


# How students' use of Generative AI will make traditional selection processes redundant



Insights and practical recommendations on how to get ahead of the seismic shift in selection that Generative AI will require TA teams to make within 12 months – backed by a survey of 2,000 students and recent graduates, as well as our internal data science-led research with UCL postgraduate researchers.



ARCTIC SHORES

# Contents

Part 1

## Why Talent Acquisition and Early Careers teams must rethink their selection process right now

**If you read only one section, make it this one.** Get the headlines on adoption of Generative AI by demographic, how ChatGPT performs vs traditional assessments, and considerations for TA teams to make now.

Part 2

## DEEP DIVE: How do 2,000 students and recent graduates use Generative AI – and feel about potential employers' attitudes to it?

Dive deeper into usage of Generative AI in the selection process, whether students think they should be allowed to use Generative AI to help them apply for roles, and their views on proctoring and in-person interviews.

Part 3

## How good is ChatGPT at completing psychometric assessments and how worried should TA leaders be?

Get an overview of ChatGPT's performance on traditional Aptitude tests, SJTs, Personality Assessments, and Task-based Assessments, plus how easy it is to use ChatGPT to outperform the average candidate with little to no specialist training.

Part 4

## What happens next? Next steps for progressive TA leaders adapting their selection process

Understand the practical steps you can take now to get ahead of the seismic shift in selection that ChatGPT will require within the next 12 months, including an overview of how to think about auditing your processes before redesigning them.



Part 1

**Why Talent Acquisition  
and Early Careers teams  
must rethink the  
selection process  
right now**

# The headlines

Introduction from  
Robert Newry, CEO & Co-founder  
of Arctic Shores



If you don't know much about ChatGPT or aren't sure its impact will be as significant as the apparent hyperbole and media frenzy are predicting, think again.

This report summarises recent Arctic Shores research combined with a survey of more than **2,000 students and recent graduates**, and the steps Talent Acquisition teams and Early Careers leaders should take next based on the findings.

---

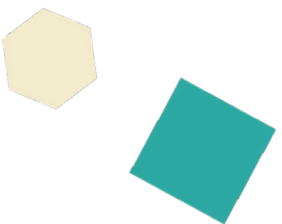
## 72% of students and candidates are using some form of generative AI on a regular basis

Our findings also reveal that the likes of ChatGPT and other Generative AI tools **are already fundamentally changing** how candidates — and specifically the next generation of Early Careers applicants — apply for jobs.

What's more, Arctic Shores research — conducted with UCL postgraduate researchers — shows that the latest, paid version of ChatGPT, GPT-4.0, **outperforms 98.8% of candidates in verbal reasoning tests.**

It also scores **in the 70th percentile on Situational Judgement Tests** (the typical threshold to pass to the next stage of the recruitment process). And in the other standard assessment type, the Personality Assessment, **ChatGPT's inherent persona scores highly for many of the desirable traits TA teams typically select for.**

None of this data is surprising when considering that ChatGPT has been shown to have an IQ of 152 (higher than the score you need to access to MENSA which sits at 150) and researchers at Microsoft tasked with determining whether it could actually reason described it as “much smarter than the average human”.



## What's different about Generative AI vs other methods candidates have used to gain an advantage in the past?

There have always been tools and practice sites students could use to enhance their applications over the years. But the cat and mouse game of gaining an advantage has been largely contained through 'Deter and Detect' techniques and, most importantly, the fact that access to practice sites comes at a cost — preventing most candidates without the means or inclination to pay for them. ChatGPT shatters this set-up.

ChatGPT is not some geeky, specialist tool that only a few will master. With its chatbot design, human leveraged reasoning, and ability to teach itself faster than the average human, it has already become a ubiquitous tool used en masse by students with little to no specialist training.

Our survey of 2,000 students and graduates found that **72% are now using Generative AI tools on a regular basis – a number that has increased exponentially in just four months** (an earlier ISE study from [Cibyl](#) reported that 51% of students were using ChatGPT in May 2023). Our survey results showed that as well as using ChatGPT, many are also branching out to other tools like Google's Bard, Wolfram Alpha, and Whisper AI.

**Generative AI is, in many ways, the language equivalent to a calculator — and as an increasing number of candidates realise its value, we can expect that the depth and breadth to which it gets used will also continue to expand.**



# How does this affect TA professionals?



Our research presented several eye-popping insights with huge implications, particularly for volume screening in Early Careers.

**While 7 in 10 respondents say they would consider using these tools to complete an application or assessment in the next 12 months, almost a fifth of candidates are already using Generative AI to help them fill in job applications or assessments.**

**This rises to nearly a quarter for Black students and those with a mixed ethnic background** – a finding consistent with an earlier study this year from [ISE](#).

The key insight is that from the perspective of **almost half of students and recent graduates, Generative AI is a positive tool which levels the playing field.** But the reality is that while it levels the playing field for some, it has the potential to disadvantage others — specifically those without the financial means to pay for the premium version of ChatGPT.

ChatGPT-4, the current premium version which is behind a paywall, significantly outperforms its free sibling, ChatGPT-3.5. Only 15% of students who use ChatGPT pay for it — and while 54% would consider paying for it in the future if they knew it would help them complete a job application, cost is still a barrier. **For 38% of respondents, ChatGPT premium is simply too expensive.**

# The ethics of Generative AI usage in selection

ChatGPT usage among the student and graduate population is here to stay.

**And students expect to be able to use it at work and in the selection process.**

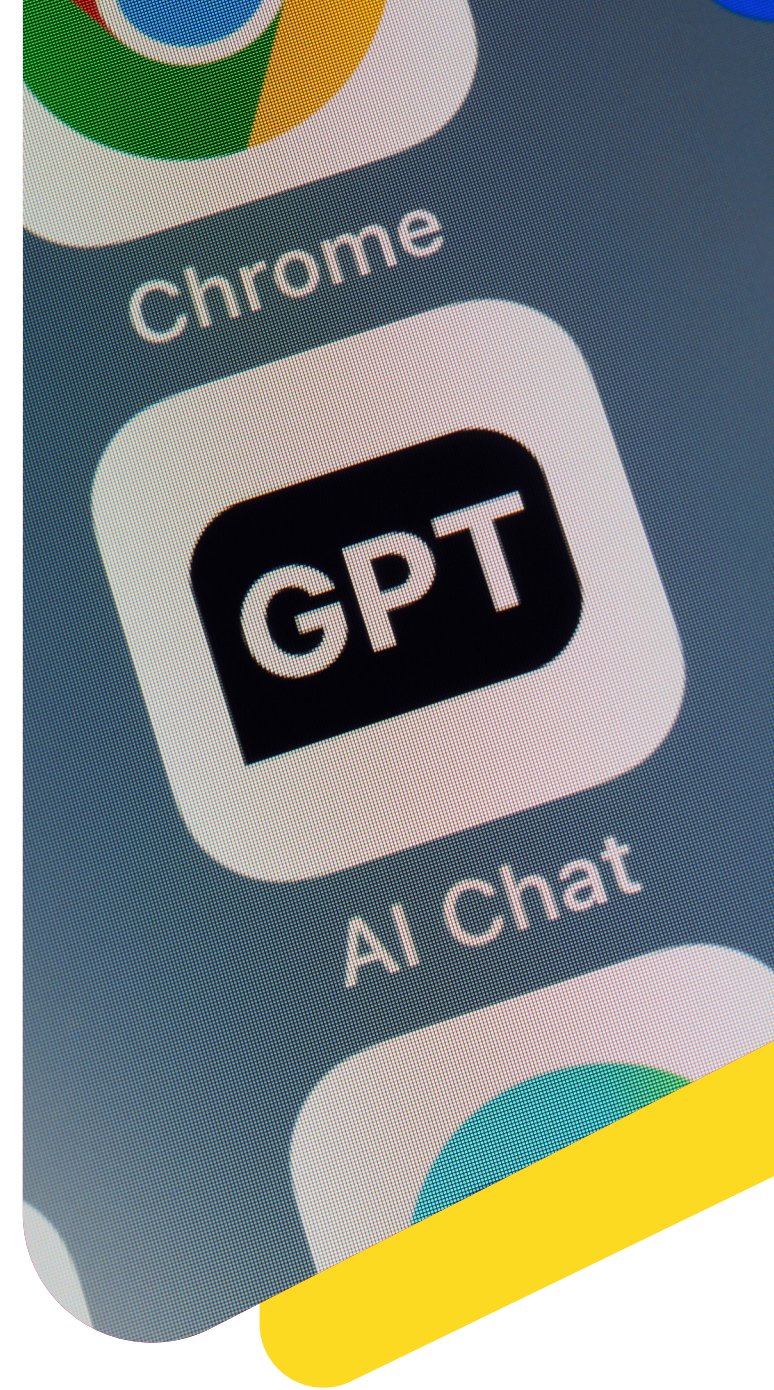
In fact, students are already using Generative AI for an average of an hour and 14 minutes a week, and that rises to an hour and 24 minutes for neurodiverse candidates — whose usage is more likely to include advanced Generative AI tools like Midjourney and Wolfram Alpha.

Almost **half (47%) of students and recent graduates believe employers should allow them to use Generative AI** as part of the selection process. And the majority may not even think about whether or not doing so would be considered cheating, with only 13% of candidates believing that usage of Generative AI in an application or assessment would be dishonest.

Our early careers population are going to be using Generative AI whether we ask them to or not — **so this puts Early Careers teams into an ethical conundrum. Should you accept candidates' use of Generative AI or strongly discourage its use in the application process?** Either way, saying nothing will leave students concerned and they will soon expect employers to have a firm position.

The starting point for every employer should be to recognise its wide use and welcome students to use Generative AI, not least to support those from underrepresented groups who use it as a means to reduce the disadvantage in their backgrounds. The challenge is that those with the financial means will soon discover the significant difference in performance between ChatGPT's free and paid versions.

TA leaders now need to think carefully about which stages in the selection process they want to encourage Generative AI usage and where to prevent it to ensure a truly level playing field for all candidates. Failing to work through these issues could set back all the progressive and positive social mobility efforts that so many employers have invested in by years.



# So what's the alternative?

**Many traditional assessment providers and employers say that the best option here is to Deter and Detect the use of Generative AI. But it's not quite as clear cut as that.**

The education sector started with a ban of Generative AI but quickly realised this was not the sensible route forward. Not least because AI detection programmes were poor at detecting AI use (Open AI shut down its own detection tool due to its poor accuracy and far too high number of false positives; studies also show that **2 in 10 times these detection methods produce a false positive**). And that's without considering Generative AI's higher usage among underrepresented groups.

Those TA leaders who don't act now to rethink their selection process potentially have some big challenges on their hands: from making already overworked recruitment teams even more stretched by having to go back to manually screening candidates at the start of the process because of higher pass rates, to candidates from underrepresented groups being unfairly penalised, to allowing candidates with the financial means to inflate their true potential above their peers by using ChatGPT-4.0.

---

**The obvious and logical answer is not simply to Deter or Detect AI usage, but to redesign your selection process in a way that recognises and embraces the use of Generative AI without undermining the effectiveness or accuracy of your recruitment methods.**

This report dives deep into the insights from our survey of students and recent graduates, as well as ChatGPT's performance on traditional Aptitude and Situational Judgement Tests, traditional Personality Assessments, and Task-based Assessments. We also share insights on the main alternatives for TA teams to consider and suggestions on what they can do next. We hope you find the insights as illuminating and valuable as we did.



Part 2

**DEEP DIVE:**  
How do 2,000 students  
and recent graduates use  
Generative AI — and feel  
about potential  
employers' attitudes  
to it?

# About the sample

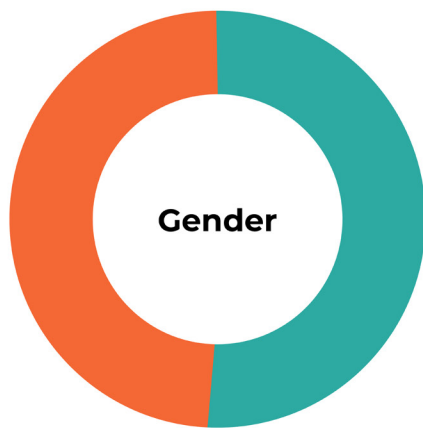
[Early research](#) from the Institute of Student Employers and Cibyl in May 2023 showed that more than half of students would use ChatGPT in the application process. We wanted to go deeper. And we wanted to see how quickly adoption was changing among the student and graduate population.

Our September 2023 survey, conducted by Opinium, asked 2,000 students and recent graduates (graduating in the last two years) about their adoption of Generative AI tools and their perception of employers' attitudes towards Generative AI too. We also wanted to see how this varied by demographic.

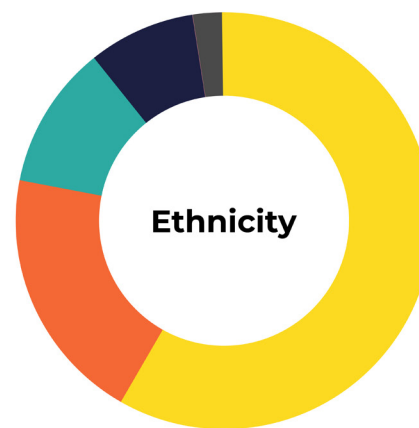
Without this data, we felt it would be difficult to understand how TA teams' approaches to hiring would need to change to maintain both an inclusive process in the age of Generative AI and a favourable candidate experience.

---

## Demographics of survey respondents

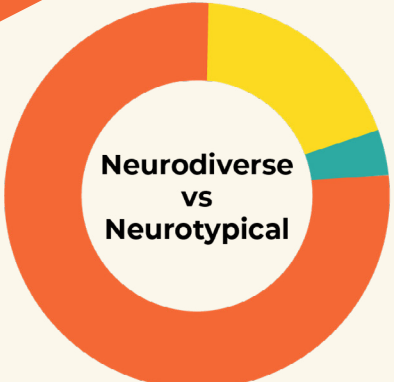


Female - 49%  
Male - 51%



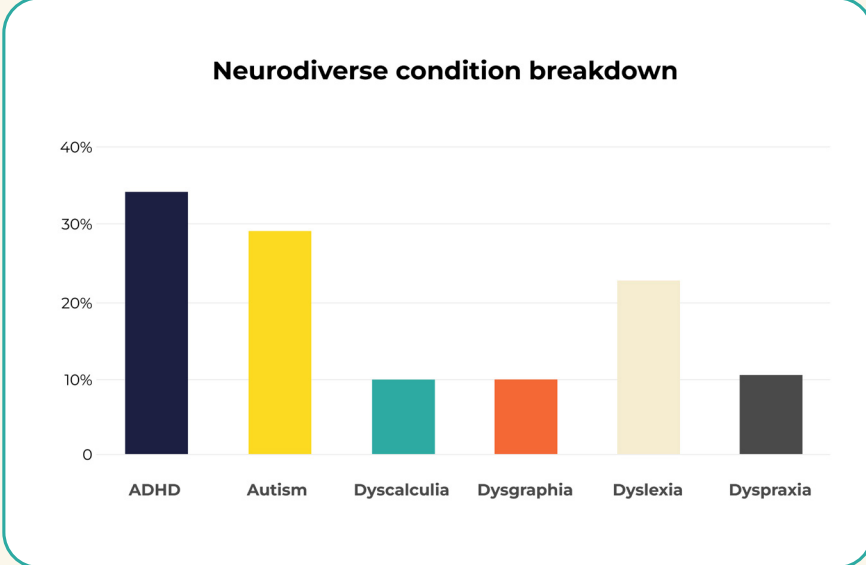
White or White British - 58.2%  
Asian or Asian British - 19.8%  
Black, Black British, Caribbean or African - 11.2%  
Mixed ethnic background - 8.6%  
Other - 2.1%

# Demographics of survey respondents



Q: Do you have a neurodiverse condition, specific learning difficulty, and/or any disability?

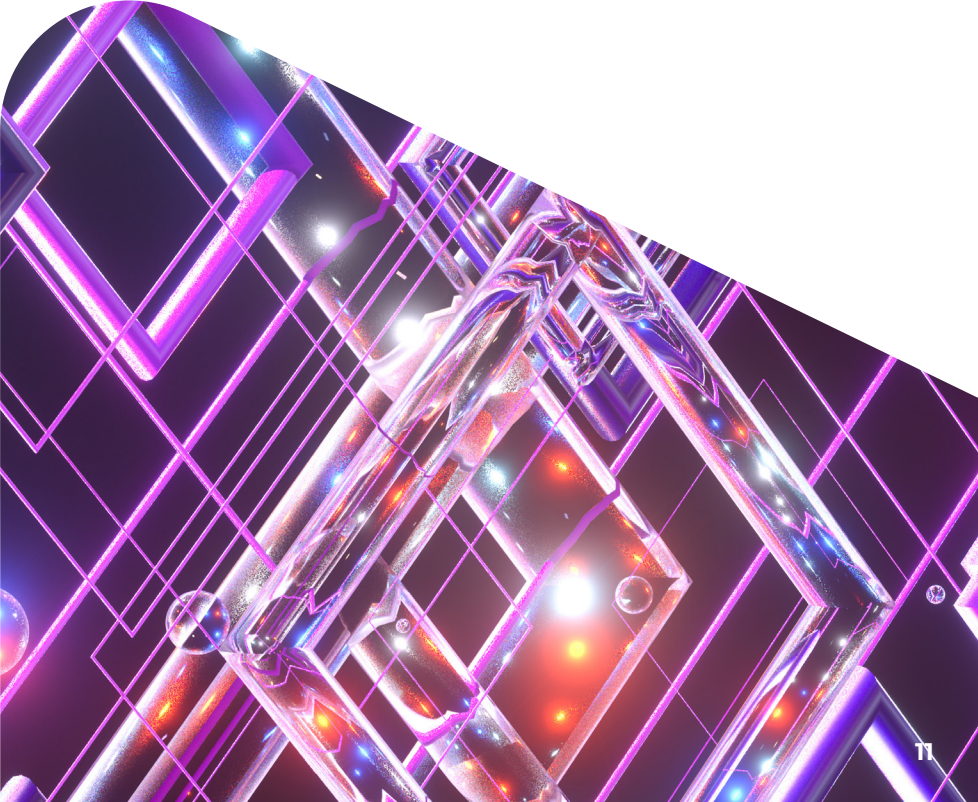
- No - 76%
- Yes - 20%
- Prefer not to say - 3%



Q: What is your annual household income?

- Up to £20,000 a year - 24.1%
- £20,001 - £40,000 a year - 30.3%
- £40,001 - £60,000 a year - 22.1%
- £60,001 - £80,000 a year - 12.3%
- £80,001 - £120,000+ a year - 11.2%

82% of our survey respondents have applied for a job in the last 12 months, and **81% expect to apply for a job in the next 12 months**. Based on that data, we're confident that this report is representative of job-seeking early careers candidates in 2023.



# Adoption is rising rapidly: 72% of students and recent graduates are using Generative AI tools as the norm

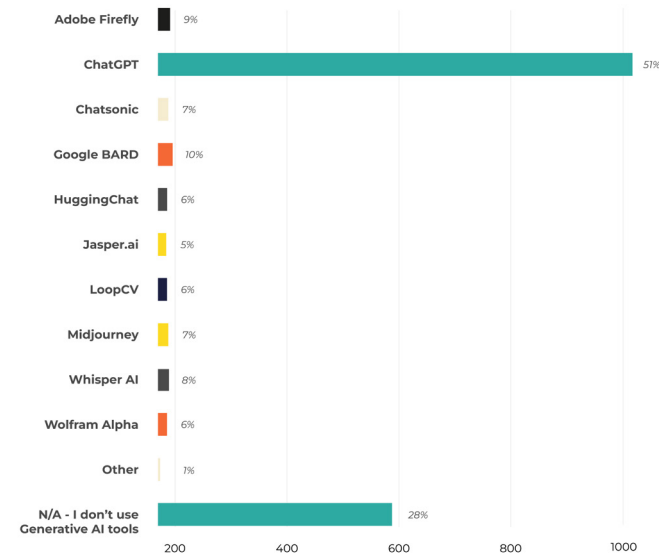
If TA Leaders and Early Careers Specialists were not sure if the adoption of tools like ChatGPT would affect them, our survey data shows that it will — without a shadow of a doubt.

It may come as no surprise that the student and recent graduate population have dabbled with new technology, but perhaps what may be more surprising is the extent to which usage of these tools is fast becoming a weekly — or even daily — routine.

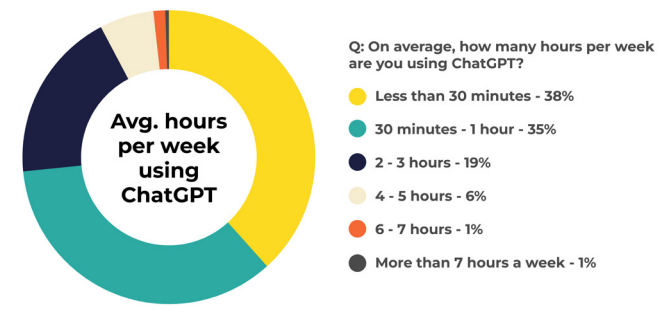
**72%** of students and recent graduates today are using Generative AI tools as the norm — **a meteoric rise in adoption** given ChatGPT only entered mainstream consciousness **back in December 2022**.

While **51%** of students and recent graduates are using ChatGPT — arguably the most commonly known Generative AI tool — **21%** are being much more experimental, expanding their usage into image generation tools, copywriting tools, and even solving complex mathematical and scientific problems.

Which, if any, of the following Generative AI tools do you use?



Many Early Careers candidates are already using this technology every week. The average candidate using ChatGPT for an hour and 14 minutes per week, while 1 in 3 are using it for more than 2 hours (and up to 7).

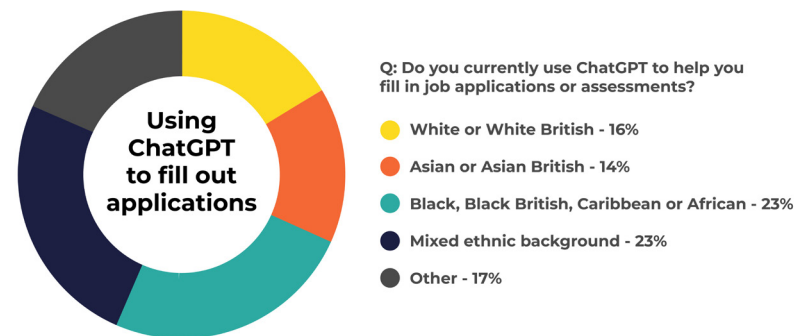


# 17% of students and recent graduates are already using ChatGPT in the selection process but 7 in 10 aren't far behind

While the number of candidates currently using ChatGPT to apply for jobs and complete assessments isn't huge, perhaps what's more of a cause for concern for Early Careers teams is that 7 in 10 candidates say they would consider using ChatGPT or other Generative AI tools for applications and assessments in the future.

**If that happens, TA teams could find that 70% of Early Careers applicants are using Generative AI as part of the application or assessment process.**

## Differences in demographic usage



**Black and Mixed ethnic background students and graduates are more likely to use ChatGPT to help with job applications and assessments (both 23%).**

# A third of candidates would be put off applying for a role if told they couldn't use Gen AI

**Just 13% of candidates said they wouldn't want to use ChatGPT in the application process because they would feel dishonest** — that means the vast majority (87%) of candidates would have no qualms at all about using ChatGPT in the selection process.

And in fact, many candidates feel as though it's their right to do so too.

**Almost half of the 2,000 Early Careers candidates surveyed believe prospective employers should allow them to use ChatGPT to improve their chances when applying for a job.** A third would not want to work for an employer who told them they couldn't use Generative AI, another third would question if they wanted to work for them, and 38% would believe the employer wasn't very progressive.

This means the tone that employers set around Generative AI usage has huge implications for their employer brand and whether or not they shrink their talent pool.



- A third would not want to work for an employer who told them they couldn't use Gen AI
- A third would question whether they wanted to work for an employer who told them they couldn't use Gen AI
- A third would view an employer as not very progressive if they told them they couldn't use Gen AI

## ▶ Is the answer then to permit candidates to use Generative AI as part of the application process?

It might not be quite as straightforward as that. Many application processes (especially in Early Careers) start with a set of application questions ('Tell us why you would like to work with us' or 'What would you bring to the role') - if these all start to look the same or meet the scoring criteria, this may no longer be an effective screening tool.

TA teams will be inundated with more candidates moving through to the next stage, with little evidence of who is genuinely appropriate and who is simply good at using Gen AI.

**The cost and efficiency of screening methods, crucial to any volume hiring programme, could be significantly undermined.**

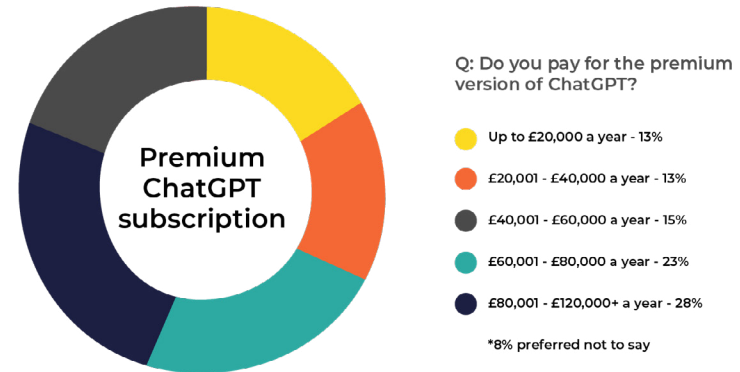
# 15% of candidates currently pay for ChatGPT-4 but 38% are priced out

There are big differences between the two ChatGPT models in terms of performance, with the premium version achieving much higher scores on traditional assessments than its free counterpart (more detail on this later in the report).

	ChatGPT-3.5	ChatGPT-4
<b>Performance</b>	Highly advanced, but less accurate responses and occasional 'hallucinations' (the term used to describe a Gen AI answer which is invented).	Offers improved accuracy, consistency, and can understand context even better. This leads to more relevant answers.
<b>Training Data</b>	Trained on vast amounts of text, but limited to data up to 2021.	Incorporates up to 20 times more data, leading to vastly more diverse datasets — around ten times larger — which enhances its relevance and knowledge base.

While just 15% of users pay for ChatGPT premium at £15 a month, 54% would consider doing so if they knew it would help them perform better in an online assessment as part of a job application. But cost is a major factor for others.

For 38% of candidates, ChatGPT Premium is currently too expensive — meaning that students who have the privilege of paying for ChatGPT will have an advantage over peers who don't have the money to spare. It doesn't take a huge leap to work out that this could have major implications for social mobility, undermining years of effort.



66% of those who pay for Premium come from a household with an income of over £40,000

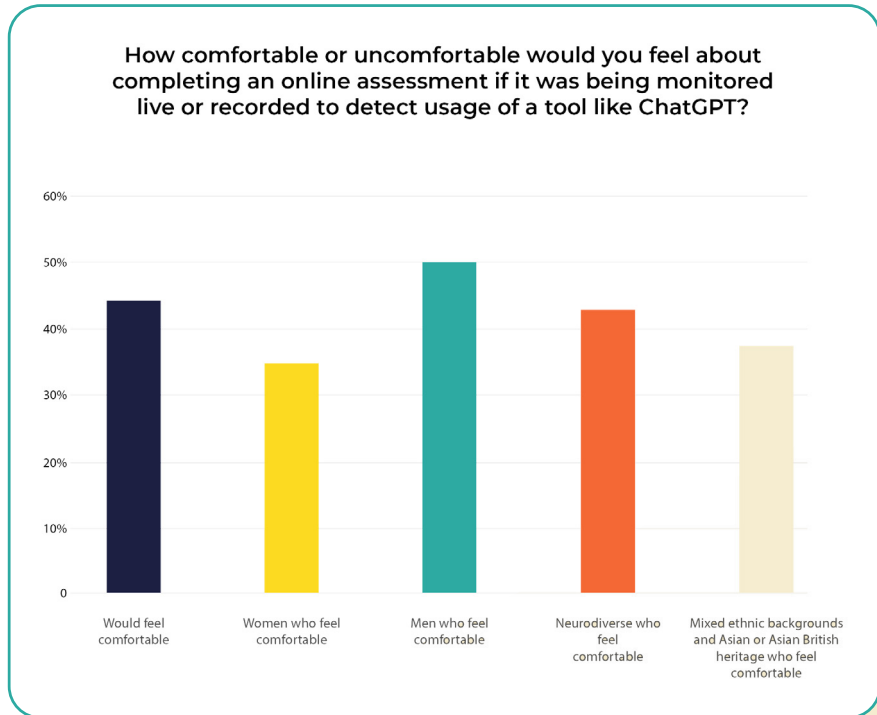


# How would candidates feel if an employer banned use of Generative AI in the application process?

Some Early Careers teams using traditional question-based psychometric assessments may have been told by their assessment providers that the best options available to them are now to Deter or Detect Generative AI usage, so we wanted to understand more about how these methods might be perceived by applicants. In most cases we found employers taking a **Deter and Detect approach would unintentionally harm workforce diversity** and the candidate experience.

## Proctoring

While the idea of a proctored online assessment may seem like the perfect solution for deterring Generative AI usage, doing so may exclude candidates from underrepresented groups – making it even harder to hit your diversity goals.

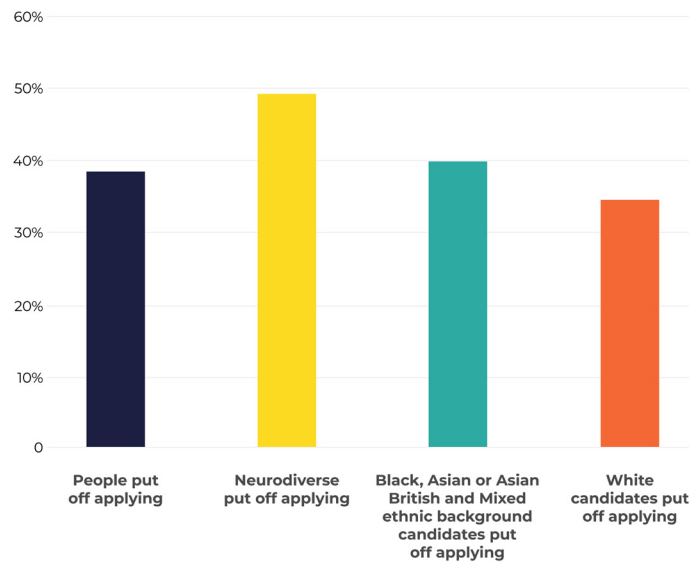




## Going back to an in-person assessment process

Aside from this being unscalable, going back to an in-person or live application screening process is also not a viable option based on how candidates feel about it.

To what extent do you agree or disagree with the following statement:  
I would be put off applying to a job if I had to do the first interview / first assessment in person



▶ This leads us to conclude that Detering or Detecting candidate usage of ChatGPT is not an option if we want to continue to improve workforce diversity and deliver an exceptional candidate experience.



# Summary: Student and recent graduate adoption and perceptions of Generative AI

## Adoption is rising rapidly

 **72%**

of students and candidates are already using some form of generative AI regularly

---

## Candidates expect to use Generative AI

 **13%**

a third would be put off applying for a role if they couldn't use Generative AI and only 13% believe using it is dishonest

---

## Generative AI is being used in applications

 **1/5**

almost a fifth of candidates are already using Generative AI to help them complete job applications or assessments

---

## It levels the playing field for some

 **23%**

of Black candidates and those with an ethnic background are more likely to use ChatGPT to help with applications and assessments (both 23%)

## But not for others

 **15%**

of candidates currently pay for ChatGPT but 38% are priced out; 66% of paying members are from households with an income of +£40,000

Our early careers population are going to be using Generative AI whether we ask them to or not. This puts Early Careers teams into an ethical conundrum – should you accept candidates' use of Generative AI or strongly discourage its use in the application process? Either way, students and graduates will soon expect employers to have a firm position. And doing nothing is a risky option.

**Part 3**

**How good is  
ChatGPT at completing  
psychometric assessments  
and how worried should  
TA leaders be?**



# Setting the scene on psychometric assessments

As we've established so far, almost a fifth of candidates are *already* using ChatGPT in the selection process. And for those who aren't, there's a strong possibility that 7 in 10 will do so in the near future.

This leads us to ask the question — if they were to try use a tool like ChatGPT to complete a psychometric assessment, would it actually make a material difference to their performance?

**For the purpose of this research, we have focused on the two key psychometric assessment formats.**

Psychometric assessment formats	
<b>Traditional</b>	These are typically text-based and question-based. And while question-based assessments have been subject to criticism for years now thanks to their questionable accuracy, cheatability, and potential to create differences in performance for disadvantaged groups, our research has shown the emergence of Generative AI tools like ChatGPT is throwing their suitability into question even more than before. It's worth noting that some question-based psychometric assessments might be 'gamified' but still fundamentally use a question format.
<b>Task-based</b>	These are a newer breed and use a series of interactive tasks to measure a person's Personality and Workplace Intelligence (Aptitude plus Emotional Intelligence) – they are focused on actions not text. There have been fewer questions about the suitability of this assessment type in the age of ChatGPT because they aren't based on 'language'. However, with the rise of image recognition AI models, that's not a guarantee.

The next chapter of this report will lay out the headlines from research conducted by Arctic Shores' Senior Data Scientist, and two UCL postgraduate researchers.

The research squad conducted a series of rigorous, systematic studies across the four main psychometric assessment types – Aptitude Tests, Situational Judgement Tests, Personality Assessments, and Task-based Assessments. Their goal was to answer the following two fundamental questions:

- ? Can candidates use Generative AI models like ChatGPT to complete psychometric assessments and tests, and outperform the average candidate?
- ? Can they do this with little or no specialist training?

# ChatGPT vs Aptitude Tests

## About Aptitude Tests

According to research, one of the best ways to predict candidates' future job performance is to measure their cognitive ability. And the best way to measure cognitive ability has historically been with a traditional, question-based Aptitude Test. There are two commonly used types of Aptitude test – Verbal Reasoning and Numerical Reasoning.

## Verbal Reasoning Tests:

### What the research showed

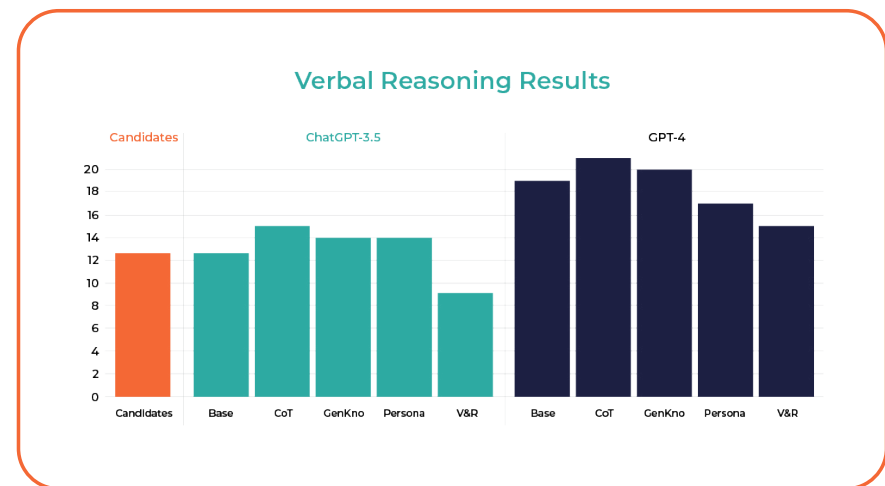
We used various 'prompting styles' (or ways of asking questions) to explain the questions to ChatGPT starting with very basic (simply copying and pasting the instructions from the Aptitude test into ChatGPT) to slightly more advanced (asking it to imagine it was an analyst or explain its thinking).

### Here's what we found:

With almost every prompting style and across free and paid versions, **candidates could achieve above-average scores on Verbal Reasoning Tests.**

- GPT-4, which is behind a paywall, scored higher than **98.8% of all candidates across a sample size of 36,000 people** – potentially setting back social mobility work years.
- Candidates can score **almost double** the human average if they can afford to pay for ChatGPT-4.

*ChatGPT outperforms the average candidate on a Verbal Reasoning Test. These are its scores across five different prompting styles.*



# ChatGPT vs Aptitude Tests

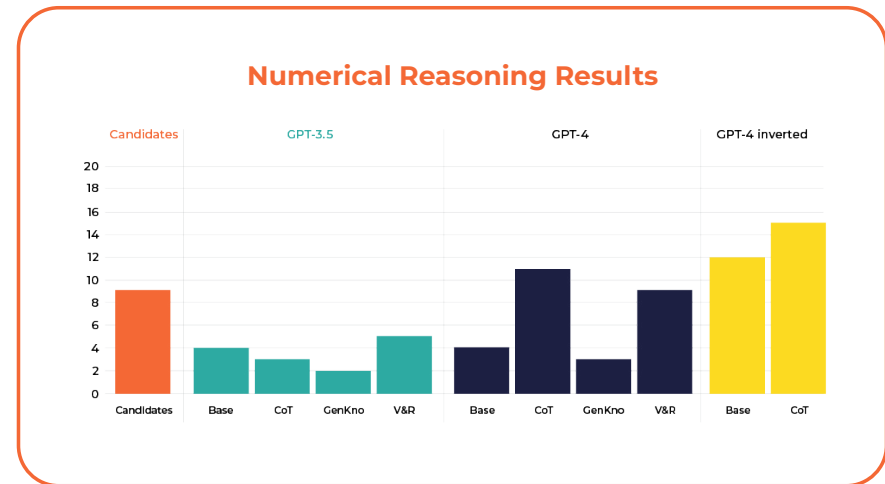
## Numerical Reasoning Tests:

### What the research showed

While ChatGPT is great at basic calculations, our researchers discovered that it's not great at complex Numerical Reasoning Tests — only outperforming the average candidate using one prompting style and using the paid version of ChatGPT.

This is not surprising, given the AI model's focus on language as it's main way of understanding and interacting with the world.

ChatGPT cannot consistently outperform the average candidate on a Numerical Reasoning Test. These are its scores across 5 different prompting styles.



### The implications

These findings pose a huge threat to the accuracy of question-based Verbal Reasoning Aptitude Tests as a sifting and selection method. The consequences could be as follows:

- 1 ▶ **An increase in Aptitude Test pass rates, but a drop in candidate quality** leading to wasted recruiter time and potentially putting hiring manager trust at risk.
- 2 ▶ A huge gap in pass rates between candidates with the financial means to pay for ChatGPT-4 vs those who can't afford it, **setting social mobility efforts back years.**
- 3 ▶ **A need to rethink which cognitive abilities we really need to test for** — if ChatGPT can complete this level of reasoning and score highly, do we need candidates to have this reasoning capability anyway? Or is it more important we test for learning agility and resilience to identify how likely candidate is to embrace this kind of technology?

# ChatGPT vs Situational Judgement Tests

## About Situational Judgement Tests

Situational Judgement Tests (SJTs, for short) typically assess things like a person's decision-making skills, organisation and planning, resilience, and communication skills. They do this by presenting hypothetical, job-related situations, and asking the test-taker to choose the most appropriate action from a set of multiple choice questions.

When scoring an SJT, candidates might be asked to rank multiple choice options from 'best to worst' or 'most likely to least likely', or they might just be asked to select 'the best' vs 'the worst' option. Either way, a test provider will always be looking for — and awarding points — based on a 'desirable' answer.

## Situational Judgement Tests:

### What the research showed

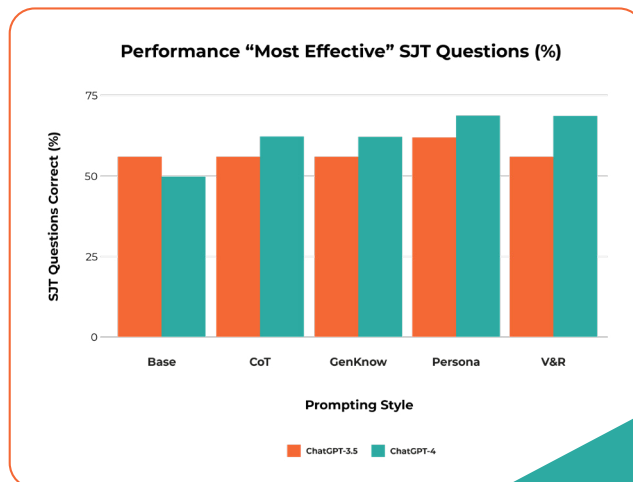
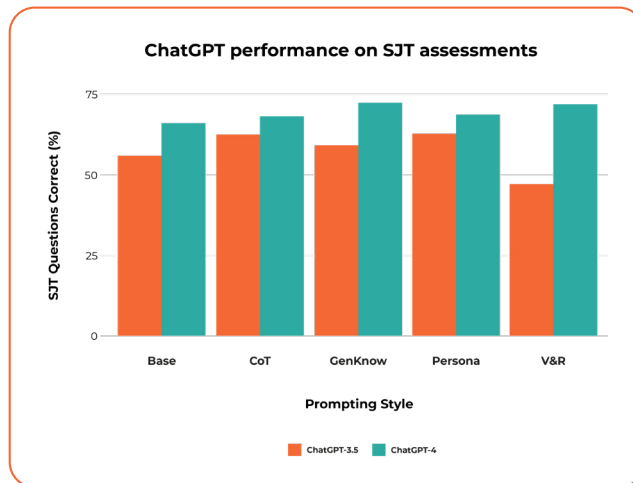
Some SJT providers have argued that ChatGPT is only good at getting 'binary right or wrong answers' correct. They state that because SJT's scoring systems don't always have a 'binary right or wrong answer', but instead rely on awarding points based on ranking, ChatGPT cannot handle that level of complexity.

### Our research showed that ChatGPT performs well on SJTs, even with non-binary scoring

- Candidates with minimal training can now use the free version of ChatGPT (3.5) to **achieve scores above the human average** on a Situational Judgement Test, getting 50–60% of the answers correct.
- Those who can afford to pay for ChatGPT-4 can **achieve scores much higher than the average candidate on a Situational Judgement Test, getting 65-75% of the answers correct.**
- ChatGPT-4 scores in the **70th percentile in comparison to the average test-taker's scores** (the typical cut off point).
- Both versions of ChatGPT performed well when selecting the 'most effective' answer. **But even if ChatGPT-4 didn't pick the 'most effective' or 'least effective' answer, it picked the next best one in most cases.**
- This leads us to conclude that while ChatGPT does perform 'better' if there is a binary right or wrong answer, **ChatGPT-4 still performs very well in a more nuanced context — even if being asked to give a solution to a problem with a 'rank order'.**

# ChatGPT vs Situational Judgement Tests

[Dive deeper into the research methodology and findings here](#)



## The implications

We can no longer rely on question-based Situational Judgement Tests as an accurate selection method. And just as with Aptitude Tests, the implications are likely to be:

- ▶ **An increase in SJT pass rates, but a drop in candidate quality** leading to wasted recruiter time and potentially putting hiring manager trust at risk.
- ▶ **A gap in pass rates between candidates with the financial means** to pay for ChatGPT-4 vs those who can't afford it, setting social mobility efforts back years.
- ▶ **A need to re-evaluate the purpose an SJT really serves in the selection process.** If ChatGPT can make judgement calls as nuanced as a human, does an SJT really tell us about a candidate's ability to succeed in the modern workplace? Or is there a better way to assess a person's resilience, decision-making and communication skills?



# ChatGPT vs Personality Assessments

Personality Assessments help employers identify how well suited a candidate's soft skills are to a particular role and to an organisation's culture. They assess a person's natural personality traits, behaviours, and preferences. They're often used alongside Aptitude Tests to give a hiring manager a comprehensive insight into the candidate's potential to succeed in a role.

While many people may be familiar with Myers Briggs or DISC profiles, both commonly used in the employee lifecycle, the majority of traditional psychometric assessments used in selection use the 'Big 5 Personality' or 'OCEAN' Model. The Big Five Model is widely accepted in the academic community and is based on decades worth of meta-data and analysis.

The OCEAN model assesses how a person's personality and temperament sit on a scale across the following top-level factors (groups under which all the individual traits can be organised):

- ✓ Openness to experience

---

- ✓ Conscientiousness

---

- ✓ Extraversion

---

- ✓ Agreeableness

---

- ✓ Neuroticism

A [study of 634 firms](#) (Wehner, de Grip, Pfeifer 2022) revealed that being Agreeable and Conscientious increases the probability a candidate will be hired. For analytical tasks, recruiters prefer more Open and Conscientious workers. While for interactive tasks, recruiters favour more Open, Extraverted, and Agreeable workers.

## Personality Assessments:

### What the research showed

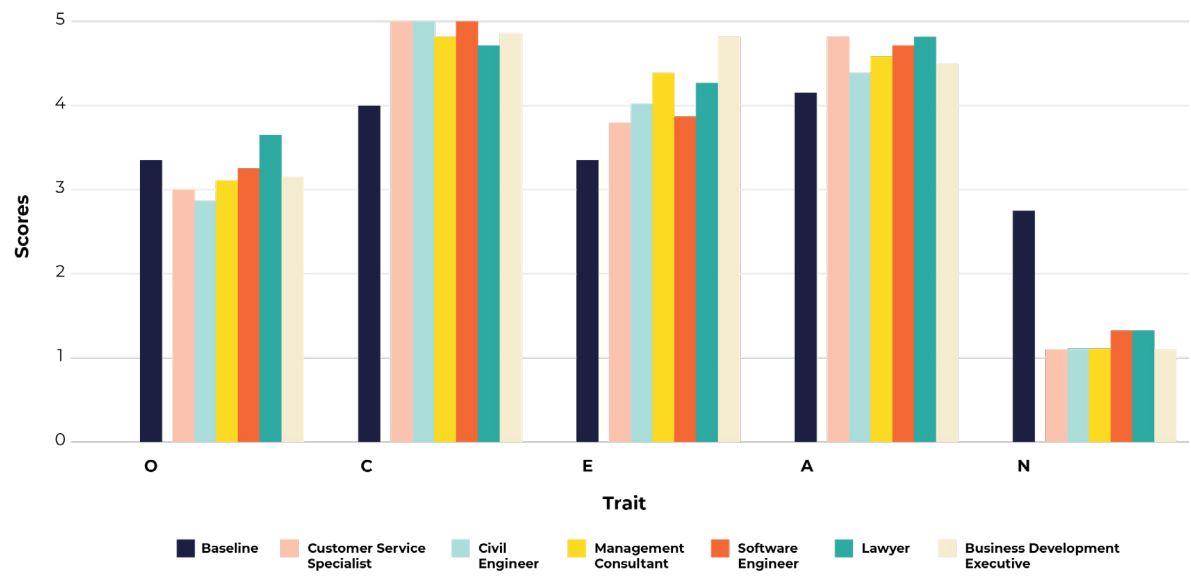
To determine whether ChatGPT could be used to 'game' the accuracy of Personality Assessments, we first needed to understand whether ChatGPT exhibits a Personality in the first place, how closely aligned it is to the most commonly desirable traits employers look for, and then how easy it would be to ask ChatGPT to change the traits it exhibits very quickly.

### Here's what we found:

- **ChatGPT's persona scores highly for many socially desirable traits across free and paid models** — it's highly Agreeable and Conscientious (a proxy for things like being co-operative and organised), but it scores lower on Neuroticism (which means it will respond in a resilient and emotionally stable way) – all traits that could make it the ideal employee.
- Given these are often the most desirable traits employers look for, **if candidates use ChatGPT to help them complete a Personality Assessment, they will score highly in the desirable traits required for many job roles without any specialist prompting.**
- **Candidates could perform even better by simply giving ChatGPT a job description, and asking it to tailor its responses to a Personality Assessment based on the desirable traits for that role** – for example, after showing ChatGPT a job description for a Business Development role, it increases Extraversion. In our experiment, it was able to tailor the traits to each of the six different roles we tested.

# ChatGPT vs Personality Assessments

Traits by job type



ChatGPT can easily adjust answers on Personality Assessments based on a job description, dialling up some traits and dialling down others

## The implications

If the adoption of ChatGPT to complete a Personality Assessment becomes widespread, we could expect to see consequences such as:

- 1 ▶ **Question-based Personality Assessments ceasing to be a useful sifting tool** if a large proportion of candidates are suddenly able to achieve desirable scores, meaning recruiters have to find new, more scalable ways to sift.
- 2 ▶ **Candidates making it through to the interview stage or even getting hired, even though they may not have suitable personality alignment with a role**, leading to wasted recruiter time, a drop in hiring manager trust, and even higher attrition rates.
- 3 ▶ **TA leaders needing to re-evaluate the usefulness of question-based Personality Assessments** in the selection process and rethink how they capture a more accurate and authentic view of a candidate's soft-skills.

# ChatGPT vs Task-based Assessment

## About Task-based Assessment

Sometimes when people talk about Task-based Assessment they mean two different things. We categorise them as follows:

### Task-based Assessments

<p><b>Gamified question-based</b></p>	<p>Some traditional question-based assessment providers have gamified the question-based format in a bid to make it feel more engaging for a candidate. Because these assessments still use text and language to guide a candidate through the journey, we believe that the research we have outlined on ChatGPT vs Aptitude, SJTs, and Personality Assessments is still relevant to them.</p>
<p><b>Task-based</b></p>	<p>This is what this next chapter of our research will cover. Task-based Assessments might feel like a game to candidates, but instead, enable Talent Acquisition teams and hiring managers to capture 1000s of data points on a candidate's potential to succeed in a role based on their actions, not their words.</p> <p>The task-based assessment format can be applied both to Aptitude and Personality. In the spirit of full transparency Arctic Shores classifies itself as a task-based psychometric assessment provider. We assess Personality and Workplace Intelligence (our more expansive interpretation of Aptitude that covers things like learning agility and facets of emotional intelligence).</p> <p>Task-based Assessments have no right or wrong answer and do not have binary scoring. Instead, they observe how a person responds to a situation to assess things like their risk appetite, how they learn and course correct, and how they perceive emotions.</p>



# ChatGPT vs Task-based Assessment

[Dive deeper into the research methodology](#)

## Task-based Assessment:

### What the research showed

To really stress test our own Task-based Assessment, we had to get more creative than when assessing ChatGPT vs question-based assessments.

#### Here's what we found:

- **ChatGPT could not complete a Task-based Assessment even when translating the instructions using an image-to-text scan in the ChatGPT App** – the only option available was to describe what the image was showing to ChatGPT and see if it could respond. However because many tasks assess reaction time and because there is no right or wrong way to approach a task, this process took a long time and did not improve candidate scores in any way.
- **Image recognition software does not help candidates in completing this assessment type** — we tried uploading images of the tasks to Google Bard to ask it to make a recommendation on how to approach them but it struggled to understand the tasks or make any useful recommendations on how to approach them, meaning that currently, this was not a useful tool.

- **We tried to ask ChatGPT to write a bot to complete a task but this did not work** — the results were very poor and even using more complex Generative AI tools to do this did not yield better results.
- **ChatGPT could offer some ideas on how to approach a task, but this was the equivalent of someone standing next to a candidate and giving them advice** — because the tasks require constant engagement, this did not meaningfully impact scores.

### The conclusion

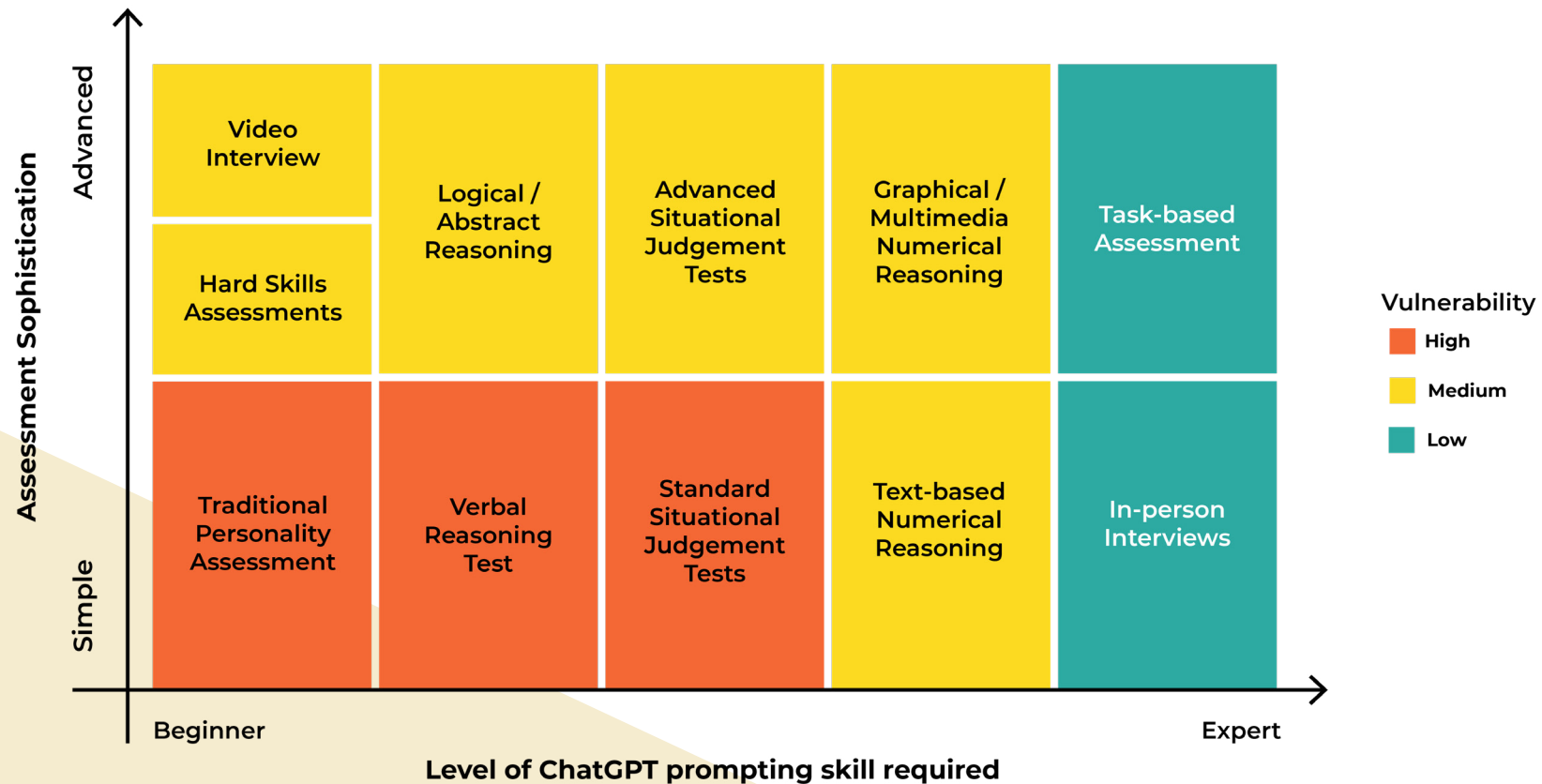
As a task-based psychometric assessment provider, it would be easy to view this conclusion and believe that we're biased. But — as a vendor who strives to eradicate bias at every opportunity and is passionate about creating a great candidate experience — we believe that Generative AI poses a big threat to the whole assessment industry and we went through a concerted effort to stress test this.

In line with this goal of transparency, we conclude that today, the Task-based Assessment approach is the most robust that the industry has to offer. Though we've also made most of our research public so TA Teams and academics can make their own minds up.

# ChatGPT Vulnerability Matrix

This matrix maps the vulnerability of common assessment types vs ChatGPT. This varies along two dimensions: the sophistication of the assessment method, and the level of ChatGPT prompting skill required.

Most traditional text-based assessment types are highly vulnerable to Generative AI, while in-person assessments and Task-based Assessments have low vulnerability.





## ChatGPT is evolving...

**On 25th September 2023, OpenAI – the business behind ChatGPT - revealed that they are now beginning to roll out new voice and image capabilities** which according to them “offer a new, more intuitive type of interface by allowing you to have a voice conversation or show ChatGPT what you’re talking about”.

*“We’re rolling out voice and images in ChatGPT to Plus and Enterprise users over the next two weeks. Voice is coming on iOS and Android (opt-in in your settings) and images will be available on all platforms.*

*Snap a picture of a landmark while travelling and have a live conversation about what’s interesting about it. After dinner, help your child with a math problem by taking a photo, circling the problem set, and having it share hints with both of you.”*

### **You can learn more here.**

Given ChatGPT’s performance on traditional question-based assessments to date, it doesn’t take a huge leap of the imagination to assume that these updates could make it even easier to use ChatGPT to complete a text-based assessment and outperform the average candidate. As soon as the updates are available, we’ll be testing the new functionality to determine whether this is the case. As always, you can [register for our TA Disruptors newsletter](#) to be notified the moment the new research drops.



**Part 4**

**What happens next? Next steps for progressive TA leaders adapting their selection process**

# What happens next?

It's now clear that Early Careers candidates can use ChatGPT to complete traditional, question-based assessments and tests and 7 in 10 expect to do so within the next 12 months. This has big implications for TA leaders and their teams. And this brings us back to the conclusion that we shared at the beginning of this report.

“

**The obvious and logical answer is not simply to Deter or Detect AI usage, but to redesign your selection process in a way that recognises and embraces the use of Generative AI without undermining the effectiveness or accuracy of your recruitment methods.**

”



This process will need to be part marathon, as this technology and its usage adapts over time – and part sprint, as we race to get ahead of the very real threat that ChatGPT poses to selection within the next 12 months. In the same way a volcanologist can spot seismic activity a year out from an eruption and use that data to evacuate and minimise the consequences ahead of time, the data shows us that **TA teams must now begin to make plans to mitigate against ChatGPT's use in the selection process.**

This presents a few options in the first instance.

If TA teams aren't in a position to redesign their selection process right now, they'll need to rely on a **Deter and Detect strategy**, noting that this will likely only be effective in the short-term.

But if TA teams are ready to move quickly to mitigate, they'll be able to opt for the **design strategy** to get ahead of the challenges that will be posed to them in the next year.



# Option one:

## Deter and Detect usage of Generative AI in the selection process

### Explanation:

The 'deterrence' element comes, unsurprisingly, from making candidates aware that they'll be monitored or prevented from using it. Careers sites have always highlighted that if an applicant is caught 'cheating', they'll be removed from the selection process. Other deterrent techniques include preventing copying and pasting or having multiple applications open but these are easily bypassed and feel clunky.

**The difficulty with Generative AI, as our research shows, is that the majority of students don't consider using it as 'cheating', any more than they would using a calculator instead of relying on mental arithmetic.**

Other businesses are taking a less severe, more transparent approach to deterring candidates. By letting them know that using Generative AI will obscure the 'real them', the hope is that candidates will be more willing to participate in the assessment without assistance. In the highly competitive and high stakes environment of job applications this is unlikely to deter many.

### Detecting candidates using Generative AI means using monitoring algorithms to 'flag' suspect responses.

These tools were effective when manipulation happened only in around 10% of applications, but as Generative AI tools become used by the majority, this will both become ineffective and expose the unreliability of detection algorithms (and potential bias).

A more draconian option offered by some vendors is online video proctoring — an option which has been around for some time — and enables a test-taker's entire assessment session to either be streamed live or be recorded by the system automatically, by auto-enabling a device's webcam. It's then up to either a test administrator, or an AI system, to closely monitor video details, examining suspicious activities during an online test or assessment.

### But remember that when it comes to proctoring only

**43%** of candidates would feel comfortable with this form of monitoring.

### And these numbers shrink to:



**Proctoring runs the very real risk of negatively affecting these groups disproportionately — raising ethical concerns about the viability of this method.**

# Option one:

## Deter and Detect usage of Generative AI in the selection process

### Effectiveness:

Traditional assessment vendors who advocate for a 'Deter and Detect' approach take the position that using Generative AI tools like ChatGPT is a form of cheating. They argue that there has always been a challenge with candidates sharing questions online or being coached on how to improve their answers and they have the means to flag and highlight such cases. Perhaps for this reason, traditional assessment vendors claim that detecting candidates using ChatGPT is both accurate and effective.

### Five reasons why Detering and Detecting Generative AI usage may not be effective

- 1 ▶ No ChatGPT detection models have been shown to work effectively as of today.** Some sources even report that 2 in 10 times these detection methods produce a false positive; meaning you risk falsely accusing 20% of your candidates of cheating, potentially harming your employer brand.
- 2 ▶ Detection models will date quickly.** It's also worth noting that given how quickly the underlying language models change and improve, there's a high chance these detection methods will become out of date.
- 3 ▶ Prevention tools are easily circumvented.** Candidates can easily use an iPhone or Android phone to scan text, feed it into the ChatGPT app, and input the suggested answer into a computer in just a few seconds.
- 4 ▶ Flagging candidates as suspicious could harm diversity.** What will you do if a candidate is flagged as suspicious? If 7 in 10 candidates are willing to use ChatGPT in their application, exclusion runs the dual risk of:
  - a. Reducing diversity and reach**
  - b. Presenting your employer brand as technophobic**
- 5 ▶ Poor candidate experience.** We have to consider the impact of harsher detection measures on the candidate experience. Existing research shows how these methods (especially online proctoring) are likely to increase candidate anxiety, blurring your view of candidates' real abilities, and reducing the diversity of your talent pool further (Hausdorf, LeBlanc, Chawla 2003).

# Option two:

## Redesign your selection process to ensure a level playing field

### Explanation:

The other option is to redesign your selection process to be progressive and embrace the age of Generative AI. This will mean reviewing each step of the process and 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 a candidate wanted to use it.

Instead of layering on more detection tech, this approach to identifying candidates most suitable for the role would involve simply replacing the vulnerable language-focused sifting methods (whether a form or a test) with a different type of assessment design.

Our research has revealed that TA teams should review the vulnerability of their traditional question-based assessments and either amend their assessment design to address the weaknesses (where possible) or consider a more modern assessment design that is Task-based.

### Three key elements required for a robustly designed selection process.

The Task-based Assessment has three components which make it more robust. TA teams could also consider looking for these elements in other assessment designs too.

#### 1 ► They should be non-verbal:

Instead of relying on language-based questions, a Task-based design can be built on visual interactions that bypass ChatGPT's linguistic prowess. We've tested automated ways of explaining tasks to ChatGPT but none of these worked. Interactive, visual tasks provide an inherent defence against Generative AI.

#### 2 ► There is no right or wrong answer:

Most traditional assessments have a definite or 'preferred' answer, and this binary nature leaves them vulnerable to Generative AI tools. However, Task-based Assessments offer a different approach - they capture how candidates respond to tasks through subtle shifts in behaviour and score every move they make, rather than simply giving a right or wrong score. While some traditional assessments may claim to lack a right or wrong answer, their robustness must be tested at the ChatGPT-4 level. Moreover, these traditional assessments lack a cognitive element.

#### 3 ► Refreshed and novel items:

Generative AI tools rely on a chatbot style input method to explain the task. Traditional assessments have just one format and style whereas Task-based Assessments have multiple styles, and within each style, a fast changing set of requirements. It would be too time-intensive to craft the prompts each time, let alone within the time limit.

# Option two:

## Redesign your selection process to ensure a level playing field


---



### Effectiveness:

The effectiveness of the selection process is dependent on two things:

- 1 | Being able to differentiate fairly and consistently between candidates
- 2 | Giving the candidate the opportunity to be their authentic selves



Our research has shown that Generative AI both supports and undermines these two key pillars of effectiveness. If they are to be achieved in a Gen AI-centric world, then the selection process has to be designed with these in mind - not just tinkered with. We need to do more than simply update our career sites with new language and instead, design our process from the ground up with the knowledge that these tools exist and candidates want to use them.

Using assessment designs that sidestep the traditional question-based, right-or-wrong format will allow organisations to open up a rich landscape of opportunities to gauge candidates' true potential to succeed in a role. After all, the main draw of the 'Design' solution is allowing for a more nuanced understanding of candidates' abilities, tapping into areas that are untouched by AI assistance, and hence, presenting a truer picture of what they're really capable of.

And where Business Psychology support is available to help explain, embed and develop this approach within their organisations, forward-thinking TA Teams have an opportunity to not only survive in the age of Generative AI — but to thrive.

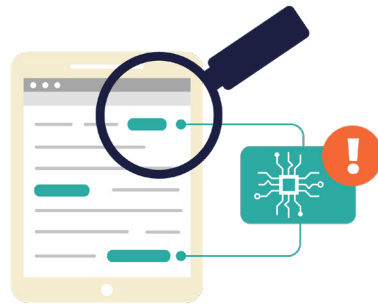
# Deter and Detect... or Redesign?

How Talent Acquisition teams can mitigate against Generative AI's use in the selection process.

## Deter and Detect

Why Detering and Detecting Generative AI usage may not be effective.

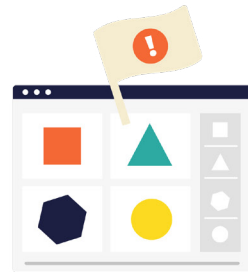
**1** No ChatGPT detection models have been shown to work effectively as of today.



**2** Prevention tools are easily circumvented.



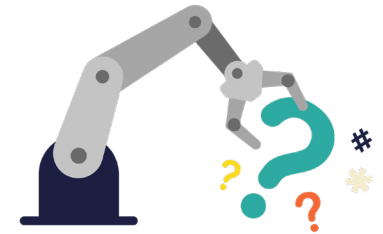
**3** Flagging candidates as suspicious could harm diversity.



## Redesign your selection process

The elements required for a robustly designed selection process.

**1** **Make your assessment visual**  
Removing text and questions makes it almost impossible for Generative AI to interact with the assessment.



**2** **Make it dynamic**  
An assessment with a fast-changing set of requirements can't be explained to an AI model.



**3** **Remove binary scoring**  
Score every step the candidate takes for a more complete picture of their ability.



# First steps for TA professionals to take now

**The impact of Generative AI and tools like ChatGPT is going to be huge and ignoring its impact is not an option. We've shared our research on how broadly these tools are being used and how with little or no training any user can 'ace' traditional assessments, especially those which are language and reasoning based.**

We don't pretend to have all of the answers. But based on the above insights and observations, we believe there are some immediate actions that Talent Acquisition professionals must consider when thinking about how to adapt their hiring processes in light of candidates' increasing use of Generative AI.



## Step 1

**Provide and deliver a Generative AI Vulnerability Audit.**

It will be too late to address issues if pass rates suddenly jump, quality becomes more variable, and diversity targets start to fall again. This is especially important as it's the top of the funnel sifting that is most at risk of distortion.

Understanding which stages are most at risk is the first step towards developing a robust future approach to avoid homogenous or inflated candidate results.



## Step 2

**Create a set of internal recommendations that outline how you plan to redesign the selection process to embrace Generative AI, rather than criminalise it.**

Once Step 1 above has been completed, you'll have the information on the areas you need to address both immediately and in the long term. Depending on the results of your audit, some obvious changes could be changing your assessment design from question-based to Task-based.

You might decide to change the wording on your career site to make your position on the use of Generative AI explicit — whether welcoming it or cautiously accepting its use. Law professor Tammy Pettinato Oltz [offers good advice on this](#).





## Doing nothing is a risky choice...

**One thing is absolutely clear for Early Careers teams. Doing nothing and ignoring the impact of Gen AI on the recruitment process is a risky choice.**

Over the coming months, we can expect more and more vulnerabilities to come to light. Talent Acquisition professionals must begin working through this challenge now to get ahead. This report and its underlying research are designed to help inform what you can do and why its so important to do that now.

No matter where you now stand in your thinking, you can join us on this journey by signing up to our TA Disruptors Newsletter, where we'll be sharing more insights and progress on the impact of Generative AI each week.

[Register here for the TA Disruptors newsletter.](#)



## 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 3 million candidates worldwide something different: a stress-free, unbiased candidate experience that truly rewards them for their time. Join the 350 pioneering companies leading the way, including Vitality, Molson Coors, Burness Paull, TalkTalk and Siemens.

**For more information, visit: [www.arcticshores.com](http://www.arcticshores.com)**

