ARCTIC SHORES

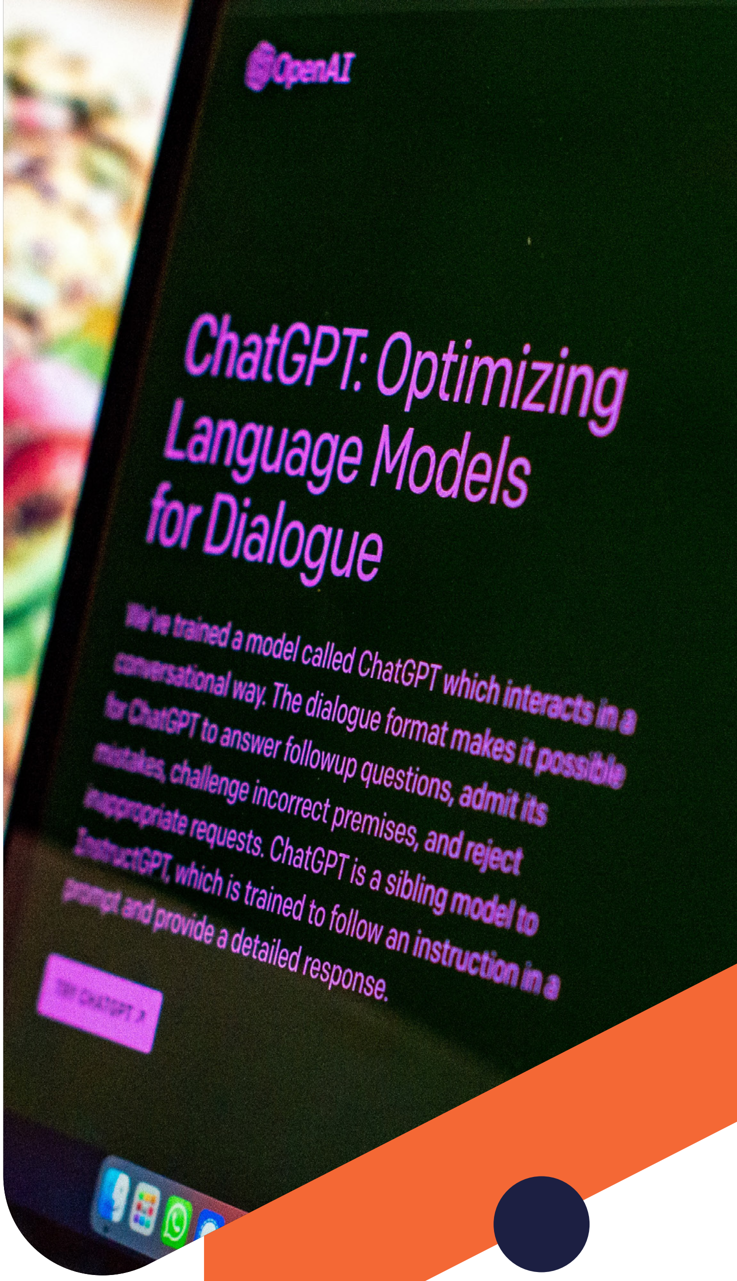
How to conduct a Generative AI vulnerability audit

Contents

Part 1

Why audit your recruitment processes?

If you already recognise the need to audit your processes then please go to section two.



OpenAI

ChatGPT: Optimizing Language Models for Dialogue

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests. ChatGPT is a sibling model to InstructGPT, which is trained to follow an instruction in a prompt and provide a detailed response.

TRY CHATGPT ↗

For anyone tasked with hiring new talent, the impact of Generative AI is unfolding with seemingly endless plot twists.

On one hand, tools like ChatGPT are automating repetitive manual tasks for recruiters. For example, writing job ads, sending rejection letters, or checking out people's past jobs and skills.

But Generative AI is also having a major impact on how candidates apply for jobs.

72% of Early Careers candidates are using some form of Generative AI every week.

And according to a survey from Greenhouse, 47% of candidates would use it to help them complete a job search.

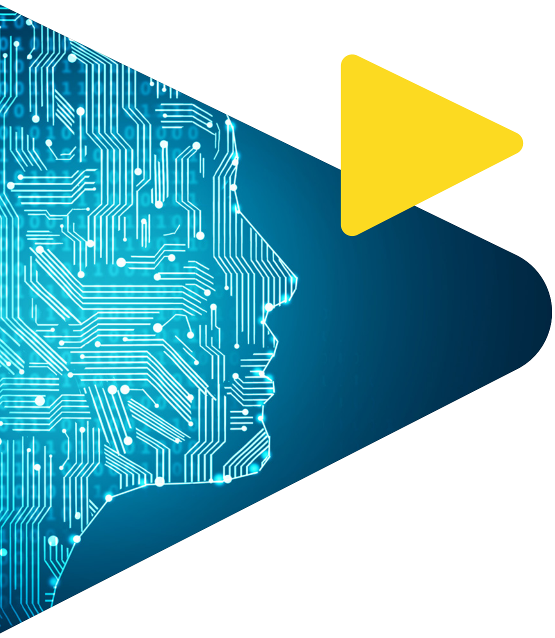
But according to TikTok, that number could be higher.

What does this look like in practice?

Candidates are using GenAI tools to send thousands of CVs at the click of a button – plus a mismatch in the quality of a candidate's application and their performance in the final stage of the interview process.

Overnight, many talent acquisition teams are finding that traditional sifting methods like CVs and application forms have become inaccurate and unscalable.

The obvious alternative for many teams would be to use a psychometric assessment – to measure candidates' true personalities and abilities...



But ChatGPT is also being used to outsmart many traditional assessments. For example, it can:

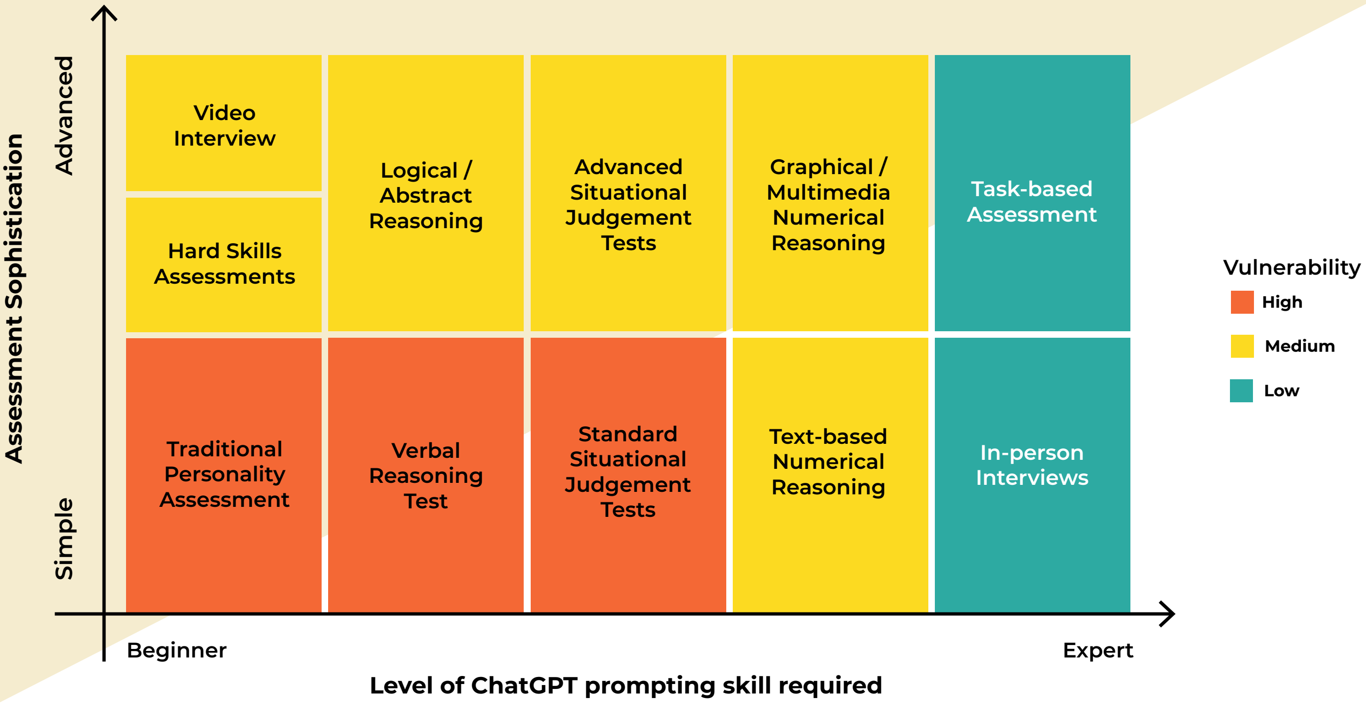
✓ **Outperform 98.8% of human candidates** in verbal reasoning tests

✓ **Score in the 70th percentile** on Situational Judgement Tests

✓ **Ace question-based Personality Assessments for any role** by simply reading the job description

Of course, not all selection tools are created equal.

For example, while many traditional question-based assessments are highly vulnerable to being completed by ChatGPT, others are designed in a way that makes them more robust.

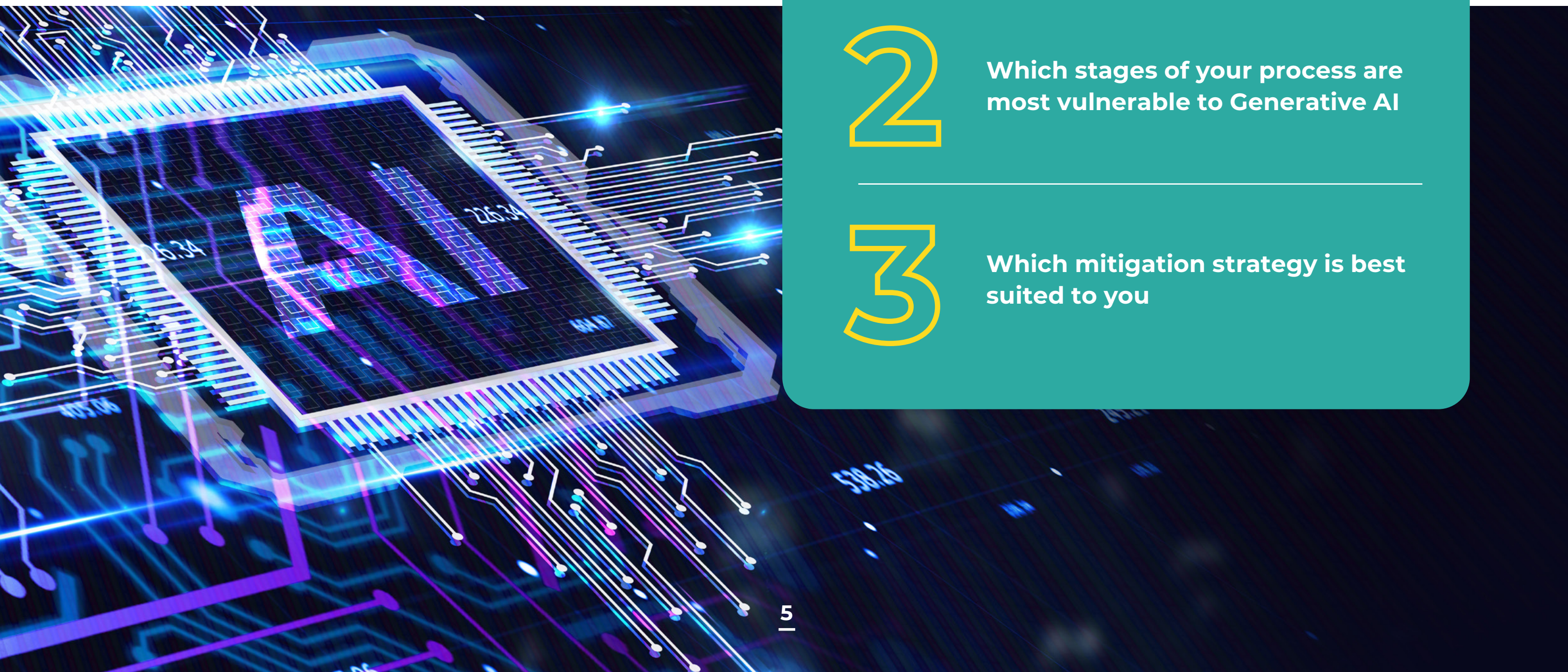


	Beginner				Expert
Advanced	🟨 Video Interview 🟨 Hard Skills Assessments	🟨 Logical / Abstract Reasoning	🟨 Advanced Situational Judgement Tests	🟨 Graphical / Multimedia Numerical Reasoning	🟩 Task-based Assessment
Simple	🟥 Traditional Personality Assessment	🟥 Verbal Reasoning Test	🟥 Standard Situational Judgement Tests	🟨 Text-based Numerical Reasoning	🟩 In-person Interviews

Vulnerability
- 🟥 High
- 🟨 Medium
- 🟩 Low

In the remainder of this practical guide, we'll explore a methodology that allows TA teams to audit:

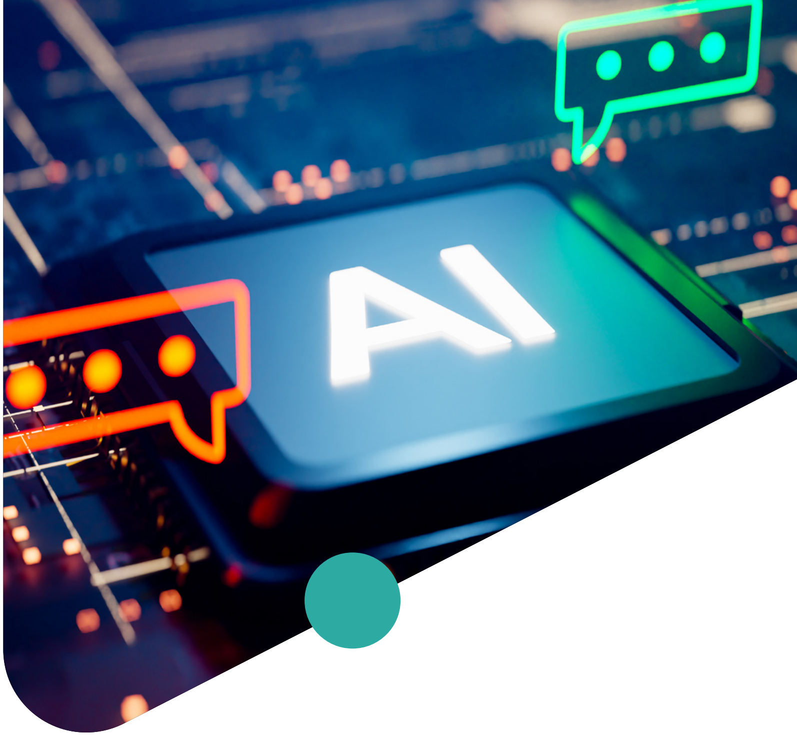
1. How vulnerable your application and selection tools are to GenAI
2. Which stages of your process are most vulnerable to Generative AI
3. Which mitigation strategy is best suited to you



Part 2

GenAI vulnerability audit: methodology

1 Analysis



A Identify if you're feeling any of the symptoms of candidate usage of ChatGPT

i. **An increase in the volume of applications.** As this video shows, there is no shortage of tools allowing candidates to apply for jobs with a single click.

ii. **An increase in applications looking similar to each other.** Linked to the above point, tools like AutoApplyAI move candidates beyond single-click applications – and allow them to mass apply for and personalise thousands of applications in one go. The difference? TA teams are seeing not only being inundated with applications, but many of the answers are looking very similar. As this LinkedIn post shows, that's not always to the benefit of the candidate when the replies are overly long, with flowery but hollow-sounding language.

iii. **An increase in candidate quality at the first sift, but a drop in quality at the interview or assessment centre stage.** Some argue that ChatGPT is a great leveller – allowing candidates from all backgrounds to complete application forms and even psychometric assessments to a level that exceeds their natural ability. The downside is that, when these candidates' true behaviours and abilities are revealed, TA teams see a marked drop in candidate quality at later interview stages. Sometimes to the extent that the interview stage or assessment centre has to be rerun.

B Compile a list of all your application forms and selection tools

Even if you're not feeling the symptoms yet, we'd recommend running the rest of the vulnerability audit anyway to ensure you understand how future-proofed your entire selection process is. To get started, you'll need to plot out every part of your process. Then you'll move on to assessing their vulnerability. Here are some common selection methods you might want to consider as your plot out your process.

Application Forms:

Video Interviews: Remote interviews, conducted via video conferencing tools to assess and interact with candidates.

Hard Skills Assessments: Tests designed to measure specific, teachable abilities or knowledge areas.

Traditional Personality or Strengths Assessments: Tools that measure personality traits to predict candidate future performance.

Logical/Abstract Reasoning Test: Evaluations of a candidate's ability to identify patterns or solve problems using non-verbal information.

Verbal Reasoning Test: Assessments of a candidate's understanding and interpretation of written information.

Situational Judgement Test: Scenarios assessing how candidates might respond to typical job-based situations.

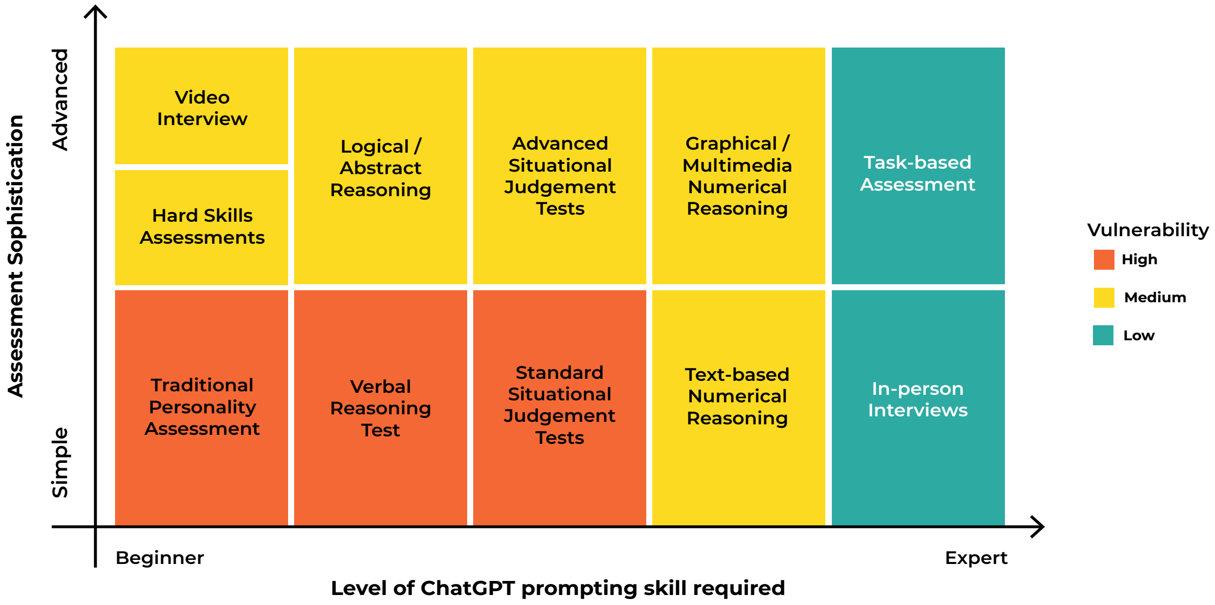
Numerical Reasoning Test: Assessments using visual data representations to test quantitative analysis skills.

Task-based Assessment: Tasks measuring either candidates' Personality fit for a role, or their Cognitive Ability — or both. The tasks are interactive and visual rather than text-based.

In-person Interviews: Face-to-face meetings to assess and discuss a candidate's suitability for a position.

Map your application and selection tools against our vulnerability matrix

This will determine which areas of your assessment process are most vulnerable to ChatGPT.



Assessment Sophistication

Advanced

Simple

Video Interview

Hard Skills Assessments

Logical / Abstract Reasoning

Advanced Situational Judgement Tests

Graphical / Multimedia Numerical Reasoning

Task-based Assessment

Traditional Personality Assessment

Verbal Reasoning Test

Standard Situational Judgement Tests

Text-based Numerical Reasoning

In-person Interviews

Beginner

Expert

Level of ChatGPT prompting skill required

Vulnerability

🟥 High

🟨 Medium

🟩 Low

ChatGPT
Capabilities
Remembers what user said earlier in the conversation
Limitations
May occasionally generate incorrect information
Allows us
Trained t

You can start to see that this approach gives you a clear view of which tools are:

- **Most sophisticated in their design**
- **Require the highest level of prompting skill to break**
- **Are most robust against GenAI**

In doing this, you're also building up a picture of which stages in your process needs to be redesigned.

PRO TIP:

To get the most out of this stage, consider assigning a score to each of the three vulnerability levels. For example:

High = 10 Medium = 5 Low = 1

Then calculate your baseline score by mapping your selection tools against the matrix and adding up each individual score.

Finally, see how replacing or removing certain tools with alternatives affects your score.

There is no magic number that we recommend here. The main thing to consider is the degree to which you can reduce your vulnerability by redesigning your process.

For example, using the scoring method above, replacing a Verbal Reasoning Test with a Task-based Assessment can reduce your score by 19 points – slashing your vulnerability to ChatGPT in a single move.

For some TA teams, this preliminary analysis will provide enough insight for you to confidently consider mitigation strategies. (See Section 3, below.)

Other teams, however, may want to sense-check this analysis, by taking a deeper dive into these vulnerabilities.

2 Optional deep dive

There are two ways to deepen your analysis from Section 1.

A Internal evaluation:

This approach allows TA professionals to sense-check their initial analysis.

First, you might want to check whether there are any parts of your application or selection process where hiring managers are reporting a mismatch in candidate quality. For example, if you're seeing a huge rise in candidates passing Personality assessments at the first sift, but then a drop in quality at the interview stage, it could indicate candidates are using ChatGPT to inflate their scores on the assessment.

Another data point could be your text-based applications. If you're receiving a much higher volume of applications - or finding that lots of the written text is very generic or lengthy - (all common signs of GenAI usage) then you might conclude that candidates are using Generative AI to complete these stages.

Alternatively, test your own application or selection process using Generative AI. Can you complete your assessments using ChatGPT or another AI tool? (If you're looking for inspiration on how to do this, check out our research series on ChatGPT vs assessment tools, which includes prompting strategies that you can use.)

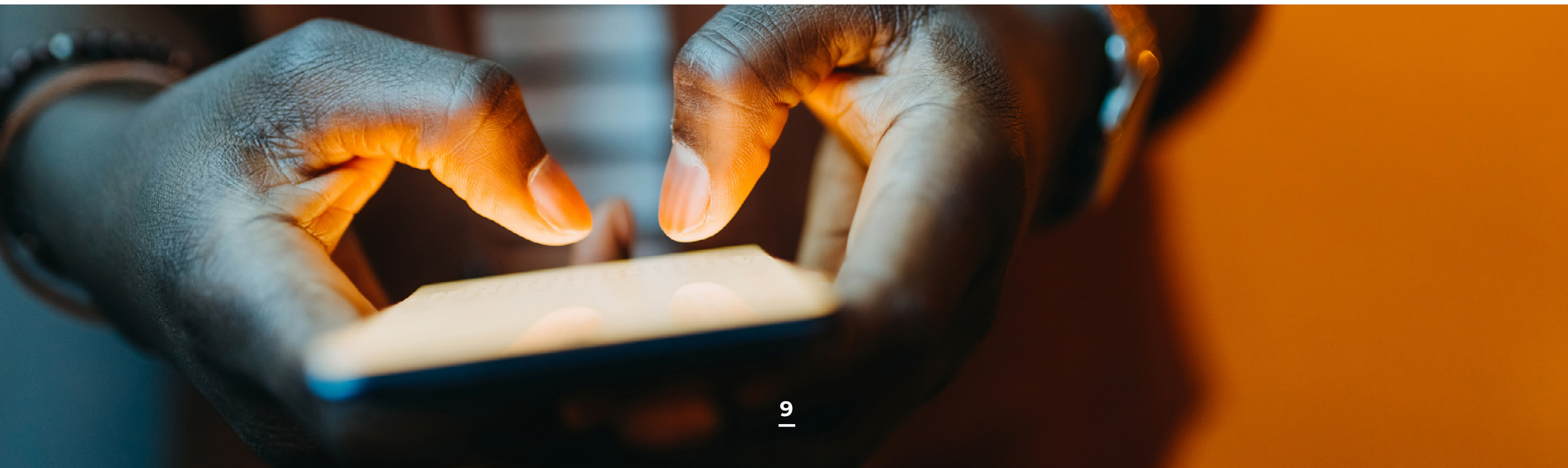
B External expert evaluation:

TA professionals who have worked in the sector for some time will know that assessments are built by psychometricians and business psychologists. From vendors with in-house specialists to external academics publishing new research breakthroughs, there's a wealth of expertise that can be called upon to understand any potential challenges with your application and selection process.

There will be a charge associated with this service, so it won't be for everyone. But for companies with limited human resources and generous budgets, this approach offers the possibility of obtaining deeper, more statistically valid feedback on potential vulnerabilities and areas for improvement.

Some companies already stepping into this space are:

- Holistic AI
- Talentspaces



3 Mitigation strategies

As detailed in the **companion blog post** there are currently three mitigation strategies to combat the influence of GenAI in your application and selection stages:

Deter

Deterring AI usage comes, unsurprisingly, from making candidates aware that they'll be negatively affected if they use Generative AI in the application or selection stages. For example, career sites have always highlighted that if an applicant is caught 'cheating', they'll be removed from the selection process. Other deterrence techniques include preventing copying and pasting or having multiple applications open. But these are easily bypassed.

Detect

Detecting candidates using Generative AI means using monitoring algorithms to 'flag' suspect responses. The main issue with this approach is that no ChatGPT detection models have been shown to work effectively as of today. Some sources even report that 2 in 1 times these detection methods produce a false positive; meaning you risk falsely accusing 20% of your candidates of cheating, potentially harming your employer brand.

Design

Design, on the other hand, involves redesigning your selection process to ensure a level playing field for your candidates while enabling you to sift efficiently and effectively at scale in the earlier stages of the recruitment process.

It also ensures that your company meets candidates' expectations of being progressive and embracing Generative AI.

This approach involves reviewing each step of your existing application and selection process. Then deciding where you're comfortable with Generative AI usage vs where you aren't.

For example, talent acquisition leaders might decide that they're comfortable with candidates using Generative AI to help them complete an application form. But want to make sure that psychometric assessments used for sifting cannot be completed using Generative AI – even if candidates want to.

In this case, the TA leader will need to look at where their current tools sit on the vulnerability matrix – and then decide whether the tools themselves need to be replaced.

If the answer is yes - say they have a question-based Verbal Reasoning Test, which is very vulnerable to ChatGPT usage, then they may look to replace it with a Task-based Assessment, which is far less vulnerable to the influence of ChatGPT.

If the answer is no - and your tools are already designed in a more robust way - then you will need to focus instead on your messaging to candidates.

For example, make it clear that you welcome the use of Generative AI at the application stage – and that the candidates' usage will be assessed. In particular, overly lengthy, unsubstantiated or convoluted answers (all hallmarks of Gen AI responses) will count against them.
Or, alternatively, their answers will be used at the interview stage.

Part 3

Conclusion

While not exhaustive, this methodology provides a structured approach for TA teams to rigorously evaluate and mitigate the threats posed by Generative AI, ensuring a fair and effective selection process.

Through a combination of technical evaluations and mitigation strategies, TA teams can better navigate the challenges and opportunities presented by Generative AI in talent acquisition.

If you have any further questions about this audit process or its implementation then please reach out to **hello@arcticshores.com** and a member of our team will be very happy to assist you.

About Arctic Shores

Arctic Shores is the market leader in hiring for potential. Our task-based assessment, powered by science, gives everyone a way to show their potential, and every employer the means to see it. Proven to counter natural bias during the recruitment process and build the diverse, successful workforce of tomorrow, our next-generation assessment widens talent pools and unearths high-quality candidates in any economic climate. We've given over two million candidates worldwide something different: a stress-free, unbiased candidate experience that truly rewards them for their time. Join the leaders in our community of pioneering customers, including Vitality, RSA, Burness Paull, TalkTalk and Siemens.

Want to learn more about how Arctic Shores could help you start hiring for potential? Get in touch with us today.

ARCTIC SHORES