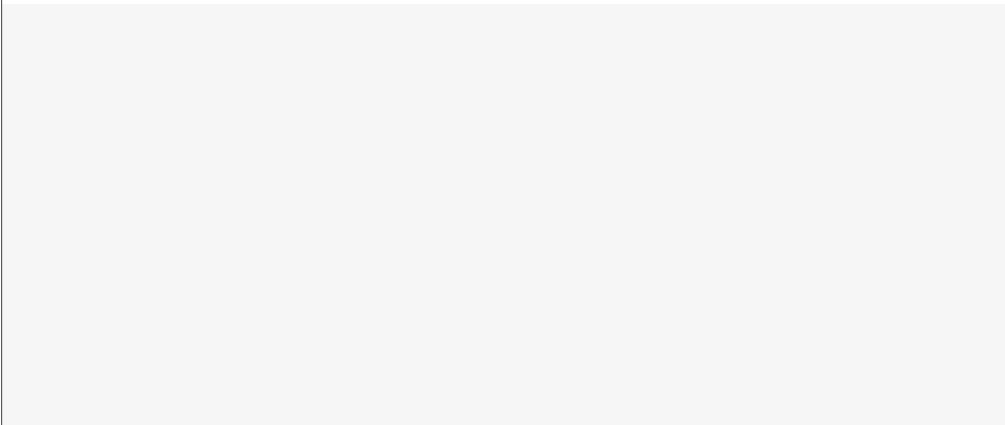
What is the current state of the data infrastructure?

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

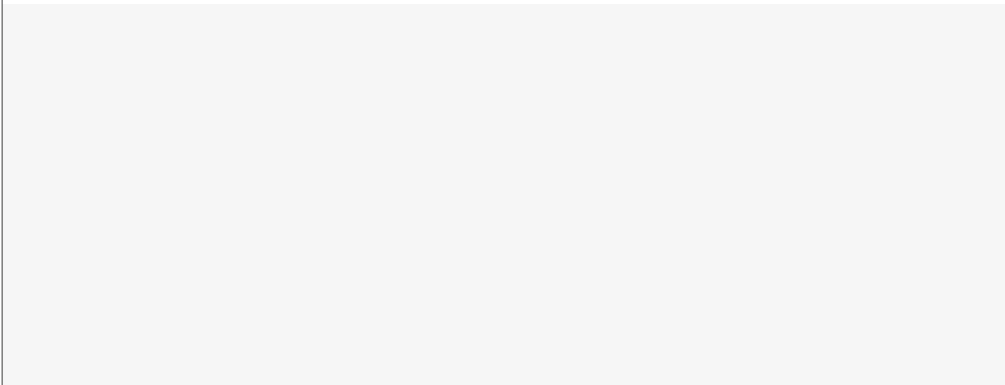
What would be the purpose of adopting such a tool and how would it change existing practices?

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:



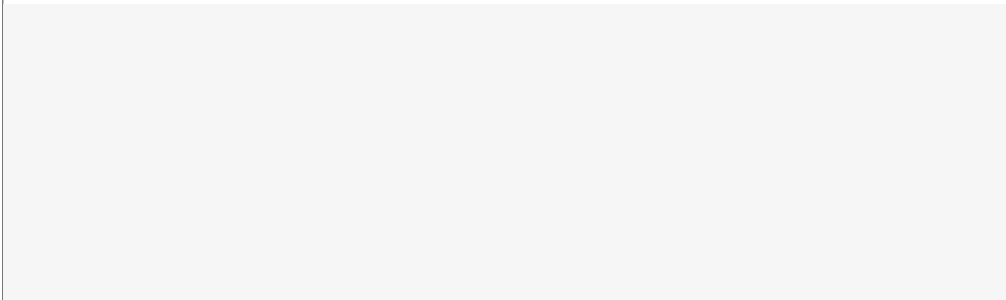
What are the core elements of the tool?

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.

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.

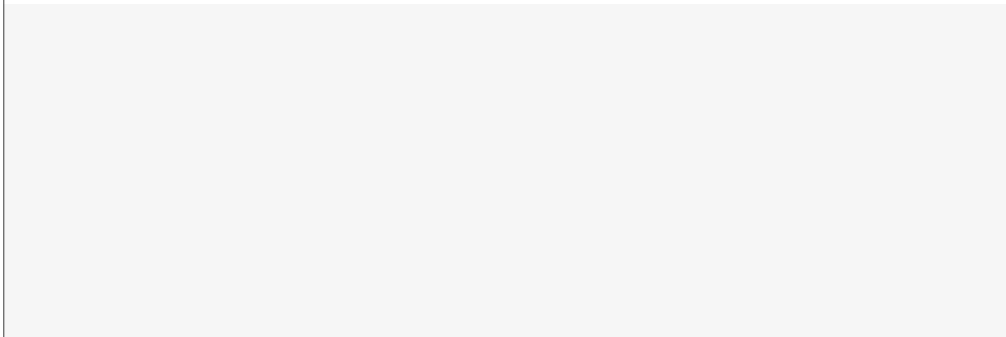
Does this risk checklist work for your organization? Is there anything that should be added or removed?



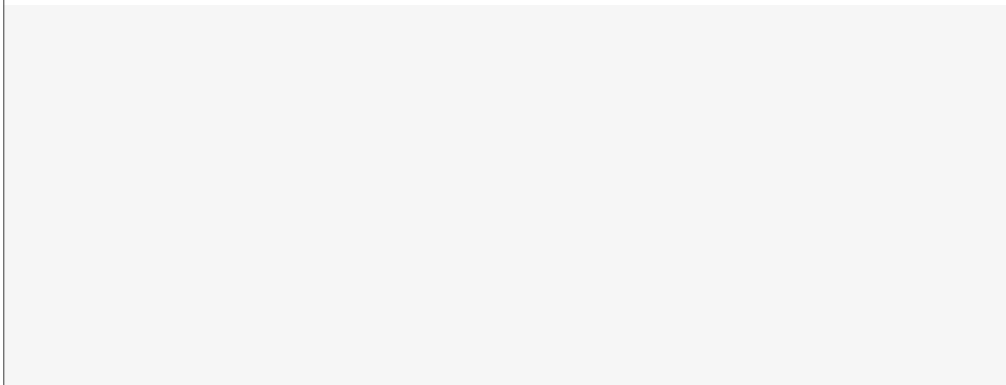
Discuss the items in the risk checklist

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?



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.

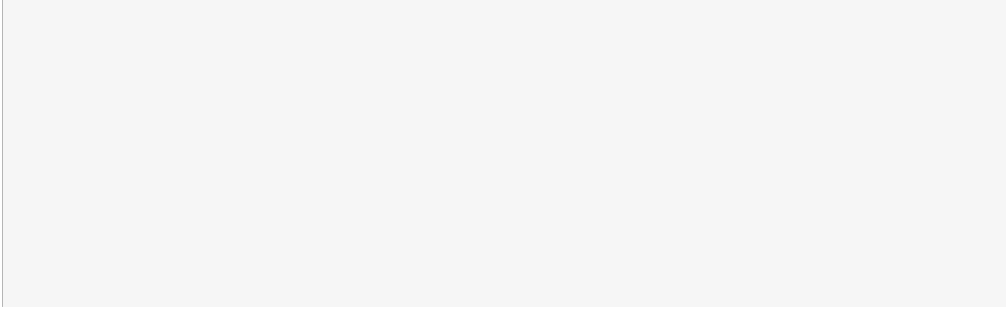
What do they capture and what don't they capture?

To what extent could bias get incorporated into these measures?

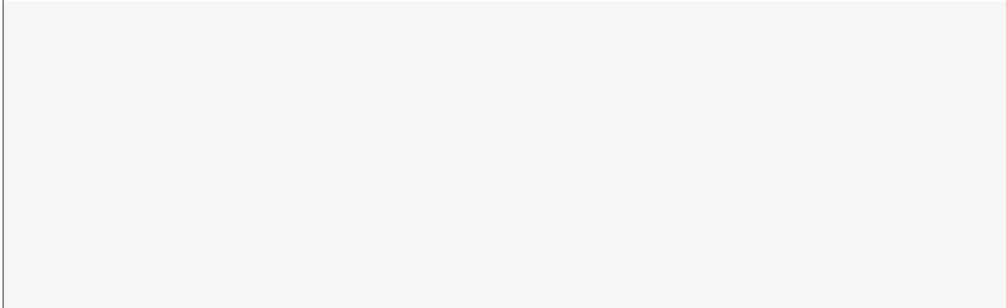
Notes on these points:

Data privacy

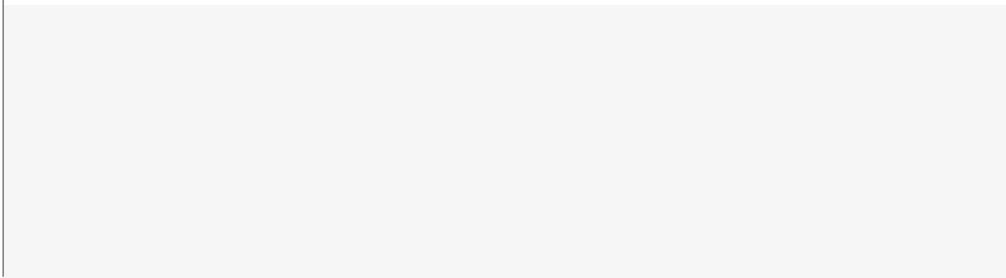
Who will provide expertise on data privacy and security?



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

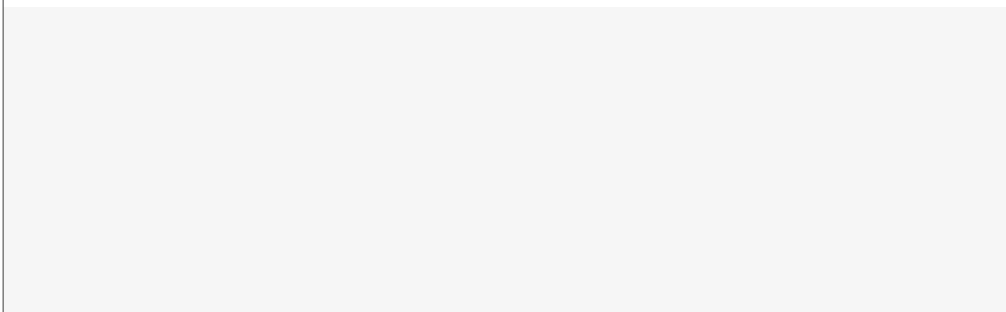
What types of data would it use?

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:



Transparency and explainability

Consider one or two example tools

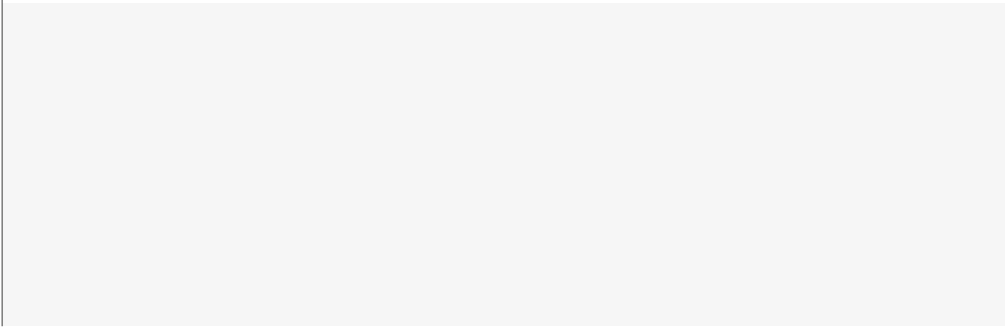
What might transparency look like in this case?

How important would explainability be for this use case?

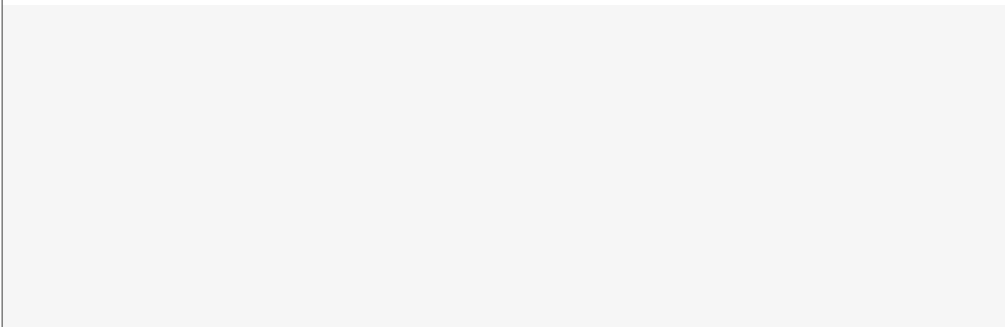
What information would be important in a global and local explanation?

Would you share local explanations with the individuals?

Notes on these points:

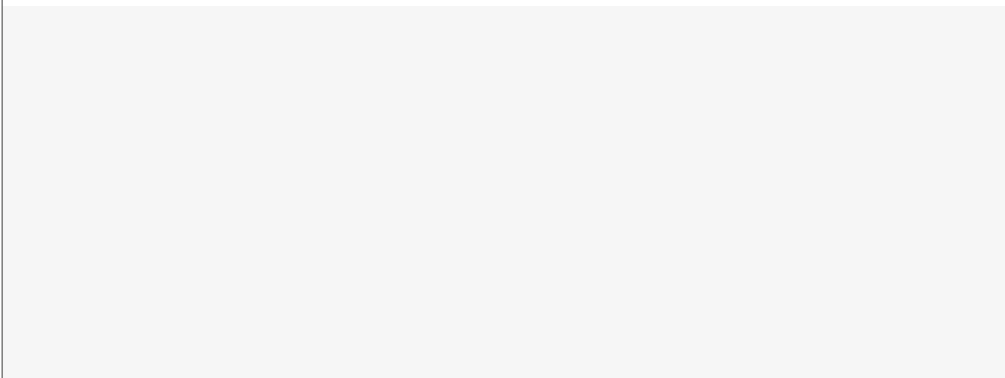


Do you want to set policies on transparency and explainability?

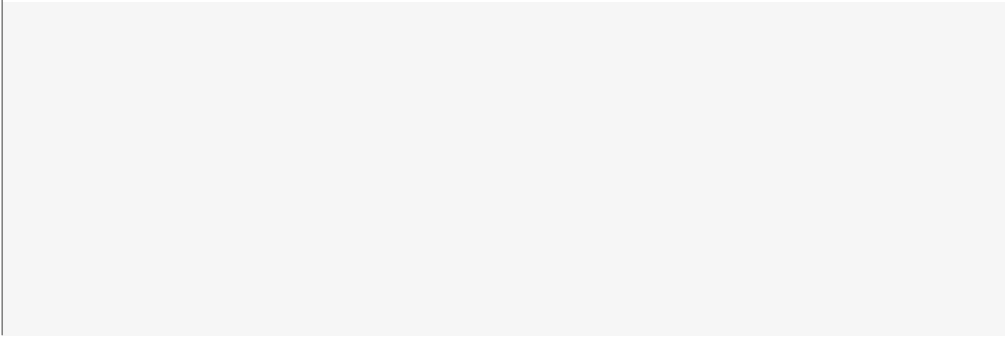


Implementation and buy-in

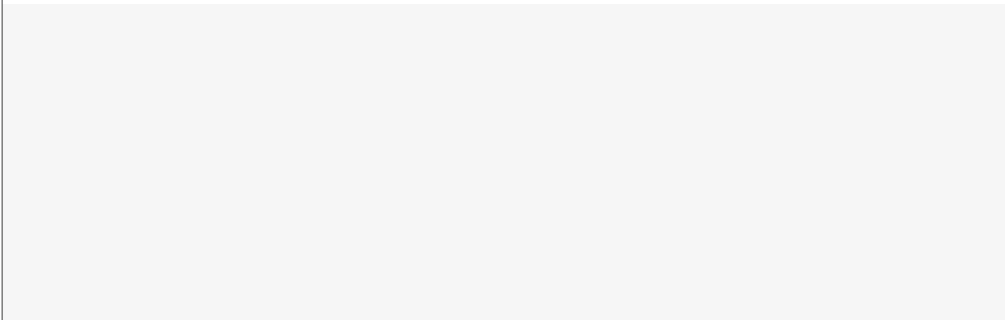
What are your current organizational resources and practices for change management? What additional resources or practices might be needed for a change that involves the adoption of an AI-based technology?



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?

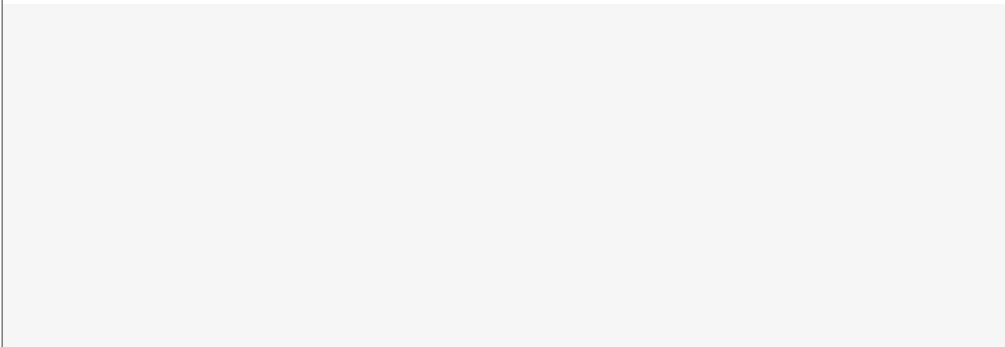


Consider setting broad policies, guidance or principles on:

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

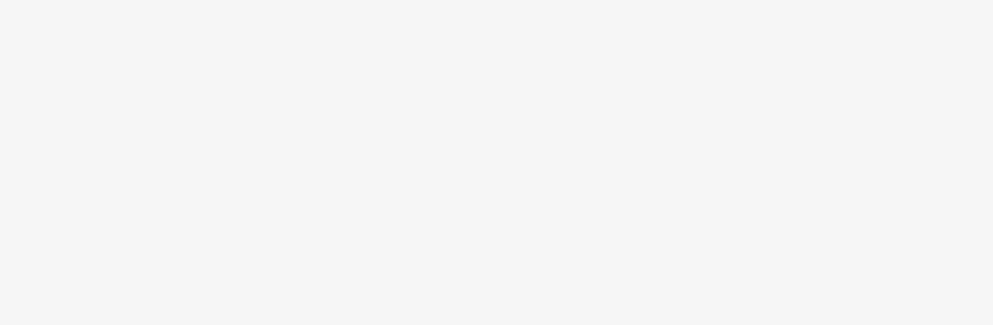
Employee involvement in decisions on the adoption and use of AI-based HR tools

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

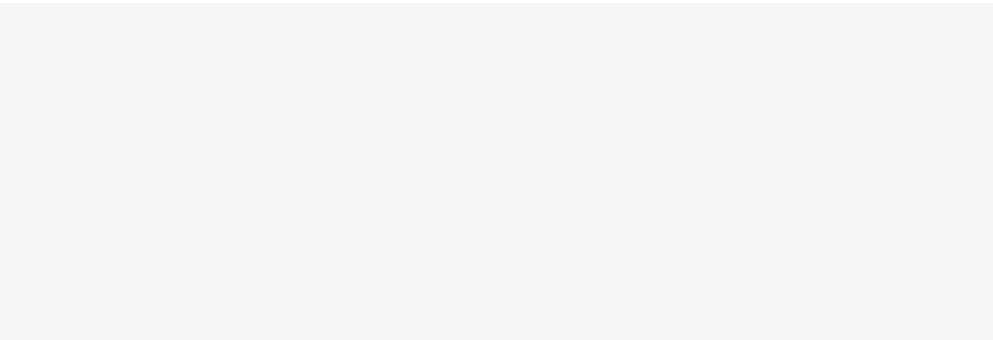


Ongoing maintenance and monitoring

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.





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