

Do Cheaters Prosper? A Look at Cheating in Hiring Workflows

By Andrew Papa and Tom Cornell

Updated April 2024

Cheating is a problem in hiring. Hiring the wrong candidate due to deception or dishonest characterization of potential is both expensive and time consuming. Even worse, mis-hires can come at the expense of top-tier, honest candidates. Fortunately, our analysis of multiple forms of cheating revealed that only a small minority of candidates choose to cheat. HireVue can help you detect and deter cheating behaviors while preserving a positive experience for the great majority of honest candidates.

There is a fine line between cheating and interview preparation that makes cheating difficult to define. To demonstrate, let's do a quick thought experiment: Does working with a career coach to practice interviewing constitute cheating? Would you say that using Google to prepare for an interview constitutes cheating behavior? How about jotting down notes from those online resources to prepare for the interview? What if the candidate has those online resources open in a different tab while taking the interview? Finally, what about actively searching for and utilizing interview responses on Google or ChatGPT while completing an interview? As you can see, there is a fine line between interview preparation and cheating.

HireVue clients deal with varying forms of cheating depending on the various job families, geographies, and talent pools involved in their hiring workflows. Companies utilizing our Assessment Simulations may be concerned with validating a candidate's identity, whereas code challenge users may be more concerned with candidates utilizing ChatGPT to generate solutions.

Cheating is ultimately defined differently by each company. HireVue intends to provide information and tools that will help you identify cheating behaviors that fit your own definition of cheating within your hiring workflows. For our purposes in this paper, we'll define cheating behavior in interviews or assessments as: "Deceptive or dishonest actions taken by a candidate to misrepresent or embellish his or her knowledge, skills, abilities, or their potential for the role." In this paper we'll take a look at the most common cheating concerns, discuss their validity, and recommend actions we can take to reduce the impact of these behaviors.

Artificially Enhanced Answers

The first form of cheating behavior, Artificially Enhanced Answers, refers to a candidate using tools such as Google, ChatGPT, or even Auto Repair For Beginners 2023 to modify or fabricate entirely their interview or assessment responses.

ChatGPT

While concern over cheating in digital interviews and assessments is not new, these concerns have been heightened due to the public unveiling of ChatGPT. In case you were washed away in a Life of Pi style ocean-faring adventure for the past year and a half, ChatGPT is a new LLM (Large Language Model). It is an AI chatbot that has garnered attention for its detailed responses and articulate answers across many domains of knowledge. The interactive chat interface and the fact that ChatGPT is open and free to the general public has resulted in significant buzz over previous LLMs. Microsoft and Google, among other tech companies, have released their own chatbots, though ChatGPT is widely regarded to be the most advanced chatbot right now.

Is ChatGPT Effective for Assessments and Interviews?

Should you be concerned that candidates are using ChatGPT to cheat? While this is a valid concern, there is relatively low risk of a mis-hire resulting from ChatGPT. Our Data Science team has tested ChatGPT (v3.5) against our Virtual Job Tryout (VJT) and our AI-Scored Assessments and found that ChatGPT generally scores in the poor to intermediate range and does not beat out top-tier candidates.

ChatGPT vs Virtual Job Tryout (VJT)

Our ChatGPT VJT Cheating Pilot tested ChatGPT-3.5's performance on two core VJTs that included situational judgment, work history, work style, and simulation items. We used two conditions: one with minimal prompting and another with some additional prompting. The responses were scored to evaluate ChatGPT's performance.

ChatGPT performed poorly on most predictors and measure types. On a scale from 1 to 5 where 1 is the lowest and 5 is the highest, ChatGPT's overall scores ranged from 1 to 2 with a majority of item level scores also ranging from 1 to 2. The only exception was Work History items related to Career Stability where ChatGPT scores ranged from 4 to 5. On Work Style items, ChatGPT provided contradictory responses, sometimes even back-to-back on the same item. Additional prompting slightly improved performance, resulting in an overall score of 2 as opposed to 1 with minimal prompting. The results suggest that ChatGPT does not effectively cheat on HireVue VJTs. Performance issues aside, utilizing ChatGPT on a VJT was also found to be extremely impractical, taking over an hour for assessments meant to last 20 minutes.

ChatGPT vs AI-Scored Assessments

HireVue's data science team also tested ChatGPT against our AI-Scored Assessment questions. The experiment involved several different ChatGPT prompts based on our interview questions and tips for getting a good interview answer from ChatGPT that we found online.

The average HireVue candidate scores in the Intermediate range on our competency evaluations (i.e., a score of 3 on a scale from Novice - 1 to Expert - 5). When given just the interview question with no further instructions, ChatGPT also scored in the range of an average candidate. When given better prompting, ChatGPT was able to achieve a score of 5, but in the vast majority of instances still scored in the average range. However, ChatGPT responses were generic even with additional prompting, so candidate input would be required to align them to their actual work or educational experience reported in their application. Overall, with a variety of input prompts, ChatGPT interview scores ranged between 2-5 on our adaptability, problem solving, and negotiation competency areas with most responses performing average.

Although generated text from ChatGPT can result in a reasonable interview score, monitoring average assessment scores before and after the ChatGPT boom shows no upward trend or shift in candidate scores. AI-Scored Assessment scores have remained incredibly stable over time, suggesting there is no use of ChatGPT to gain a meaningful advantage at scale. Based on these results, we can infer that most candidates aren't using ChatGPT to cheat and those that do use ChatGPT effectively in their AI-Scored Assessments need to significantly tailor their responses to fit their background. Despite the evidence that ChatGPT is not significantly supporting cheating behavior today, we do take this concern seriously and are exploring product concepts to mitigate this issue as we expect ChatGPT to improve in future versions. We'd love to hear about your concerns around ChatGPT and how we can improve our product; please reach out via your HireVue Customer Success representative if you'd like to discuss this concern further.

ChatGPT vs Standardized Testing

Aon, a company that offers talent assessments, also completed a study by running ChatGPT prompts against assessment exercises. Aon found that ChatGPT performed well against knowledge-based tests (numeric, verbal, and logical reasoning) but generated mixed results against talent assessment exercises like HireVue offers today:

“For situational judgment tests, chatGPT is not bad at picking out most/least effective response options, although when presented with three or four desirable responses it often chooses a different response when asked multiple times. Gamified assessments that involve motion-based elements or multiple solution steps are either impossible to cheat with chatGPT or tend to result in unusable responses. In terms of prompt engineering, all of these types of assessments require quite sophisticated approaches to prompt engineering that don’t always work across the various assessment methods used.”¹

-Jouko van Aggelen, Chief Portfolio and Innovation office, Aon Assessment Solutions

Overall: A Top-Tier Candidate Prevails!

Overall, ChatGPT (v3.5) generally performs comparably to the average candidate when utilizing advanced prompting by an experienced ChatGPT user; although it is impractical to utilize for assessments. Even in the age of ChatGPT, there’s no substitute for top-tier talent!

Recognizing Artificially Enhanced Answers

Candidates that do choose to use the average responses generated from ChatGPT will likely demonstrate irregular speech patterns common to reading a text when responding to interview or AI-Scored assessments. This is something a human reviewer can sense when watching a candidate’s response—the same way you can tell when someone is reading from a teleprompter. People tend to use bigger and more formal words when writing compared to when they are speaking unscripted. AI generated answers tend to be more general and sometimes internally inconsistent. Finally, people tend to use more filler words such as “ah”, “um”, and “ya know” when they are speaking off-the-cuff, compared to when reading a scripted response.

What Tools Are Available to Reduce Artificially Enhanced Answers?

Multi-Stage Workflow Validation With A Live Interview

One of the best ways to reduce cheating behavior of all types is to utilize a multi-stage workflow in which a candidate’s knowledge, skills, and abilities are validated by an interviewer in a live interview setting. HireVue’s Talent to Opportunity (T2O) platform offers Builder which suggests structured interviewing questions; these competency-based questions elicit responses which will help your interviewers verify that past interactions with a candidate were genuine and honest representations of their abilities.

Code Challenge Similarity Score and Review Information

HireVue code challenges have a variety of tools to help identify cheating behavior. The code similarity score uses the MOSS algorithm to identify similar code in code challenges to help detect cheating. Additionally, the code challenge reviewer experience offers playback of the candidate completing the challenge and informs the reviewer when the user left the window and/or pasted text. This will help reviewers detect if a candidate used ChatGPT to form an answer. Although, many organizations are utilizing ChatGPT as a “copilot” to boost productivity; using ChatGPT augment or start an answer may not match your organization’s definition of cheating.

¹ <https://www.linkedin.com/pulse/cheat-thats-question-jouko-van-aggelen-5izre/>

Follow up with an “Explain Your Answer” Question

When assessing a candidate on technical skills, best practice is to follow up with an On-Demand Video question asking them to explain their thought process and answer provided in the previous question. Candidates that answered the question honestly, without artificial enhancement, will be able to walk you through their thought process and answer.

Time-Limited (Spontaneous) Interview Questions

Spontaneous On-Demand questions in our Talent to Opportunity (T2O) platform can help stop candidates from artificially improving their performance using ChatGPT or other resources. Spontaneous question mode requires candidates to provide their answers immediately, with no retakes and with limited prep time, removing the ability to formulate a response with ChatGPT. However, this comes at a heavy cost for both candidates and the hiring team. Research on time-limited On-Demand responses indicates that candidates are more likely to misrepresent their skills, knowledge, and abilities, in part due to the increased time pressure.^{2 3} Ultimately, this behavior means that time-limited questions are not the most effective way to identify potential in a candidate. Reducing a few potential instances of cheating also reduces the effectiveness of the question for the vast majority of candidates who don't cheat. Additionally, these questions cannot be made WCAG-compliant, meaning they penalize candidates who require Accessibility-based accommodations and generally provide a poor candidate experience.⁴ So we recommend only implementing this option after careful consideration of the drawbacks.

Choosing Smart Questions

The best way to neutralize the impact of ChatGPT and similar tools is to ask candidates questions requiring them to apply their skills, abilities, and experiences rather than questions which can be answered by copying and pasting from ChatGPT or other internet sources.

When an interviewer asks a candidate to “estimate how many windows are in New York City” they’re asking the question to see how a candidate solves problems with a high level of ambiguity. A successful candidate will take the interviewer through their estimation process and will account for underlying assumptions in their thinking. Whereas, the candidate using ChatGPT (v3.5) may respond with the actual number or an attempt to obtain an actual number. ChatGPT is unable to apply an individual's skills or speak to a candidate's relevant historical experience, so we can limit the impact of ChatGPT by choosing smart interview questions.

Like all technology, ChatGPT continues to evolve, improve, and embed itself in our lives. We will continue to analyze the efficacy of ChatGPT and other technology and enhance our hiring solutions accordingly.

²<https://onlinelibrary.wiley.com/doi/full/10.1111/ijisa.12341>

³https://www.researchgate.net/publication/366067148_Is_More_Always_Better_How_Preparation_Time_and_Re-recording_Opportunities_Impact_Fairness_Anxiety_Impression_Management_and_Performance_in_Asynchronous_Video_Interviews

⁴<https://www.w3.org/WAI/WCAG21/Understanding/no-timing>

Shared Responses

Another form of cheating is candidates sharing answers to interviews or assessments.

Do Candidates Use Shared Responses?

Our Data Science and IO Psychology teams completed a study in early 2024, in which they ran a script on 1,000 transcribed candidate responses to the same AI-Scored assessment questions and flagged responses that were identical or nearly identical to other candidate responses. The script scores string similarity using the longest, non-consecutive common sequence of words in a pair of candidate responses. This type of analysis does not identify similarity by looking for perfect matches, or perfect sequences of words or sentences, but rather by identifying core commonalities in word choice and order that can't be significantly masked by swapping out individual words like company name, role title, person's name, etc..

In a sample size of 1000, <1% were flagged for responses of high similarity. These results indicate the vast majority of candidates are honest when responding to virtual interview questions. Entry level roles also saw a much lower rate of genuine shared responses compared to more skilled roles.

Candidates that were flagged for similar or identical responses were typically only flagged on one question, not multiple questions. We can infer that candidates utilize a shared script on one question, maybe in an attempt to briefly gain an edge, then revert back to the norm of honesty. Most importantly, candidates that were flagged for identical responses in our study performed poorly on the assessment anyway and were not selected to move forward in the hiring process. This means the candidates who are cheating are not doing a good job of it, and there's a low risk of hiring an unqualified candidate that utilized shared responses in this study.

The above script sharing study was also completed against On-Demand responses in 2023. Again, we identified surprisingly few incidences of genuinely shared answers between candidates. In the United Kingdom, n=2,100 candidates resulted in <1% showing any evidence of collaboration with another candidate. In China, n=3,000 candidates resulted in <2% showing any evidence of collaboration with another candidate. And in Japan, n=4,500 candidates resulted in no candidates showing any evidence of collaboration with another candidate.

The above research makes it clear that candidates don't typically utilize responses that are shared from others. But in high volume hiring a small percentage can yield more instances of cheating. What can we do to ensure candidates aren't tempted to share answers?

"i have learned how to accept rejection in my last job working as a sales representative you don't know how many times i had to you know and it wasn't just a simple no some people complained some said bad words some even hung up without saying a single word"

"i had to learn to accept rejection in my last job working as a sales representative in a call center you cannot imagine how many times i heard no on the phone every day and it wasn't always a simple note some people complain some people said bad words some people add up without saying a single word"

Reducing Shared Responses

First, as we noted above for ChatGPT, we should utilize smart interview questions that require a candidate to apply their skills, knowledge, abilities, and experience for success. Next, follow up with a live interview stage after your On-Demand interview. In a live interview setting, it becomes immediately apparent if a candidate has misrepresented their knowledge, skills, or abilities in a past interview by utilizing a shared answer. Alternatively, consider altering your questions often! Swapping questions in your hiring process is the proverbial belt and suspenders that will help keep candidate responses honest. Builder (available on T2O and HV4) offers numerous interview questions with behaviorally anchored rating scales (BARS) targeting each competency. Consider using Builder questions for immediate access to a library of structured interview questions that measure competencies proven to be important for success in the role for which you're hiring.

Builder

[< Back](#)

Competencies

Prescreen

Skills & Experience

To ensure your interview is fair, job-relevant, and suitable, when selecting questions, choose only those questions related to success on the job.

Adaptability

Show less

☐ Tell me about a time when you had to adapt to a change at work. What feedback did you receive? What did you learn?

☐ Tell me about a time you had to deal with a lot of ambiguity or unexpected changes. How do you adjust to change or uncertainty? What strengths do you have during times of change or uncertainty?

☐ Describe a situation in which you had to adapt your style, approach, or ideas to meet the needs of others.

☐ Give an example of a situation when your work was not going as planned, and you had to change your

Impersonation: The Hiree is Not the Interviewee

Our third and final category of cheating behavior is impersonation, where some individual impersonates the candidate and completes the interview or assessment on their behalf. Unfortunately, this form of cheating can be surprisingly common in some geographies. We've seen individuals offer to complete assessments for money.

In some areas, hiring organizations will complete multiple rounds of interviews and assessments with an individual only to find an entirely different person reporting for work on day one. HireVue takes these issues seriously and has work planned toward the end of 2024 to address these issues in the product.

Macro level trends and statistical methods can help determine if cheating is taking place. If assessment pass rates are increasing, we can try to identify drivers of those increasing rates overall and within individual exercises of the assessment. We can also study completion time trends; decreasing completion times indicate candidates are completing assessments more quickly than the comparison population. Neither of these trends, higher scores and lower completion times, necessarily indicate cheating is taking place. In fact, these changes are more often due to candidate population differences over time. But, they are some initial data points we can use to take a deeper look. Logically, one could make the assumption that impersonators completing assessments for money tend to score well and may move through assessment questions more quickly having a general response pattern they follow. Our IO Psychology team ran a study in 2023 to investigate that assumption by flagging the fastest 10% of VJT assessment completion times that also scored the highest score of a 5. In this study, 1.9% of completions were in the fastest 10% of completion times and also achieved a top score. It is important to note that while this response pattern is logically aligned with cheating behavior we are unable to conclude that this population did cheat. Ideally this is a topic of further research in the future. Candidates that were flagged for these suspicious response patterns only demonstrated those suspicious patterns in one exercise of the VJT assessment. And, only 0.2% of candidates demonstrated those patterns in 2 or more exercises. This finding is consistent with the results of our shared scripts study, where candidates utilized shared scripts for 1 assessment question and then reverted back to their normal honest behavior.



How Can We Reduce Impersonation?

Include A Live Interview In Your Workflow

First, as noted previously, a live interview following an assessment is the best way to confirm the candidate is a good fit for the role and is genuinely representing their own human potential. Structured live interview questions are a great defense for the few candidates that do choose to cheat while also supporting a positive candidate experience for honest candidates.

Pair Your Assessment With An Introduction Video

If you notice suspicious patterns occurring, you can validate that the candidate assessed is the person you hired by requiring an introduction video within the assessment stage. An introduction video is a great way for a candidate to speak to their skills, abilities, and past experiences in their own words. Alternatively, you could ask the candidate to upload a photo during the application process and then compare that to the introduction video or On-Demand throughout the hiring process.

Smart Interview Questions

Finally, it's worth repeating that smart interview questions can help reduce cheating behaviors by making a good response difficult to be completed by a confederate candidate. While I know my colleague Andy Bonk very well, it would be very hard for me to complete a thorough competency-based, structured interview by impersonating him and applying his knowledge, skills, abilities in responses.

What Else Can Be Done To Curb Cheating Behaviors?

Right to Retest Language

A valuable cheating deterrent occurs before the candidate even reaches the assessment; anti-cheating messaging can be included in your terms and conditions presented to the candidate at the beginning of the interaction. Specifically, candidates can be asked to agree to a number of related terms, such as, "I will take this assessment by myself without assistance from others." Language along these lines can include statements dissuading any number of nefarious behaviors, including the creation of answer keys or providing false information, and may include the ramifications of failure to comply. Right to retest language informs candidates that you're aware of the potential for cheating and have processes in place to manage it.

Transparency in the Candidate Experience

Cheating behavior seems to be fueled by fear and uncertainty. Fear that you, as a candidate, won't measure up and be successful in obtaining the role. Or fear that you'll come off poorly to the interviewer. A powerful way to help curb the desire to cheat is by providing a transparent experience for the candidate. If candidates know exactly what will be expected of them, they can prepare adequately, perform their best in the interview or assessment, and ideally feel less motivation to cheat. A great way to provide transparency is to let the candidate know what you're measuring in your various interview and assessment activities. Plus, laying out the hiring process in advance can help the candidate feel more comfortable throughout their interactions with your company. Finally, an added benefit of increased transparency is your recruiters will have fewer confused candidates wondering about the next step in their hiring journey.

HireVue's Continued Research and Enhancements

Our research concludes that the vast majority of candidates complete interviews and assessments candidly and honestly. For the very rare candidate that does cheat, they generally perform worse than the top-tier honest candidates and do not cheat on the entire test or interview. Additionally, using the tools provided in T2O and HV4, you can take steps to reduce the impact of cheating while simultaneously building onto HireVue's excellent experience for all candidates.

As cheating methods and technology continue to evolve, our research and tools to mitigate cheating will also evolve. At HireVue, we're dedicated to helping you identify and reduce cheating behaviors in your hiring processes. We'd love to hear your thoughts as we continue to refine our cheating mitigation roadmap. Here are a few concepts we're exploring for this domain:

- Disable copy/paste in candidate experience
- Warning or removing ability to change tabs during interview
- Additional interview data in the reviewer experience such as: recording count, completion time, records of changing active browser tabs, IP Address, login count, etc.
- Screen capture and image capture of the candidate during the assessment or interview and presented on review page
- Candidate photo upload for comparison throughout interview workflow

If you have questions, concerns, or feedback about cheating mitigation or our roadmap, please contact us via your account representative.