2025 Explainability Statement

This is HireVue's Explainability Statement. This document is intended to provide information on how the Artificial Intelligence (AI)-based assessments within our Talent to Opportunity Platform™ work, including when, how, and why we use this technology to assist our customers in making their hiring decisions. It is separate from our privacy policy, which is available at: https://www.hirevue.com/legal/privacy.

Please note that this is a 'living document' which will be updated from time to time, based on updates to our systems and processes. HireVue considers the ethical development of AI along with data security and privacy to be core values. Part of ethical AI development involves taking an evidence-based approach to evaluate our AI-based tools. HireVue is actively engaged in research with academic collaborators to better understand how our AI tools work, ways to improve them, and we open source our techniques such as algorithmic debiasing to offer methods to improve AI tools more broadly (see a list of published peer-reviewed articles here). Further, we routinely have third-party audits conducted on our tools to ensure they meet professional and AI regulatory standards or requirements. In addition to its research, audit activities, and the creation of its Expert Advisory Board to help guide ethical AI Development, HireVue developed this Explainability Statement to explain HireVue's processes and in an effort to assist our customers in fulfilling their obligations as 'data controllers' in compliance with data protection laws including EU / UK General Data Protection Regulation (GDPR), and as a 'provider' under the EU AI Act as a vendor of AI tools used for employment decisions.

If you have any queries, we can be contacted at press@hirevue.com.

What is our Talent to Opportunity Platform?

Our Assessments

HireVue transforms the way organizations discover, engage, and hire the best talent. Connecting companies and candidates anytime, anywhere, HireVue's industry leading end-to-end Talent to Opportunity Platform™ features video interviewing, assessments, and conversational AI. HireVue has hosted more than 40 million video interviews and over 200 million candidate assessments for over 1100 customers around the globe.

HireVue offers a broad suite of traditional and AI-scored assessments, including video interviewing, simulations, and online game-based challenges. These can be combined in a 'modular' fashion for specific roles. Using different assessment types allows us to measure different competencies and capabilities. For example, we can combine an assessment of teamwork skills with another that evaluates problem-solving and decision making. Customers who choose HireVue's modular system work with our Industrial/Organizational **(IO)** Psychologists to conduct a job analysis to determine the competencies required for a specific job role they are looking to fill. The job analysis process guides the final selection of job-relevant assessment content to include for a particular role.

What is our Talent to Opportunity Platform?

Our AI

AI, in the broadest sense, means technologies which are capable of undertaking or facilitating tasks that would otherwise require human thought or reasoning. Within this very broad definition there are many different techniques and applications. The aim of this Explainability Statement is to explain which AI techniques we use, why we use them, and what factors they take into account.

At HireVue, we use AI to score our interview- and game-based assessments. The purpose of our AI-scored interview assessments is to give recruiters a standard, structured, and fair way to screen many candidates, in a shorter time with greater accuracy and at a lower cost than traditional human-led interviews. Our AI-scored interviews don't replace recruiters. They simply help recruiters and talent acquisition teams assess more candidates quickly, consistently, and accurately. Recruiters and hiring managers are provided materials and training on what competencies are measured in the interview and how to interpret interview assessment competency scores (we provide further detail on this later).

The purpose of our game-based assessments (GBAs) are to offer an engaging way to measure job-relevant competencies or attributes such as critical thinking skills, as well as personality characteristics (i.e., how people think, act, and feel in work situations). Machine learning was used to train our scoring models to identify relevant patterns between candidates' behavior in games and various cognitive abilities and personality profiles important for success in specific job roles. The use of machine learning allows us to generate scores for games that optimally balance the goals of: (1) accurately measuring each job-relevant competency, and (2) minimizing scoring differences across demographic groups (see the section below on 'How do we use AI in game-based assessments?' for further details).

How do recruiters use the results from HireVue's AI?

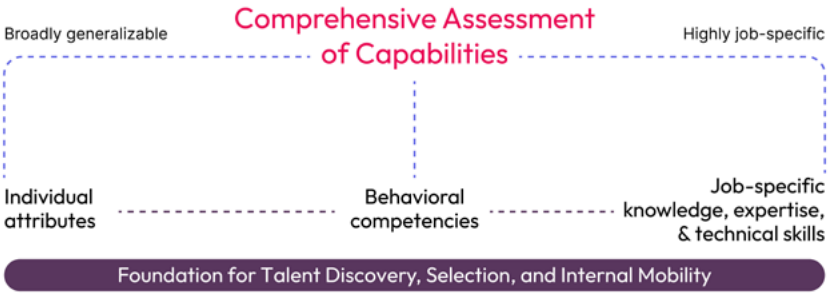
HireVue provides a tool which assists employers in evaluating candidates in their own hiring process, but the ultimate decision as to what action is taken based on that information always remains with the employer. In EU and UK law, this means HireVue is a 'processor' of personal data, whereas the employers using our platform are the 'controller' of data, because they take the ultimate decisions on the purposes and means of processing. Since HireVue's platform does not make recruitment decisions, if the candidate wishes to query the decision-making in the recruitment process then that challenge needs to be made to the hiring company which uses HireVue's platform (according to its own configuration; see below) and ultimately makes the final recruitment decision. As set out in detail below, HireVue provides end users with candidate assessment reports (see Appendix A), including a recruiter report that shows how candidates performed on assessments and a candidate feedback report that can be shared with individual candidates to offer insight into their performance.

Why do we use AI?

Choosing the right candidate

The aim of any hiring process is to find the right candidate(s) for a job. There can be hundreds of factors involved in making good hiring decisions. Even entry-level, hourly jobs require a unique combination of competencies, cognitive abilities and personality types – few of which will be clear from an application alone, or a CV or résumé. Historically, there were two main ways to assess job fit: (1) human-led interviews and (2) questionnaires marked by examiners. Both are potentially problematic, as we explain below when comparing these methods to our technology.

To identify the attributes that are most important for job performance, we draw on over 100 years of research in the field of IO Psychology, the study of human behavior in organizations and the workplace. Our methods use AI to produce a single comprehensive assessment of each candidate, which organizations can then use to make better, more informed hiring decisions.



Our AI-Scored Interviews

Our AI-Scored Interviews have multiple advantages for candidates and employers:

- **Mitigating bias**. Any hiring process involves the risk of bias – the tendency to give systematic undue preference to certain characteristics not related to job competencies, or to discriminate against particular groups. Bias in human interviews, without rigorous rater training, is well-documented but can be difficult to spot until it is too late to correct. By contrast, our AI-scored interviews assess job-relevant competencies while minimizing the potential for bias. To do this, our AI scoring models are designed to mitigate or reduce differences between demographic groups while still maintaining the prediction of job-related competencies. When our AI-scored interviews are used in a hiring process, this leads to more fair and equitable treatment of demographic groups (i.e. when used for a pass/fail type decision, small group differences in scores are likely to yield highly similar passing rates among demographic groups).

We follow legal guidelines at all stages when developing, testing, and monitoring AI assessments, and in many cases we test for group differences beyond those required by law. These protections include, but are not limited to, the '4/5ths Rule' mandated by the US Equal Employment Opportunity Commission, according to which if the selection rate for a certain group is less than 4/5ths of the group with the highest selection rate, that can be considered evidence of 'adverse impact' on the group with lower scores. We perform additional checks using well-established ratio and statistical metrics for group differences (the technical terms for which include 'Cohen's d', 'Fisher's Exact', '2 Standard Deviations', and others). Additional information on our bias mitigation strategy can be found here:

- How to Advance Diversity Hiring with Big Data
- Rottman et al. (2023) - New Strategies for Addressing the Diversity–Validity Dilemma With Big Data

Why do we use AI?

Our AI-Scored Interviews

- **Consistency.** Each human interviewer will have slightly different hiring preferences, based on their own unique background and experiences. There may even be differences between the evaluations made by the same interviewer, depending on circumstances such as the time of day or other pressures of which the interviewer is unaware. A given candidate can be 'lucky' or 'unlucky' depending on who happens to interview them, and when. These differences in interviewer preference can lead to significant variations in results for candidates with exactly the same competencies. This problem is sometimes called 'noise' or 'unconscious bias'. Unlike interviews conducted by humans, our AI models are completely consistent across candidate pools. All candidates are asked the same questions and have the same opportunity to answer them. Their answers are all assessed and scored using the exact same algorithm and evaluation criteria to ensure consistent and objective treatment of all candidates. Our system avoids the danger of a particular human interviewer scoring a candidate well or poorly based on personal preferences that have nothing to do with job competency.

- **Equality of opportunity.** Instead of needing to be available for an interview at a specific time or place, candidates can record their responses to HireVue interview questions at a time of their choosing, using a computer, tablet or smartphone. In the same fashion, recruiters can review and compare candidates' interviews at any time. Allowing all candidates to undertake video interviews enables the hiring organization to consider a wider pool of applicants, some of whom would be excluded because of an inability to attend a particular interview slot (for example, because of other work or care commitments). For candidates in need of special equality of access accommodations, our system is set up to have well-defined alternative assessments.

- **Better candidate experience.** Unlike a traditional interview, our AI-scored interviews allow candidates multiple attempts to answer each question, if they feel that the first attempt did not go as well as they would have wanted. In addition, our AI-scored interviews allow for clear feedback to be given to every candidate no matter how they scored as soon as the interview is complete– something which would be time consuming and difficult for human interviewers to do for every applicant. We have provided a sample 'candidate feedback report' in Appendix A, a document that can be shared with individual candidates to offer insight into their performance.

- **Good data means good decisions.** The result of our AI interview techniques is a highly accurate assessment of specific competencies, mitigated for bias. Our AI-scored interviews provide excellent insight into behavioral competencies such as adaptability, communication, and problem solving. Our AI-scored interviews can also be used to improve the hiring process over time, because data collected during such interviews can later be mapped against the performance of those who were hired. This type of data-driven comparison is extremely difficult to accomplish using traditional human-led interviews because the relevant data is not collected in a systematic way. Relatedly, customers can choose to give greater weighting to certain competencies (for example, Teamwork) in a defined and structured way in our interviews – something that would be difficult to do with human interviewers since it's difficult for humans to disentangle different attributes of an interviewee.

- **Costs savings for customers.** Compared to using human interviewers to screen all candidates, HireVue's customers are able to obtain major cost savings using our AI through reduced time to evaluate and hire employees as well as eliminating travel costs associated with in-person interviews. In our experience, organizations using HireVue experience a significant return on investment.

Why do we use AI?

Our Game-Based Assessments

Game-based assessments (GBAs) refer to an assessment method for measuring an attribute or psychological trait using game design principles that are intended to lead to a psychological state for the candidate known as gameful experience (Landers et al., 2019). HireVue GBAs may be deployed as cognitive ability only, or a combination of cognitive ability and non-cognitive competency domain measures such as the Big 5 personality. Each HireVue game takes only a few minutes to complete. At HireVue, our approach to game design starts with identifying the psychological construct we want to measure, and then building a game that is designed to elicit behavior related to the target construct(s).

Our cognitive GBAs are designed to be a general measure of cognitive ability. We design each game to target a specific ability (e.g., Working Memory with our Digitspan game). While each game is designed to measure a specific ability, the specific abilities are related and combine to form a general measure of cognitive ability.

Portrait is HireVue's GBA of the Big Five personality traits using images and adjectives. It measures Openness, Conscientiousness, Extraversion, Agreeableness and Emotional Stability. One of the most widely recognized ways to categorize personality traits is on the basis of the Big Five model (e.g. Goldberg, 1992; McCrae & Costa, 1987).

We also offer a candidate feedback report that makes it clear to candidates what is being tested for the position for which they are applying (for example, compassion for a Registered Nurse role, or service focus for a Customer Service Representative role). See Appendix A for a sample candidate feedback report.

In addition to minimizing bias and gathering richer data, our GBAs have several advantages over traditional questionnaires:

- **Speed.** Traditional multiple choice cognitive skills tests last 30-45 minutes, as opposed to approximately 6-15 minutes for our GBAs.
- **Flexibility.** Our cognitive games adapt in real time based on a candidate's performance. If a candidate completes one level in a game, the next level they will be asked to complete will be more difficult. If they fail a task, they will be given an easier one. This allows for more detailed data to be gathered on individual candidates than would be possible using a static test.
- **Improved candidate experience.** Based on feedback by 1.5 million candidates who have taken HireVue's AI-scored assessments: 80% enjoyed the experience and appreciated the opportunity to differentiate themselves; 85% thought it reflected well on the employer's brand; 70% rated the experience as 9 or 10 out of 10; and 89% said it respected their time.

How did we design our AI?

How do we use AI in video interviews?

There are three stages involved in the AI technology we use in video interviews, each of which involves a different system: (1) transcribing spoken words to text, (2) understanding the meaning of the text, and (3) assessing/scoring the candidate's answers based on expert human rater evaluations of answers to similar competency-based interview questions.

Importantly, our AI relies only on what is said by the candidate and does not use any video analysis or other audio characteristics (meaning that we do not assess a candidate's facial expressions, body language, their background and surroundings, or tone-of-voice).

1. Changing speech to text

First, we convert the candidate's speech to written text, using a third-party speech-to-text transcription system developed by a company called Rev.ai. This technology recognises the sound of words based on its experience and learning from over 50,000+ hours of human-transcribed content across a wide range of topics, industries, dialects and accents. We have provided more details about the transcription accuracy of Rev.ai in the section below on Third Party Providers.

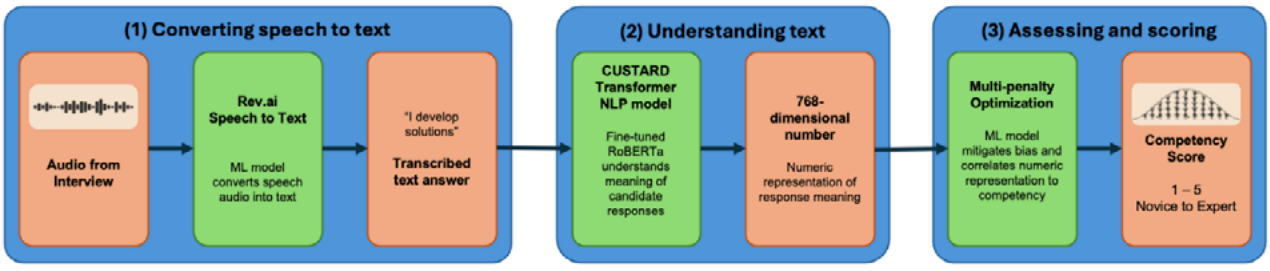
Rev.ai, in common with our own AI systems, uses a technique called 'machine learning'. Machine learning is a form of data processing that identifies statistical patterns from data sets. Rather than being programmed with predetermined responses to a set of conditions, a machine learning system is designed to develop its own responses to those conditions under a training regime. For instance, a simple machine learning system might learn to differentiate between the spoken words cat and dog with training data that includes many audio examples of different people saying 'cat' or 'dog', which are then labeled before being fed into the system. After the AI has been trained on enough examples of training data the system will build a predictive model that can distinguish between the two words. The principles which a system has derived from the training data are called a 'model'.

Machine learning systems are particularly good at undertaking complex tasks where the rules can be difficult to specify with precision (such as understanding language) as well as for tasks involving the computation of very large amounts of data. For these reasons, machine learning is now very commonly used for tasks which involve understanding human language.

2. Understanding words and sentences

Second, based on the transcribed text, we use a form of AI called 'natural language processing' (NLP) in order to understand candidates' answers, as summarized in the diagram below.

Example AI-scored Interview Process for Problem-Solving Competency



How did we design our AI?

We have developed our own NLP model, '**CUSTARD**', which is based on a state-of-the-art language model called Robustly Optimized Bidirectional Encoder Representations from Transformers, or '**RoBERTa**'.

RoBERTa was adapted from a model designed by Google called BERT. BERT refers to Bidirectional Encoder Representations from Transformers and these models consider word context. More specifically, they learn about important information embedded in language, creating dimensional representations of text that are context-dependent. For instance, the word "green" in the following three sentences would be represented with different embedding vectors: (a) The company has started a new green initiative to reduce their carbon impact, (b) The family bought a new green colored car, (c) Due to the recent rain, the grass was very green. In this illustration, the vectors for the word "green" used in sentences (b) and (c) refers to a color and would be closer than the vectors for the word "green" used in sentences (a) and (b) where "green" refers both to a color and being environmentally conscious.

CUSTARD starts with this RoBERTa language model as its base and is further fine-tuned on interview data. The language analyzed by CUSTARD is processed by a 'deep neural network', a technology which comprises a collection of connected nodes or 'neurons' which can attribute a particular weight or significance to various features of the language presented to it. Table 1 below outlines the steps we took starting with RoBERTa as our base, ending with 19 fine-tuned CUSTARD language models.

Table 1. CUSTARD Fine-Tuning Steps

Step	Description	Data Source
1	First the system learns to understand English words in context by predicting masked words in a large number of documents; specifically, the RoBERTa model we used as a base was trained on a large corpus of text	160 GB of text data which is over a billion sentences or approximately 30 billion words
2	Our in-house RoBERTa model refines this general RoBERTa language model to the interview domain by predicting masked words over a million of interview transcripts	One and a half million unlabeled asynchronous job interview transcripts
3	Building on step 2, the model is fine-tuned to the task at hand by predicting expert human interview ratings on communication across tens of thousands of interviews	Over 30k interviews rated on the communication competency
4	Lastly, to fine tune to a specific competency domain, a final layer is added to the CUSTARD language model by predicting thousands of expert human interview ratings on a specific competency domain	Over 1k interviews rated on each of the 19 specific competency domains (e.g., Adaptability)

The process outlined above yields a total of 19 CUSTARD language models, each fine tuned to one of our competencies. The output of a CUSTARD language model is a numerical value – known as a 'vector' – which the model has generated based on passing a candidate's transcribed interview response through the deep neural network. Unlike many simpler NLP methods, our system is especially effective at understanding the meaning contained in response to a question, regardless of the specific vocabulary used. This ability to generalize makes it more difficult for candidates to "game" the video interview process by mentioning particular words or phrases in their responses. CUSTARD is capable of differentiating between the usage of the same word in different contexts. This is particularly important where the same word can have different meanings depending on the words around it. For example, the word "bank" is used in two different senses in this sentence: "Joanne went to the river bank today, and she visited the bank to withdraw cash on the way home."

3. Assessing and scoring each candidate

Third, once a CUSTARD model for a competency has understood and assigned numerical values to the candidate's response to an interview question, this numerical value is fed into a 'multipenalty optimized model' (a machine learning system). The multipenalty optimized model has been trained to score those responses against the relevant competency.

We have developed a separate AI model and set of questions for each competency. A sample of competencies we can measure include adaptability, problem solving, communication and willingness to learn. There are 19 interview-based competencies which we can cover, and we are increasing this list over time based on scientific research and our own data.

How did we design our AI?

Following best practices in structured interview design, of using guides to support consistent expert evaluations, we created a Behaviourally-Anchored Rating Scale **(BARS)** for each competency for which our questions were designed to elicit candidate interview answers. Creating scoring models for each competency based on expert BARS ratings, we have modeled the most accurate way of fairly and consistently rating structured interview responses of expert human evaluators.

HireVue's AI system scores each of the candidate's responses according to a BARS for each competency. Our BARS content and guides are based on data from thousands of real-life interviews, covering a diverse range of interviewees and job types. BARS for each of our competencies have five rating levels, from 'novice' to 'expert'. In Appendix B we provide the BARS used for our 'Communication' competency. Additionally, in Appendix B, we provide an example of how an interview answer can be scored along the response timeline as each statement relates to an anchor in the BARS guide.

As mentioned earlier, the models we use to assess candidates through interviews have been trained on expert human evaluations of structured interview responses using these BARs. Our AI-interview scoring algorithms are based on sophisticated analytic techniques to craft correlational-based models that mimic trained expert human rater judgements. The assessment scores provided by our AI-scored interviews are highly similar to the evaluations expert interviewers would provide, but without the unconscious biases.

To create the assessment scores for each BARS, HireVue collects thousands of expert human rater evaluations of standardized interviews and uses these ratings to train the models to score candidate interview responses. Our assessment development work and rater studies conducted over the past 6 years, have drawn upon over 900,000 applicant video interviews scored and evaluated for bias (see Appendix C).

More specifically:

- We collected scoring data from interviews for different levels of roles, type of companies, and geographic locations.
- We trained diverse teams of around 60 expert raters to evaluate the responses in each of those interviews based on specific competencies according to an evaluation guide based on using a BARS.
- The expert raters then manually scored each response in the interviews against each competency, with 2-3 separate evaluators scoring each candidate's answer.
- During the training process, we held regular calibration discussions to ensure consistency in scores from each rater. We also filtered any unreliable data (for example where there were audio issues or insufficient words in a response). We also re-scored or removed responses where rater evaluations varied significantly. Based on the above training, our multipenalty optimized regression model is able to score candidates' responses, by comparing them to the manually scored responses during the training exercise. As compared to a simpler regression model, a multipenalty optimized regression model helps to ensure that the algorithm generalizes well to unseen data, rather than 'overfitting' to the training data, as well as reducing subgroup differences. Overfitting can occur when a model is trained to be highly accurate for the examples it has seen before, but which then results in the model being inflexible and not able to generalize as well (which, in this case, could mean it is unable to recognise different but similar candidate responses).

How did we design our AI?

How do we use AI in game-based assessments?

We use AI in our game-based assessments to assess a candidate's cognitive ability or personality. Our technology works as follows:

- The candidate's score and other key game metrics including the ratio of levels lost and won, the total number of levels played, selected answers and/or highest level completed are fed into a ridge regression model (as explained above). A regression model is useful in understanding the relationship between different variables – cognitive ability or personality traits. An accurate regression model can predict or assess the value of a dependent variable (e.g. cognitive ability) based on a set of independent variables (e.g. the game performance).
- Our game-based assessment regression models were trained and mitigated in a similar way to our multipenalty optimized models used in the video interview process. Through multiple panel studies, we asked hundreds of individuals to undertake our game-based assessments and they were scored based on the aspects noted above. We then asked the same individuals to undertake traditional cognitive assessments (typically based on questionnaires), which gave us an accurate and reliable indicator of their cognitive ability. This data was then used to train our regression model to spot relevant patterns between candidate's behavior in games, and different types of cognitive ability.

Ethical AI & Consent

Our Ethical AI Principles

The following five principles guide our thoughts and actions as we develop AI technology and incorporate it into our products and technologies. HireVue practices will continue to evolve as we work with our customers, job-seekers, technology partners, ethicists, legal advisors, and society at large to ensure we are always holding ourselves to the highest possible standards.

1. We are committed to benefiting society.
2. We design to promote diversity and fairness.
3. We design to help people make better decisions.
4. We design for privacy and data protection.
5. We validate and test continuously.

Full details can be found on our website: https://www.hirevue.com/why-hirevue/ai-ethics.

AI Consent

Before completing a HireVue AI-scored assessment, applicants review an AI consent statement and have the option to opt in or out of the use of AI in evaluating their responses. This consent process is consistent with our Ethical AI principles and HireVue's commitment to AI transparency and explainability. In our AI Consent Statement (see example in Appendix D), the candidate is informed of the following:

- Where and why AI is used
- How it was developed
- How it evaluates responses
- How we monitor AI fairness and take corrective action when needed
- How the hiring team makes the final decision
- How opting out of AI evaluation will not exclude a candidate from participating the hiring process

Ethical AI & Consent

For our AI-scored interviews, candidates who opt out of AI evaluation complete the same HireVue assessment experience, but their responses are not scored by AI. Instead, recruiters manually review and rate these candidates' interview responses based on the same behaviorally anchored rating scales describing low, medium, and high performance on the relevant competency(ies). Similarly with our AI-scored games, candidates can opt out of AI and be evaluated by other, non-AI scored elements. Additionally, when a HireVue assessment is complete, candidates are provided with a Candidate Feedback Report (see example in Appendix A) which provides the candidate insight into how they performed on the assessment offering additional transparency into what is assessed.

What is the user journey for interview candidates using our AI platform?

System Configuration Impacts Candidate Experience

Recruiter or user training is provided by HireVue to our customers that details the following information concerning the HireVue assessment(s) deployed in their system. Main topics covered in the training are: how the assessment was designed, what the assessment measures and how it links to the target job, configuration settings for the assessment, how the assessment is scored and results presented in feedback reports, sample candidate communications, and detailed HireVue system or platform navigation.

As mentioned above, the assessment which applicants take will match the competency requirements of the position for which they are applying. An example AI-scored assessment will consist of 4-6 video interview questions (delivered asynchronously) and 2-3 game questions. Thus the entire candidate time to complete the video interview plus games is typically 15-25 minutes. Each interview question is designed by HireVue's IO Psychology team to elicit behavioral responses related to a specific competency (e.g., customer service). The games are designed to measure general mental abilities (e.g. numerical reasoning) or personality areas (e.g., conscientiousness).

How the candidate experiences the assessment is configurable by the company deploying the assessment. HireVue system consultants can assist with this configuration or setup and provide best practice recommendations. The main configuration decisions are:

- Preparation Time for each Interview Question (0 - 5 minutes, or Unlimited Prep Time): No Restrictions on Preparation Time Recommended (1 Minute Minimum)
- Interview Question Retries (Yes/No): Unlimited
- Candidate Feedback Report (On/Off): Recommended On
- Evaluation Transparency Screen (On/Off): Recommended On
- Reusability of Assessment (On/Off): Recommended On
- Rating Guidelines (On/Off and by Question): Recommended On and with 5 Star Guidelines (BARS) On
- Candidate Assessment Result Tiers (On/Off): Recommended On with Result Tier Labels Reflecting Client Use Case (e.g., Top/Middle/Bottom labels)
- Competency and Assessment Numeric Score (On/Off): Recommended Off
- Data Retention: Recommended 2 years, but follows company policy

What is the user journey for interview candidates using our AI platform?

The scoring of the interview and games assessment follow the technology described above and a report of results is presented in the HireVue system for recruiters and hiring managers to view. This report provides a description of the assessment taken, the competencies evaluated, a description of how the candidate scored on each competency, and an overall assessment result as compared to all the other candidates that completed the assessment for that position. Additionally, a Candidate Feedback Report can also be generated and sent to a candidate (see Appendix A for examples of these reports).

Communications to the candidate are managed by the employer using the HireVue system but typically consist of email or text messages informing the candidate how they did in the specific step in the hiring process and what to do next (e.g., "you have completed the application and now please complete the video interview or assessment by clicking this link"). These messages are customizable by each company and recruiter, but template messages are provided by HireVue to facilitate consistency in candidate communications. An example email text informing the candidate they have completed the assessment/interview, what happens next, and whom they can contact with questions follows:

Dear [Name], We have successfully received your interview for [Position]. There is no further action on your part for this interview and a representative from [Company] will contact you about the next steps. We are working very hard behind the scenes to complete the recruitment process and will update you as soon as we can. If you would like feedback on your assessment results please let us know and we will be happy to send you a report. In the meantime, if you have any questions please feel free to email me. Thank you again for participating, we wish you the best of luck and thank you for your time. Kind Regards, [Name and Email]

How did we choose our AI suppliers?

Third Party Suppliers

Our only supplier of AI components is Rev.ai, which supplies our transcription system. Prior to adopting Rev.ai's transcription system, we tested its accuracy compared to other transcription systems (e.g. one offered by Amazon) using Word Error Rate (WER) which is the standard metric for evaluating transcription accuracy. In the HireVue analysis we found the English language WERs in the United States for Rev.ai's system were less than 10% on average, whereas this average WER was 15-25% for the other transcription systems we tested. Incidentally, the estimated human transcription WER is approximately 5-10% (listening to recording and typing text). However, it is neither economically feasible nor time-efficient to use human transcribers when processing the millions of interview responses so they can be auto-scored with our AI algorithms.

Furthermore, we analyzed the WERs by country of origin to evaluate the impact of accents and by ethnicity. To check the accent impact we evaluated Rev.ai's accuracy in transcribing speech from native English speakers versus non-native English speakers with a variety of accents. The Rev.ai WERs were also lower than alternative services. We also evaluated the WERs for the transcription services by ethnicity of the applicants which yielded similar results such that Rev.ai outperformed the alternative services (e.g. Rev.ai WER 7-10% range White, Black, Hispanic, Asian applicants compared to 15-30% WER range for other services; same ethnicities). Though already best in class, Rev.ai's software continues to be improved over time, and we will incorporate such periodic improvements into our AI system as they are made.

How did we choose our AI suppliers?

Data sources

We do not obtain any data from third parties. Instead, our AI assessment systems are trained on HireVue's own data, which has been screened and checked in order to ensure that it is suitably diverse and representative. As mentioned previously, over the last 6 years we have conducted various rater studies to build and improve the algorithms and check/mitigate them for group differences (i.e., differences in scores by demographic classes and the potential for adverse impact or differential passing rates if a score is used for decision making purposes), and bias (i.e., differential accuracy by demographic classes). In these rater studies we have 99,411 expert ratings on video interviews across 19 competencies. Table 2 below shows a balance of demographic characteristics in our latest rater study. Appendix C contains the full table showing high levels of representation in the study of various gender, race, age, job level, industry and geographic type of applicants.

The developmental process also includes conducting adverse impact or bias analyses (detailed later) from which we sample from over 900,000 applicant records to check for and mitigate group differences based on gender, age, and race/ethnicity.

Public AI Models

As noted above, our NLP model, CUSTARD, is adapted from RoBERTa, which was designed by Facebook. RoBERTa was adapted from Google's BERT model. The BERT and RoBERTa models were designed by major technology companies and are widely used in NLP across different industries. We are confident that they represent the state of the art. As set out below, we have adapted these models to generate further improvements.

Explainability of the models and its limitations

The scores provided by our multipenalty optimized models can be explained by looking at the input variables (known as 'features') and assessing their relative importance to generating the output score. As explained above, our CUSTARD model calculates an embedding (i.e. list of numbers, or vector) to each answer. The CUSTARD embedding reflects 768 dimensions or features of the input sentence. We know the CUSTARD features correlate with expert ratings of interview responses (in technical terms, there was an average correlation r=0.69 across all of our interview scoring algorithms which is equivalent to or higher than values obtained in published research studies on asynchronous video interviews (Liff et al., 2024) and essay scoring (Campion et al., 2016) using AI or machine learning to score - with a score of 1 being a perfect correlation, and 0 being no correlation – in a study of 99,411 evaluated interview responses). Interpreting and explaining a candidate's score is then essentially a task of describing the Competency being measured and the level at which the candidate scored. See Appendix A for an example of a report provided to recruiter end users which includes individualized explanations of HireVue's test scores for each candidate.

Additionally, to help further explain assessment results beyond the Candidate Feedback Report, we tweak model inputs and measure changes in the resulting score to determine the relative importance of individual features. The result is an ordered list of input features and their relative strength (positive or negative). If each word were analyzed separately, it would be possible to deduce high-level patterns in topics that top performing candidates displayed (e.g. the word "team" is a strong positive input for the teamwork model). However, as noted above, often the meaning of a word will change depending on its placement in a sentence. In our CUSTARD model, individual words are not treated independently, and each word is understood in context. Therefore, in order to explain our models in context,

Explainability of the models and its limitations

we take a similar approach as above, but instead of looking at words, we look at the effect on model scores of dropping individual sentences and phrases. Once we have a set of example sentences and their relative effects (positive or negative), we analyze these phrases for patterns and topics that have large effects on model scores (e.g. “handle change very easily” and “deal with challenges” are scored as important features in the adaptability model). The results show that our models are well-aligned with the BARS used by human evaluators to create the training data.

Finally, candidates are provided with the option to contact the hiring company who is the controller of their data and application (this is a configurable option in the system). In their email communications to the candidate following the interview/assessment, many companies will inform the candidate they can contact the recruiter concerning any follow up questions they might have. Please see an example text of this communication in the above section entitled ‘What is the user journey for interview candidates using our AI platform?’

When and how is the AI system tested?

How did we test the AI in our video interviews and game-based assessments?

Our video interviewing is subject to robust testing to ensure that it accurately and reliably predicts a candidate’s competency scores. To evaluate the performance of our models, we use test data not previously seen by our models. We predict candidates’ scores on this test data using the relevant models and compare these predicted scores against their respective human scores to get an estimate of model accuracy.

How do we test for and avoid or mitigate bias?

Once a competency model has been chosen by a customer, and before it is used to assess any actual candidates, we test it for adverse impact and other metrics related to fairness. As noted above, we consider there to be adverse impact when applicants from one or more protected groups (e.g. gender, ethnicity) are selected at substantially different rates. The categories we check include some of those listed by the U.S. Equal Employment Opportunity Commission, which are generally the same as other anti-discrimination criteria in other countries (for example the UK Equalities Act 2010 ‘protected characteristics’). For example, if the passing rates of one ethnic group is significantly lower than another group then we investigate to determine which input variables have a strong relationship with ethnicity, and less impact on the model performance. Following such investigation we adjust the relevant variables to eliminate such bias.

All models used in our assessments must pass all our adverse impact tests while maintaining satisfactory performance in identifying the relevant competencies. More information on our efforts to identify and remove bias can be found here: How to Advance Diversity Hiring with Big Data.

How do we keep the AI system up to date?

There are two aspects to our updating programme: (1) adjusting AI systems used by individual customers to control for fluctuations in candidate populations over time, and (2) general updates to all of our systems to improve their functioning, efficiency and accuracy.

General Updates to All Our AI Systems

We update our models about once a year, based on a combination of human consultation and model tuning. These updates may be based on various different requirements but broadly speaking they fall into two types:

Updates at customer requests

We hold individual review meetings with each customer to discuss the functioning of our assessments, typically on a quarterly basis. In addition we hold renewal meetings to make more significant changes, typically on an annual basis. At these meetings, the following types of changes might be requested:

- Feedback from customers (e.g. we may be requested to shorten the questions asked).
- Changes in the role being recruited for, thus changes in the AI-based assessment to match the new role.
- A decision by the customer to measure different competencies or adjust the weighting of each competency to reflect changes in the role requirements for various reasons (e.g. a shift from employees working in the office to working from home).

Updates based on technological and scientific developments

As explained above, our AI systems combine insights from different scientific fields, in particular data science and AI, as well as IO psychology. Since these fields are constantly developing, we work hard to ensure that our systems continue to reflect the latest science. We cannot update our systems daily for such developments, as we need to go through various stages of detailed work to determine whether and if so, how best to implement any changes (which includes looking at potential impacts).

These updates are made based on:

- Improvements to technologies we use for assessing candidates (for example NLP models), whether those developed by third parties such as Rev.AI, or our own internal models.
- Adverse Impact data (where available).
- Changes based on HireVue's own test data produced by paid volunteer mock candidates (e.g., Panel studies for game assessments).
- HireVue's upgrades are based on developments in IO psychology and other scientific research (further details of which are discussed below).

Whenever we make an update to our technology, we put it through the same rigorous checks and procedures as when it was first developed (detailed above) to ensure that the system remains effective and trustworthy.

Updates to Customer Systems

We also monitor customer systems after they have been deployed, on an ongoing basis. Our checks include the following:

Distribution of Scores

When our systems are properly calibrated, we see a mostly unchanging distribution of scores. If we start to see the results skewing higher or lower, this could indicate a problem in the AI model or a significant change in the applicant population due to candidate sourcing or job market fluctuations.

Although the model is static once deployed for each interview cohort, because these models are using live data the results of assessments can vary depending on the input. Normally we would expect to see a 'Bell Curve' shaped distribution of scores, with a small number very low, a small number very high, and the majority clustered around the middle. If we started to see that Bell Curve distribution shifting (for example more candidates than

How do we keep the AI system up to date?

we would expect getting very low scores) then that might be a reason for checking whether any updates need to be made to the AI system.

To maintain fairness, the AI model used to judge any given cohort of candidates remains the same. For example, if a company wants to hire 50 workers over a period of 2 months, the model for those 2 months would be static whilst those 50 people were being selected. If, 6 months later, the same company wants to hire another 50 candidates and an updated model is available, this updated model could be used with this new cohort of candidates. There would be no danger of unfairness since each candidate pool would be competing under the same rules and criteria.

Adverse Impact Monitoring

In addition to the major efforts we take to avoid any bias in the design of our AI system, we also monitor and seek to correct any adverse impact in the system after models have gone live. Our processes are similar to those used pre-deployment, but unlike testing the systems using historical candidate data, when we seek to correct adverse impact once our systems are in use, we are dependent on recent candidate demographic data provided to us by our customers – specifically as to whether individual candidates have relevant protected characteristics.

Where a customer provides us with such diversity data then we are able to run an analysis on the candidates' scores against the protected characteristic data, to check whether candidates with those characteristics score higher or lower than average, and if so on which parts of the assessment. We do this by amending the model (as described previously) to ensure that there is no significant adverse impact towards groups within the particular sample, then re-launch the model. Where customer data can be used, we would typically perform such checks on an as-needed basis for each customer. In most cases these checks will be undertaken annually.

Text Scoring Requirements

For our AI text scoring to run on a candidate's interview question response, there are two primary requirements. First, the candidate must consent to the use of AI to score their responses. Second, our AI system must detect enough content in a candidate's response. A candidate's response may not include enough content for various reasons - some may simply fail to give an adequately long answer or respond to a question in the allotted time, some candidates may fail to speak clearly enough to be understood, and some may have technical issues with their microphone input. If either or both of these requirements are not met, the candidate's response will not be scored with AI and instead it will be flagged for human review.

We have found that these requirements are not met by a small proportion of candidates in any given candidate cohort. The rate at which our text scoring requirements are not met is monitored for each customer. If we start to see numbers consistently exceeding the expected rate, then we investigate and take corrective action.

Who is responsible for monitoring?

Multiple HireVue teams are involved in monitoring:

- Product Manager and Engineering team (the technical implementers of a system): monitors incidental score drift, unexpected changes in thresholding rates, and completion rates.
- IO Psychology team: monitors scores of competency models, and account-level adverse impact and validity concerns.
- Data Science team: supports Product and IO Psychology teams in scoring-related inquiries.

What happens when we spot a potential issue?

We maintain a special internal procedure for the rare occasions if system or scoring anomalies arise. Steps include pausing interview scoring based on approval by HireVue directors, communication with all relevant

How do we keep the AI system up to date?

HireVue personnel, and communication with all affected customers. We retain all raw data necessary to rescore interviews when problems are found and fixed. We have a policy of not altering any candidate scores, even if we think they may not properly reflect our competency criteria, unless we have first spoken to the relevant customer.

Who might be affected?

Who are our stakeholders?

Our stakeholders can be split into three main groups, within which there are further sub-categories:

Customers. Our customers are the companies which use our services. Key customer groups include: management executives and board members; personnel involved in the hiring process, such as human resources, diversity and inclusion officers; legal departments; and existing employees.

Candidates. Within the overall pool of candidates (and potential candidates) for any given job, there are certain further groups: ethnic minorities; those with atypical speech; older candidates; those with neurodiverse characteristics (for example autism); people with disabilities that might affect their ability to undertake interviews / game-based assessments (for example, those who have visual or speaking impairments); and those with other characteristics protected by employment law.

External Groups. Of the external groups likely to take an interest in HireVue's activities, the key players are: governments and regulators; and AI ethics, civil liberties and social justice NGOs and campaigners. Examples of the NGOs we have worked with to improve our systems and processes are mentioned in the following section.

How do we manage risks?

What (actual and perceived) risks are there to stakeholders from our AI use, and how do we manage them?

Fears of Baked-in Bias

It is a common criticism of AI assessments of any kind that there may be some form of bias hardwired into the relevant system.

As set out above, HireVue takes extensive technical steps to check for and mitigate bias at all stages of its AI system lifecycle, both generally and on specific customer projects. HireVue's systematic bias reduction should be assessed in comparison to traditional hiring processes, where multiple studies have shown that in traditional hiring processes (1) there is often systemic bias against particular groups, and (2) such bias can be hard to detect on an individual basis, and therefore hard to mitigate or avoid.

In order to ensure the robustness of our internal bias mitigation processes, we have commissioned external audits from respected third-party experts – discussed further below in the section on Oversight. As such, we are confident that our consistent focus on reducing bias means this will be more of a perceived than actual risk, especially when compared to human-only processes.

Missing Exceptional Candidates

It could be argued that our systems might be less capable of giving high scores to atypical or exceptional candidates, whose answers to interview questions are radically different from most candidates. However, there are several points to make here. First, it cannot be guaranteed that those candidates would have been selected by humans, if their answers were so unusual. Second, our AI is an aid to human decision-making but does not replace it. Personnel from our customers are always able to review the actual video interviews before making a hiring decision, and could always call the relevant candidate for an in-person interview. Third, all of our AI assessments involve multiple questions

How do we manage risks?

including other assessment content such as game-based assessments. These features provide a useful control against exceptional candidates being missed, since: (1) it is unlikely that a candidate would answer every single question in an extremely unusual but otherwise brilliant way, and (2) as there is only one correct way of approaching the games, a customer could decide to follow up with a person who has scored exceptionally well on the games but poorly on the interviews.

Accessibility

Some candidates may have concerns over the accessibility of our assessments, if physical or other disabilities prevent them from answering interview questions or completing any other assessment content When our assessments are deployed, candidates are provided with information in advance on what each part of the test will involve, and asked if there is any reason why they would not be able to take such tests – an 'Accommodation Request'. An example is shown below when an assessment includes AI-scored interview questions, but this can be altered at a customer's request to provide further information.

"This interview contains questions you must answer within a given time limit as well as a game-based assessment. If you require extra time due to a qualified disability, click the "Request Accommodations" button below to relax the time limits. If you require a different form of accommodation, please reach out to your contact at BB Data. You will be given extra time to answer the questions. Due to their strict time requirements, the game-based assessment section will not be presented. A representative from [COMPANY NAME] will contact you if any further action is required from you after completing this interview. Otherwise, click cancel to return to the previous screen."

Any Accommodation Requests for any adjustments or modifications to the standard selection process are directed to the relevant department of the employer managing the hire process. The employer will then decide whether an exemption will be granted (this process is outside of HireVue's control). As an example, for an AI-scored interview, candidates might be scored on their recorded answers with relaxed time limits.

We also engage with various stakeholders, including individuals from groups representing neurodiverse candidates, and continue to work with these groups to ensure that so far as possible our testing is fair and open to all types of candidates. External participants in our work have included: (1) Integrate Autism Employment Advisors, representing neuro-atypical candidates, and (2) Jopwell and re:work training, representing minority candidates. We are committed to further engagement with these and other representative stakeholder groups in the future. As an example of our continued work and research in this area, we recently published a peer-reviewed scientific paper relating to research of candidates on the autistic spectrum undertaking our game-based assessment (Willis et al., 2021), available here: https://www.mdpi.com/2079-3200/9/4/53.

Addressing Anomalies

As noted above, we have systems in place to detect and address anomalies which arise after our systems go live, on a customer-by-customer basis. If a customer raises a concern about a scored interview, it is reviewed by experts in our science teams and proper action and rescoring is executed if necessary.

Internal and external oversight

Internal Oversight

Diverse Expertise

Our AI based assessments are built by joint work amongst the Science team which consists of our Data Science, IO Psychology, and Product/Engineering departments. Each of these departments involves individuals with different expertise and backgrounds, who are able to contribute unique perspectives and oversight to the process. Moreover, there is no 'traditional background' for our employees, particularly within data science – where individuals may have degrees in fields including applied physics, economics and finance, astrophysics and bioengineering, as shown in their bios, available here: https://www.hirevue.com/our-science.

The Science Team at HireVue meets regularly to discuss various topics such as 'Research Updates', 'Show and Tell' discussions and 'Assessment Planning'. Our collaborative approach between these different fields helps us to promote strong internal scrutiny of our systems, and to avoid 'groupthink'.

Internal Training

Science Team: In addition to the varied external expertise of each individual team member, HireVue also carries out extensive internal training on the build and use of our AI. For example, a new IO hire will undergo 4-6 weeks of intensive reading, lecture, and partner discussions with other members of the IO team to ensure they are experts in the various aspects of our AI assessments, consulting procedures, and analytic techniques followed.

Rater Team: As detailed above, the raters who help to train our AI models undergo extensive internal training, separate to that of HireVue's staff.

Other Teams: Other departments whose work is connected to the assessments in less direct ways (sales, implementation consultants, and account managers) all undergo onboarding activities that include AI assessment overviews given by our data science, IO and product teams. Finally, additional webinars, video recordings, and marketing collateral are provided to new joiners as they become more involved with our assessment product and customers. We also undertake ongoing team training to ensure that all relevant HireVue staff are kept up to date with the AI systems in use and development.

Internal Accountability

We maintain strong internal accountability structures covering each stage of the AI system's development. The IO consultant on a particular project and the data science team member who builds the relevant models are responsible for properly validating models and ensuring that there is no significant adverse impact. The product manager for the Assessments product and an engineer from the Automated Assessments team take responsibility for scoring errors.

Above each team lead, each of our respective departments in the science organization (data science, IO psychology and product) have Executive Leaders who oversee their respective teams and technical work. These Executive Leaders are directly responsible to the CEO, who in turn reports to the HireVue Board of Directors.

External Oversight

Expert Advisory Board

Our Expert Advisory Board consists of outside experts in relevant fields of IO psychology, Assessments, Legal, and AI and meets twice yearly. Its current members include a partner at a major law firm, IO psychology specialists working in industry and a distinguished professor of management. In addition to the Expert Advisory Board's planned meetings, individual members are often consulted on an ad hoc basis by different HireVue departments, based on their specialist knowledge.

Internal and external oversight

External Audits

Finally, we have obtained external audits from various different expert organizations. These have included:

- AI Technology: a detailed analysis of our AI technology and algorithms and how they affect a range of diverse stakeholders. Conducted by O'Neil Risk Consulting & Algorithmic auditing, the audit concludes that *"[HireVue] assessments work as advertised with regard to fairness and bias issues."* More information on the audit and its recommendations can be found here:

  https://www.hirevue.com/press-release/hirevue-leads-the-industry-with-commitment-to-transparent-and-ethical-use-of-ai-in-hiring

- IO Psychology: a detailed analysis of the psychological measurements and job fit frameworks used in our AI-based assessments. Conducted by Landers Workforce Science LLC, the audit concludes: *"In general, HireVue reaches or exceeds industry standards for the creation of high-stakes assessments, and this audit exposed no weaknesses that critically undermine HireVue's approach."* More information on the audit and its recommendations can be found here:

  https://www.hirevue.com/press-release/independent-audit-affirms-the-scientific-foundation-of-hirevue-assessments

- AI Procedures: an independent review of the consulting procedures and design controls we place on our software when using AI-based assessments. Conducted by a traditional audit firm, the audit assessed whether HireVue did not meet, met, or exceeded relevant standards in 10 areas. The audit concluded that of these 10 areas, HireVue exceeded the standards in all areas apart from 2 – where it only 'met' standards. In order to exceed standards in these 2 areas audit recommended, (1) improved recording of customer approvals, and (2) better linking (in a macro-database) of our work to external databases and competency frameworks for particular roles - such as those operated by O*NET under the sponsorship of the U.S. Department of Labor/Employment and Training Administration. Overall this was a 90% achievement rate of the standards assessed. Since this audit concluded, HireVue has addressed the two 'met' standards by instituting controls (document trails and folders) for customer approval of assessment models and released a macro-database linked to O*NET that is used for every customer assessment implemented.

- Methods for Measuring Bias: an audit to analyze our data sets and the procedures we follow to measure and mitigate discrimination or bias. The auditing of our assessment solutions is conducted continuously by our Data Science and IO staff, as well as by third-parties on an annual basis.

These audits have confirmed a very high level of fairness. We are always striving to improve though, and where recommendations for improvements have been made, we have already implemented or are in the process of implementing them.

How do we use and protect your personal data?

Data Privacy

Our systems collect and record different types of candidates' personal data on behalf of our customers; the potential employers. In such cases, we are acting as a 'data processor' and are collecting and processing candidates' personal information on their behalf and in accordance with their instructions. That is an important distinction because it means that the majority of the obligations under the EU and UK's GDPR are required to be fulfilled by our customers, and not HireVue. In addition to explaining HireVue's processes, one of the roles of this Explainability Statement is to assist our customers in fulfilling their obligations as 'data controllers'.

We never collect sensitive data, like protected health information, financial information and we also do not collect dates of birth. The full details of our Privacy Policy are set out on our website: https://www.hirevue.com/privacy#what-does-hirevue-do

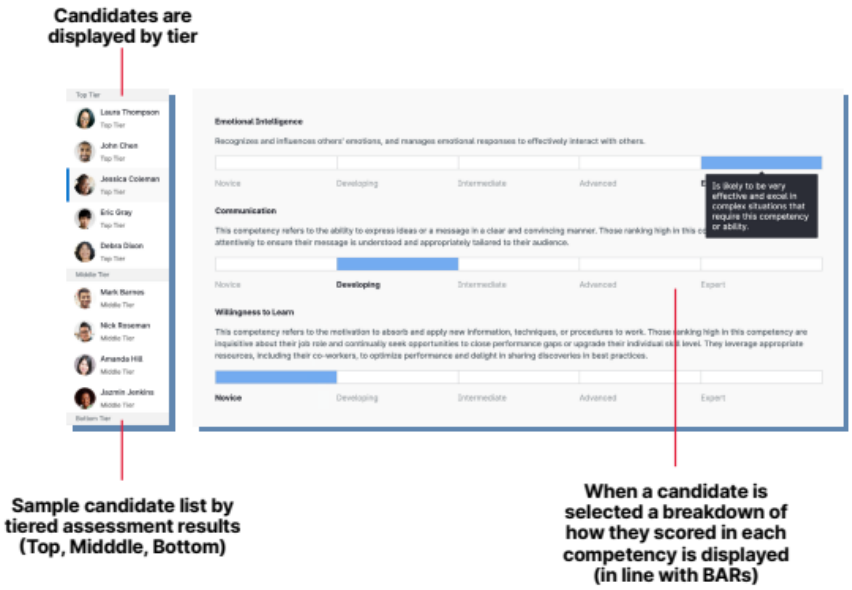
Data Protection and Resilience

We maintain state of the art cyber-security protections for all data (personal or otherwise) stored on our systems.

Appendix A: Candidate Assessment Report Examples

Recruiter Report Example. The image below provides an example of the information shared with recruiter end users to provide insight into candidate assessment scores.

Screenshots showing click downs when viewing candidate assessment results to aid customer's decisions on job candidates.



Appendix A: Candidate Assessment Report Examples

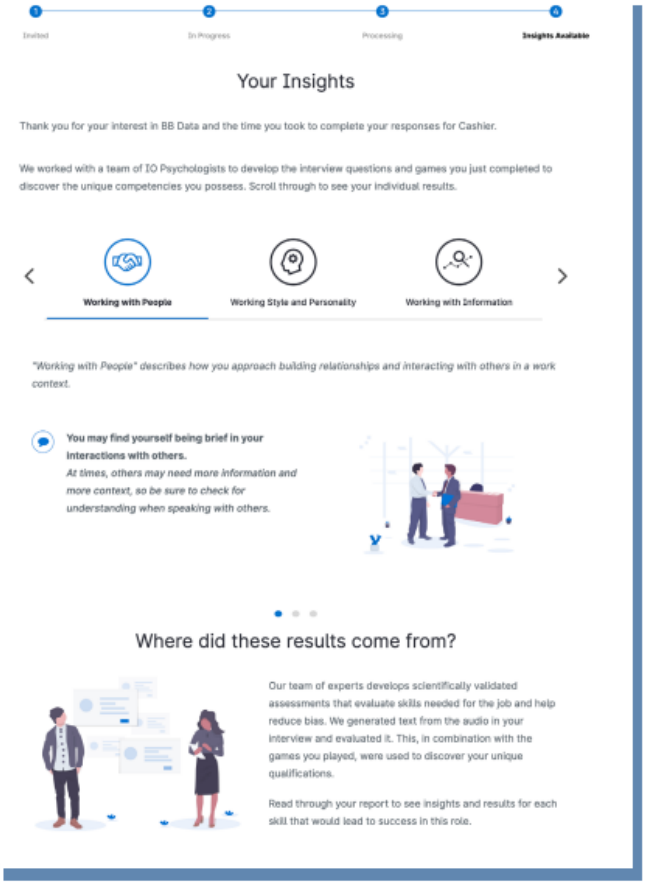
Candidate Feedback Report Example. The image below provides an example of the information that can be shared with candidates to provide insight into their assessment performance.

Sample Candidate Feedback Report. Sent to Candidate from Customer.

Note: The same Candidate Report Template is provided whether or not a candidate is successful in progressing to the next stage.

Feedback in each Competency is based on the assessment results (low, average, or high) whereby the feedback wording for the candidate reflects the competency result.

Candidates are not told if they were low, average, or high on the competency. Wording is positively stated.



Appendix B: Behaviorally-Anchored Rating Scale for Communication Competency and Example of Rated Statements

This competency refers to the ability to express ideas or a message in a clear and convincing manner. Those ranking high in this competency communicate in a respectful and considerate manner and are able to listen attentively to ensure their message is understood and appropriately tailored to their audience.

Key Behaviors	Novice	Developing	Intermediate	Advanced	Expert
	Proficiency Level Rating Guidelines:				
	Candidate is unlikely to be successful in situations requiring this competency.	Candidate is likely to demonstrate this competency in simple situations or in a limited capacity.	Candidate is likely to demonstrate this competency well, but may need assistance in more difficult situations.	Candidate is likely to demonstrate this competency effectively in moderate to complex situations.	Candidate is likely to demonstrate this competency with extreme effectiveness in moderate to complex situations.
	Behavioral Examples at Novice, Intermediate, and Expert Proficiency Levels:				
Delivers Clear & Concise Message	Message delivered is disorganized, lacks a clear explanation of purpose and importance, and is not delivered in a logical sequence.		Message delivered is reasonably organized, has a clear purpose and importance, and is delivered in a logical sequence.		Message delivered is well organized, has a clear explanation of purpose and importance, and is delivered in a logical sequence.
Exhibits Professionalism	Communicates in a way that is not considerate of others		Communicates in a way that is polite toward others.		Communicates in a respectful and considerate manner.
Uses Appropriate Communication Style	Fails to adhere to accepted communication styles appropriate to the media being used.		Mostly adheres to accepted communication styles appropriate to the media being used.		Skillfully uses other accepted communication styles appropriate to the media being used.
Shares Information	Does not openly communicate ideas effectively. Interacts and shares information only when asked.		Communicates clearly and effectively with teammates and others. Shares information in a timely manner.		Proactively seeks improvement in communication skills within the work or academic place. Encourages others by facilitating an environment that fosters sharing information and knowledge.

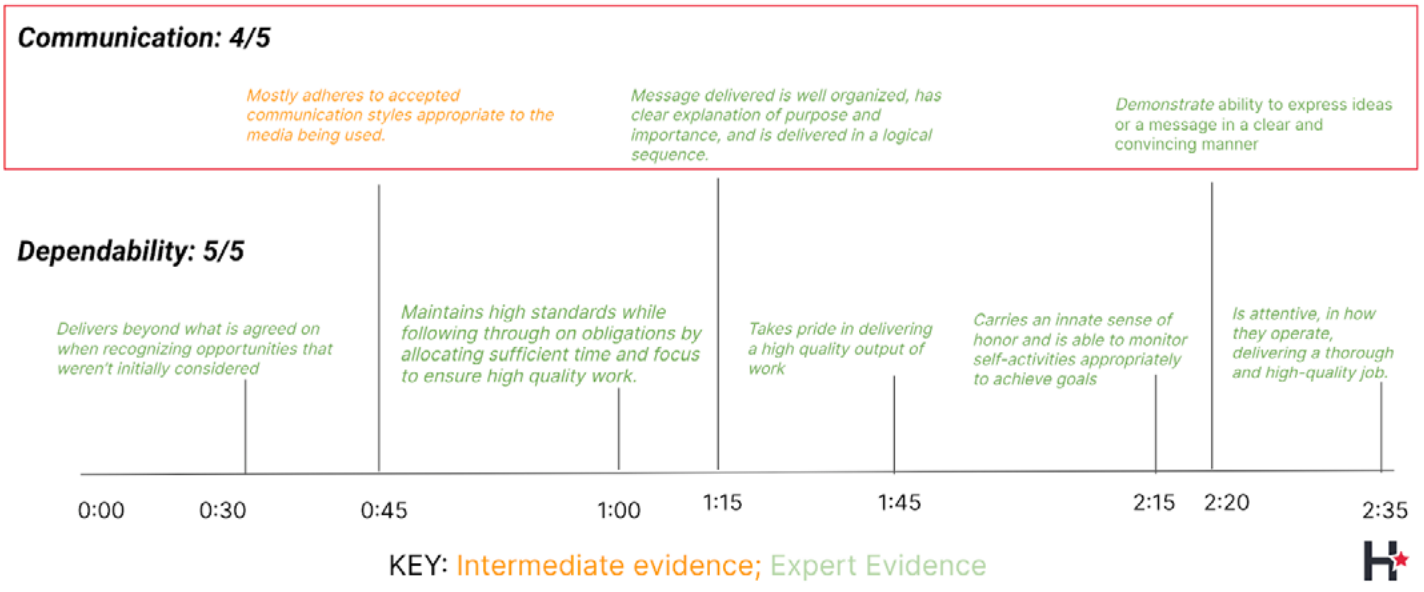
Continued on next page

Appendix B: Behaviorally-Anchored Rating Scale for Communication Competency and Example of Rated Statements

Verifies Understanding	Fails to understand the message and does not seek feedback or clarification to ensure understanding and correct interpretation.		Mostly understands and correctly interprets messages from others. Seeks feedback or clarification when there are challenges with comprehension.		Has a detailed understanding of messages from others and proactively seeks feedback and follows up on the message to confirm correct interpretation.
Engages Others	Fails to engage with others. Excessively dominates group discussions to promote their own ideas. Suppresses or ignores other people's ideas or feedback.		Maintains attention to others in group discussions and shows interest in their ideas and feedback.		Engages with others in group discussions and has a free flowing exchange of dialogue by proactively seeking the ideas of others.
Tailors Message to Audience	Does not effectively adjust their message to match the needs of the audience.		Moderately adjusts their message to match the needs of the audience.		Is highly effective at communicating by matching their message to the needs of the audience.

BARS Guide Expert Human Rater Evaluations

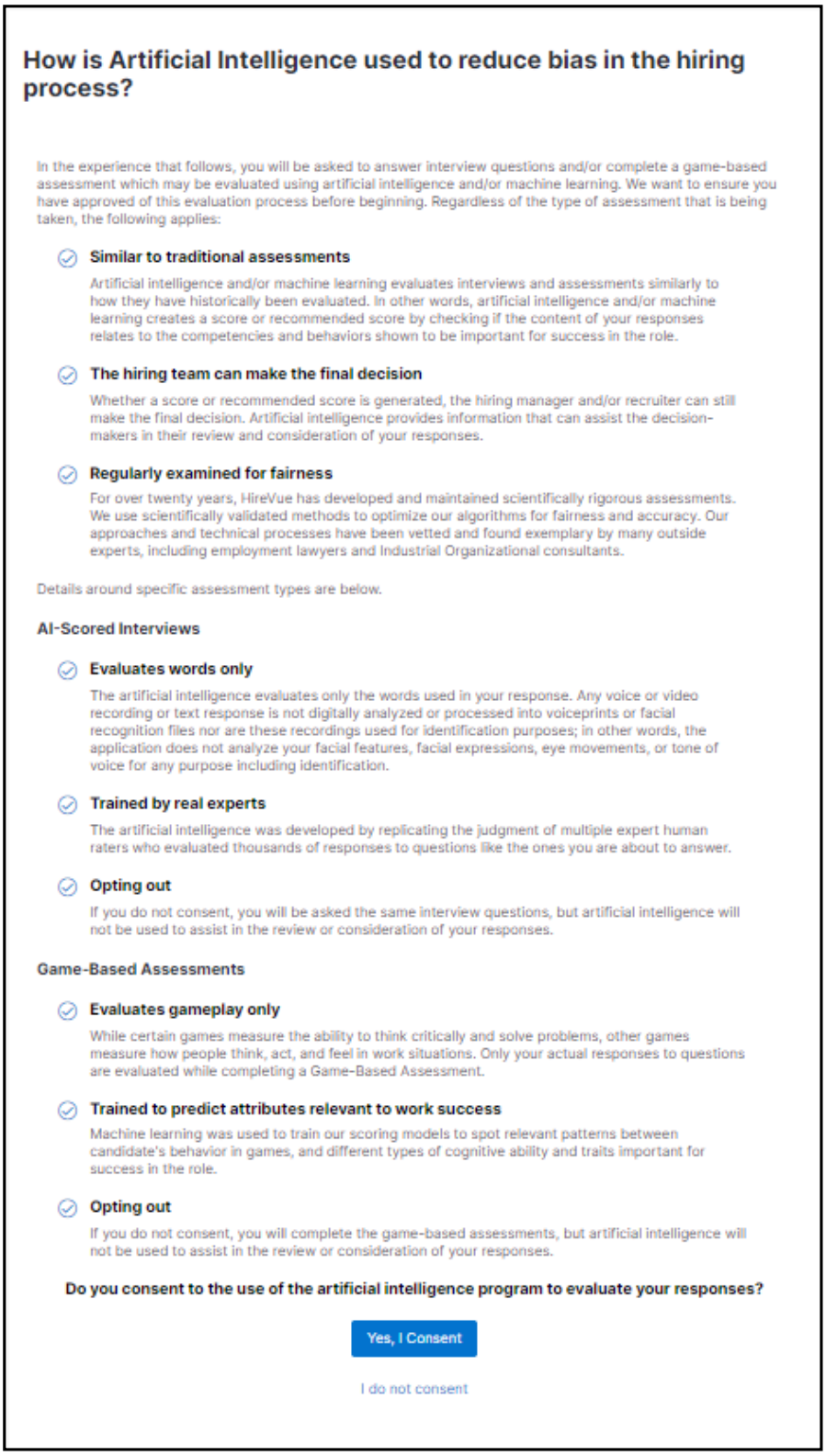
Q3: *Tell me about a time when you struggled to keep a commitment you made due to other important priorities. Please describe the specific situation, your actions, and the outcome.*



Appendix C: Applicant Demographic, Role Level, Industry, and Geographic Representation for Latest Rater Study Sample

Gender		
Male	40,629	49%
Female	42,538	51%
Age		
Under 40	67,668	85%
Over 40	12,246	15%
Race/Ethnicity		
White	34,966	42%
Black	16,662	20%
Hispanic	21,723	26%
Asian	9,806	12%
Level		
Non-Manager	76,658	77%
Manager	22,703	23%
Industry (Top 10)		
Healthcare	9,438	9%
Retail	13,100	13%
Hospitality, Recreation, & Leisure	7,403	7%
Insurance	6,435	6%
RPO & Sourcing	4,403	4%
Transportation	6,051	6%
Banking & Finance	9,352	9%
Consulting Services	5,934	6%
Food & Beverage	2,956	3%
Technology	4,837	5%
Geographical Regions		
Africa	2,138	2%
Asia	6,593	7%
Australia and New Zealand	5,083	5%
Europe	8,515	9%
Latin America and the Caribbean	1,063	1%
Northern America	71,715	75%

Appendix D: AI-Interview & GBA Consent Form



How is Artificial Intelligence used to reduce bias in the hiring process?

In the experience that follows, you will be asked to answer interview questions and/or complete a game-based assessment which may be evaluated using artificial intelligence and/or machine learning. We want to ensure you have approved of this evaluation process before beginning. Regardless of the type of assessment that is being taken, the following applies:

Similar to traditional assessments

Artificial intelligence and/or machine learning evaluates interviews and assessments similarly to how they have historically been evaluated. In other words, artificial intelligence and/or machine learning creates a score or recommended score by checking if the content of your responses relates to the competencies and behaviors shown to be important for success in the role.

The hiring team can make the final decision

Whether a score or recommended score is generated, the hiring manager and/or recruiter can still make the final decision. Artificial intelligence provides information that can assist the decision-makers in their review and consideration of your responses.

Regularly examined for fairness

For over twenty years, HireVue has developed and maintained scientifically rigorous assessments. We use scientifically validated methods to optimize our algorithms for fairness and accuracy. Our approaches and technical processes have been vetted and found exemplary by many outside experts, including employment lawyers and Industrial Organizational consultants.

Details around specific assessment types are below.

AI-Scored Interviews

Evaluates words only

The artificial intelligence evaluates only the words used in your response. Any voice or video recording or text response is not digitally analyzed or processed into voiceprints or facial recognition files nor are these recordings used for identification purposes; in other words, the application does not analyze your facial features, facial expressions, eye movements, or tone of voice for any purpose including identification.

Trained by real experts

The artificial intelligence was developed by replicating the judgment of multiple expert human raters who evaluated thousands of responses to questions like the ones you are about to answer.

Opting out

If you do not consent, you will be asked the same interview questions, but artificial intelligence will not be used to assist in the review or consideration of your responses.

Game-Based Assessments

Evaluates gameplay only

While certain games measure the ability to think critically and solve problems, other games measure how people think, act, and feel in work situations. Only your actual responses to questions are evaluated while completing a Game-Based Assessment.

Trained to predict attributes relevant to work success

Machine learning was used to train our scoring models to spot relevant patterns between candidate's behavior in games, and different types of cognitive ability and traits important for success in the role.

Opting out

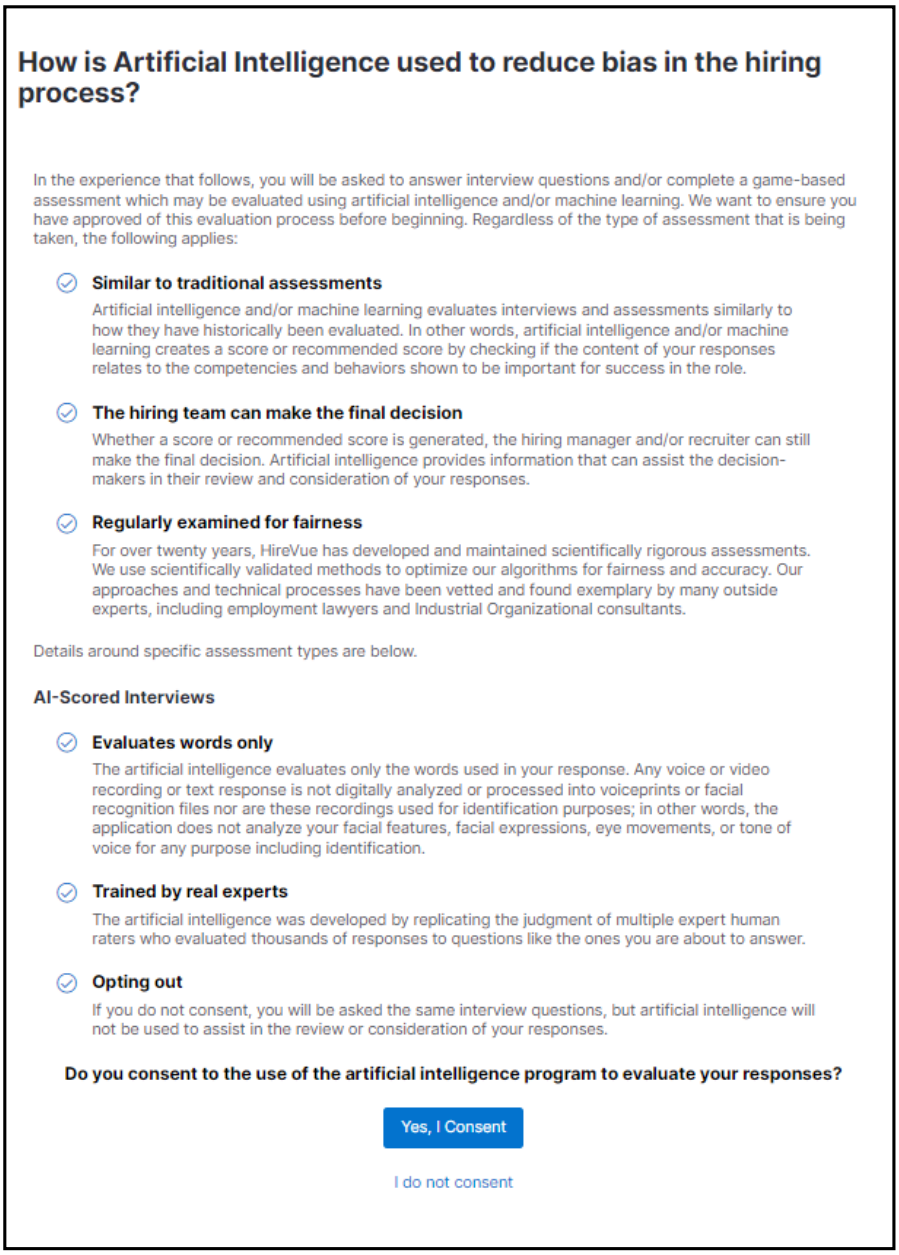
If you do not consent, you will complete the game-based assessments, but artificial intelligence will not be used to assist in the review or consideration of your responses.

Do you consent to the use of the artificial intelligence program to evaluate your responses?

Yes, I Consent

I do not consent

Appendix D: AI-Interview & GBA Consent Form



How is Artificial Intelligence used to reduce bias in the hiring process?

In the experience that follows, you will be asked to answer interview questions and/or complete a game-based assessment which may be evaluated using artificial intelligence and/or machine learning. We want to ensure you have approved of this evaluation process before beginning. Regardless of the type of assessment that is being taken, the following applies:

Similar to traditional assessments

Artificial intelligence and/or machine learning evaluates interviews and assessments similarly to how they have historically been evaluated. In other words, artificial intelligence and/or machine learning creates a score or recommended score by checking if the content of your responses relates to the competencies and behaviors shown to be important for success in the role.

The hiring team can make the final decision

Whether a score or recommended score is generated, the hiring manager and/or recruiter can still make the final decision. Artificial intelligence provides information that can assist the decision-makers in their review and consideration of your responses.

Regularly examined for fairness

For over twenty years, HireVue has developed and maintained scientifically rigorous assessments. We use scientifically validated methods to optimize our algorithms for fairness and accuracy. Our approaches and technical processes have been vetted and found exemplary by many outside experts, including employment lawyers and Industrial Organizational consultants.

Details around specific assessment types are below.

AI-Scored Interviews

Evaluates words only

The artificial intelligence evaluates only the words used in your response. Any voice or video recording or text response is not digitally analyzed or processed into voiceprints or facial recognition files nor are these recordings used for identification purposes; in other words, the application does not analyze your facial features, facial expressions, eye movements, or tone of voice for any purpose including identification.

Trained by real experts

The artificial intelligence was developed by replicating the judgment of multiple expert human raters who evaluated thousands of responses to questions like the ones you are about to answer.

Opting out

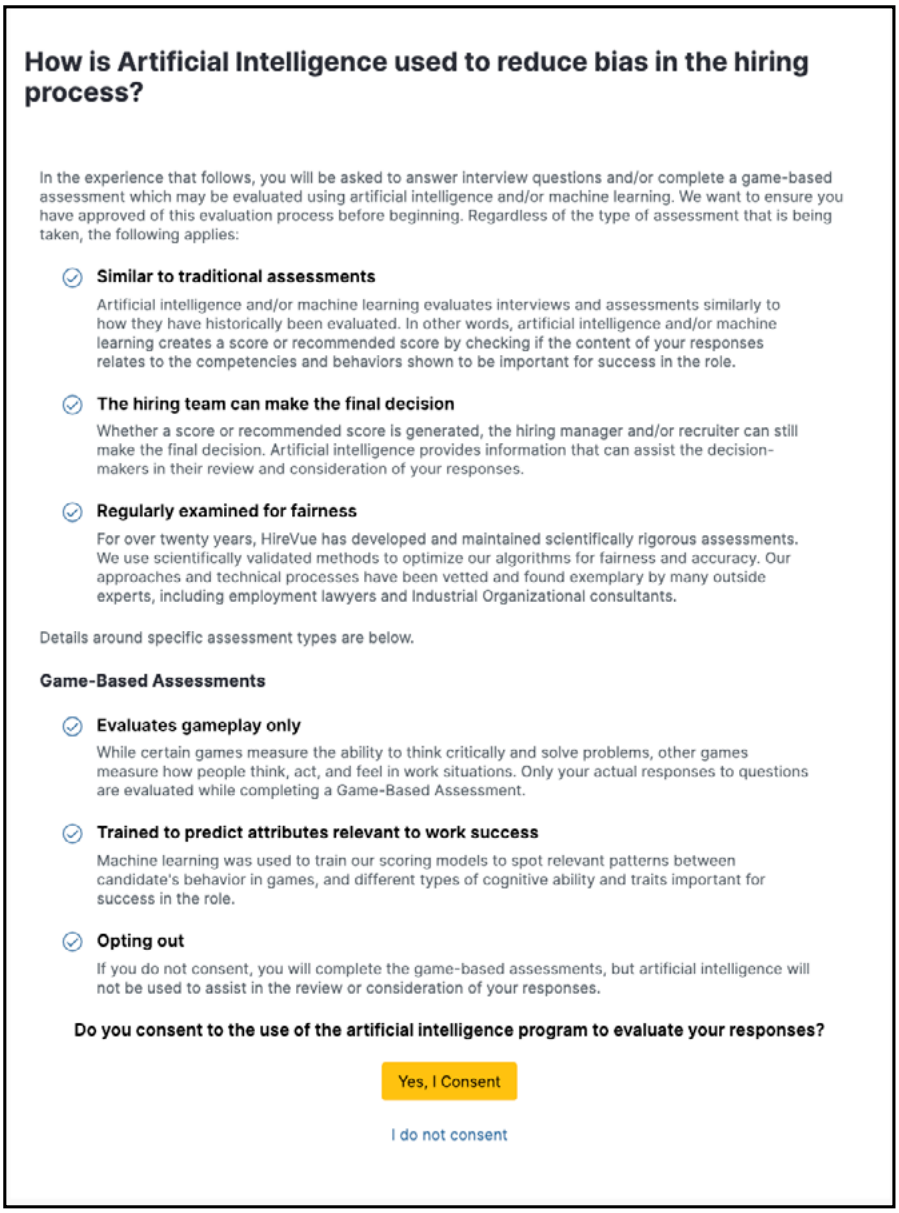
If you do not consent, you will be asked the same interview questions, but artificial intelligence will not be used to assist in the review or consideration of your responses.

Do you consent to the use of the artificial intelligence program to evaluate your responses?

Yes, I Consent

I do not consent

Appendix D: GBA Only Consent Form



How is Artificial Intelligence used to reduce bias in the hiring process?

In the experience that follows, you will be asked to answer interview questions and/or complete a game-based assessment which may be evaluated using artificial intelligence and/or machine learning. We want to ensure you have approved of this evaluation process before beginning. Regardless of the type of assessment that is being taken, the following applies:

- **Similar to traditional assessments**
  Artificial intelligence and/or machine learning evaluates interviews and assessments similarly to how they have historically been evaluated. In other words, artificial intelligence and/or machine learning creates a score or recommended score by checking if the content of your responses relates to the competencies and behaviors shown to be important for success in the role.
- **The hiring team can make the final decision**
  Whether a score or recommended score is generated, the hiring manager and/or recruiter can still make the final decision. Artificial intelligence provides information that can assist the decision-makers in their review and consideration of your responses.
- **Regularly examined for fairness**
  For over twenty years, HireVue has developed and maintained scientifically rigorous assessments. We use scientifically validated methods to optimize our algorithms for fairness and accuracy. Our approaches and technical processes have been vetted and found exemplary by many outside experts, including employment lawyers and Industrial Organizational consultants.

Details around specific assessment types are below.

Game-Based Assessments

- **Evaluates gameplay only**
  While certain games measure the ability to think critically and solve problems, other games measure how people think, act, and feel in work situations. Only your actual responses to questions are evaluated while completing a Game-Based Assessment.
- **Trained to predict attributes relevant to work success**
  Machine learning was used to train our scoring models to spot relevant patterns between candidate's behavior in games, and different types of cognitive ability and traits important for success in the role.
- **Opting out**
  If you do not consent, you will complete the game-based assessments, but artificial intelligence will not be used to assist in the review or consideration of your responses.

Do you consent to the use of the artificial intelligence program to evaluate your responses?

Yes, I Consent

I do not consent

References

Campion, M. C., Campion, M. A., Campion, E. D., & Reider, M. H. (2016). Initial investigation into computer scoring of candidate essays for personnel selection. Journal of Applied Psychology, 101, 958-975.

Liff, J., Mondragon, N., Gardner, C., Hartwell, C., & Bradshaw, A. (2024). Psychometric properties of automated video interview competency assessments. Journal of Applied Psychology. Advance online publication. https://doi.org/10.1037/apl0001173

Rottman, C., Gardner, C., Liff, J., Mondragon, N., & Zuloaga, L. (2023). New strategies for addressing the diversity-validity dilemma with big data. Journal of Applied Psychology, 108(9), 1425 -1444. https://doi.org/10.1037/apl0001084

Willis, C., Powell-Rudy, T., Colley, K., & Prasad, J. (2021). Examining the use of game-based assessments for hiring autistic job seekers. Journal of Intelligence, 9(4), 53. http://dx.doi.org/10.3390/jintelligence9040053