

## 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 Hiring Experience Platform™ work, including when, how, and why we use this technology to facilitate our customers in making their hiring decisions. It is separate from our privacy policy, which is available [here](#). 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 ethical development of AI along with data security and privacy to be core values. In

addition to 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).

If you have any queries, we can be contacted at 9800 South Monroe Street, STE 810, Sandy, UT 84070.

---

### WHAT IS OUR PLATFORM?

---

#### **Our AI**

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 hiring platform features video interviewing, assessments and conversational AI. Hirevue has hosted more than 26 million video interviews and 5 million AI-based candidate assessments for over 700 customers around the globe.

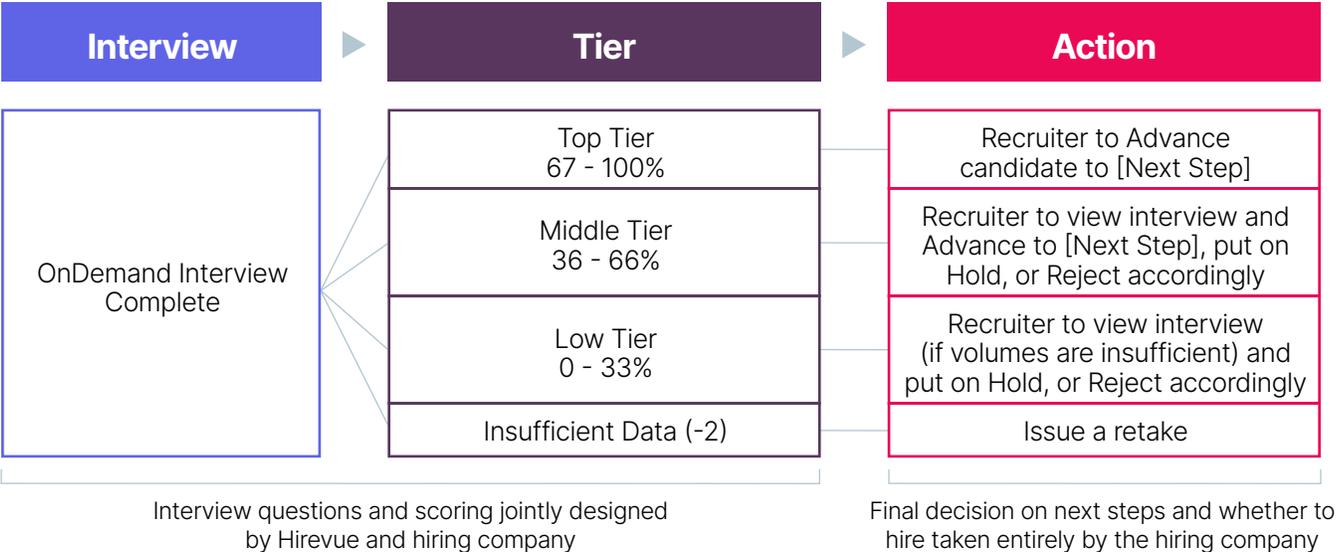
Hirevue offers a broad suite of AI assessments, including video interviewing and online game-based challenges. These can be combined in a 'modular' fashion for specific roles. Using more than one type of assessment allows us to measure different competencies. For example, we can combine an assessment to test teamwork skills with another to test problem-solving. Customers who choose Hirevue's modular system work with our Industrial/Organizational (IO) Psychology experts to decide the competencies required for each role they are looking to fill. We ensure best practice interview approaches are followed and

use state of the art technology to transcribe answers, which we then analyse to assess each competency in accordance with our customers' instructions.

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

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.

# EXPLAINABILITY STATEMENT



### How do recruiters use the results from Hirevue's AI?

As the diagram above shows, Hirevue provides a tool which assists employers in finding the best candidates in the recruitment 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 each candidate with an individualised Candidate Insights Report setting out their assessment scores (**see Appendix A**). In addition, where the hiring company requests it to do so, Hirevue can also provide further information on the underlying data in respect of each candidate.

# EXPLAINABILITY STATEMENT

## WHY DO WE USE AI?

### Choosing the Right Candidate

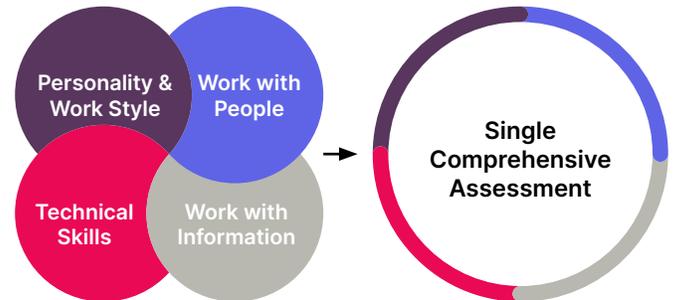
The aim of any hiring process is to find the right candidate(s) for a job. There can be hundreds of factors involved in making good hiring decisions. Even entry-level, hourly jobs require a unique combination of competencies, cognitive abilities and personality types – few of which will be clear from 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 behaviour in organisations and the workplace. Our methods use AI to produce a single comprehensive assessment of each candidate, which organisations can then use to make better, more informed hiring decisions.

### Our Video Interview Assessments

Our Interview Assessments have multiple advantages over traditional assessment techniques, both for candidates and employers:

- **Avoiding 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 is well-documented but can be difficult to spot until it is too late to correct. By contrast, with our AI systems we are able to detect and avoid or mitigate any bias with great accuracy. 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).

In designing AI-based assessments, we can minimise any data points that lead to a bias to ensure proportional outcomes for all relevant groups. For instance, if we see that a disproportionate number of men score higher than women, we can determine what behavioural measures are causing the bias and change the algorithm to remedy that score difference. In traditional job assessments, removing questions that are found to cause adverse impact is usually not feasible without significant accuracy reductions, because every question is important for the accuracy of the assessment. By contrast, because our AI interview assessments assign weights to all words, word types, statements, and contextual phrases that are predictive of the competency being rated or measured, we can minimise

# EXPLAINABILITY STATEMENT

---

## WHY DO WE USE AI?

---

any data that contributes to adverse impact with minimal impact on the accuracy of our assessments. For example if the word “aggressive” is predictive of the competency ‘Drive for Results’, yet more men than women use the word, it might add bias against women in the competency score. By lowering the weight assigned to the word “aggressive” the bias against women will be decreased and the prediction of the competency still remains.

- **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 assessments 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 by the algorithm according to the same tests. Our system avoids the danger of a particular human interviewer scoring a candidate well or badly 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 even 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 organisation 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 for a variety of individuals.
- **Better candidate experience.** Unlike a traditional interview, our AI interview assessments can allow candidates to take multiple attempts to answer each question, if they feel that the first attempt did not go as well as they would have wanted. The number of re-takes allowed is set by the employer, but our recommended setting is 1-2 retries in addition to the first attempt. In addition, our AI interview assessments allow for clear feedback to be given to every single candidate on where they scored well and badly, as soon as the interview is finished – something which would be time consuming and difficult for human interviewers to do for every applicant. We have provided a sample ‘candidate insights report’ – a document which can be provided by recruiters to candidates as feedback - in **Appendix A** to this Explainability Statement.

# EXPLAINABILITY STATEMENT

## WHY DO WE USE AI?

- **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 interview assessments provide excellent insight into attributes like interpersonal skills, communication skills, personality traits, and overall job aptitude. Our AI interview assessments 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 organise 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 which would be more difficult to do if only human interviewers were involved, since it is very difficult for a human to disentangle the different attributes of an interviewee.
- **Costs savings for customers.** Hirevue's customers are able to obtain major cost savings using our AI, which reduces the time taken to hire, staff time, and travel costs compared to using human interviewers to screen all candidates. In our experience, organisations using Hirevue experience a significant Return on Investment.

### Our Game-Based Assessments

Our game-based assessments are a series of short online games. Each game takes only a few minutes to complete, and different games give insight into a range of cognitive skills, including numeracy, problem-solving, and attention, as well as non-cognitive abilities, such as personality, empathy and influence. We also offer a post-assessment report that makes it clear to candidates what is being tested for in relation to the position for which they are applying (for example, working memory skills for Call Centre or Registered Nurse roles).

In addition to minimizing bias and gathering richer data, our game-based assessments have several advantages over traditional questionnaires:

- **Speed.** Traditional multiple choice cognitive skills tests last 30-45 minutes, as opposed to 6-15 minutes for our tests.
- **Flexibility.** Our games adapt in real time based on a candidate's performance. If a candidate completes one task in a game, the next task 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 unitary test.
- **Improved candidate experience.** Based on feedback by 1.5 million candidates who have taken Hirevue's interviews: 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. By contrast, long multiple-choice tests can be boring and demoralising for candidates.

# EXPLAINABILITY STATEMENT

## 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 what that text means and (3) assessing/scoring the candidate's answers following expert human rater evaluations of answers to the same competency-based question.

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, accents and inflections. 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 labelled 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.

# EXPLAINABILITY STATEMENT

## HOW DID WE DESIGN OUR AI?

### How do we use AI in video interviews?

#### 2. Understanding words and sentences

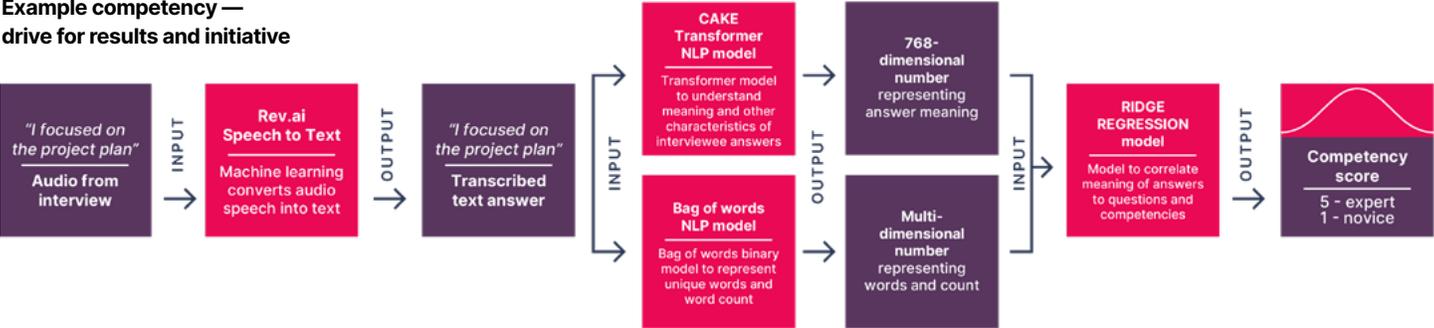
Second, based on the transcribed text, we use a form of AI called 'natural language processing' (NLP) in order to understand candidates' answers, as summarised in the diagram below.

We have developed our own NLP model, 'CAKE', 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. CAKE starts with this base model and is further fine-tuned on interview data (as explained further below). The language analysed by CAKE 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. Specifically, the system training happens in two steps: (1) by predicting masked words in a large number of documents, the system learns about the English language, and (2) further refining this model with interview transcripts to understand the nuances of language which might be expressed in job interview scenarios. The output of the neural network is a numerical value – known as a

'vector' – which the model has attributed to the particular answer to an interview question that has passed through the 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 complexity makes it more difficult for candidates to "game" the video interview process by mentioning particular words or phrases in their responses. CAKE is capable of understanding individual words as well as their context. 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."

In addition to the CAKE model, we also used an older and simpler NLP method, called "binary bag of words." This looks at all the words in the answer, with no consideration of the grammar and order of the words. These features add to our ability to explain since we can look at the relative weights of different words in the model. For example, we see things like saying the word "team" contributes positively to a candidate's score for the teamwork model.

### Example competency — drive for results and initiative



# EXPLAINABILITY STATEMENT

## HOW DID WE DESIGN OUR AI?

### How do we use AI in video interviews?

#### 3. Assessing and scoring each candidate

Third, once the CAKE model has understood and assigned numerical values to the candidate's response to an interview question, this numerical value is fed into a 'ridge regression model' (a machine learning system) along with the "binary bag of words" analysis. The ridge regression model has been trained to identify responses of a similar nature and then score those responses against the customer's chosen competencies.

We have developed a separate AI model and set of questions for each competency. A sample of competencies we can test include adaptability, problem solving, communication and willingness to learn. There are over 20 which we can cover, and we are increasing this list over time based on scientific research and our own data.

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 modelled the most accurate way of fairly and consistently rating structured interview responses without the unconscious bias injections of normal human interviewers.

Hirevue's AI system scores each of the candidate's responses according to a BARS for each competency. The BARS guides or content are based on data from thousands of real-life interviews, covering a diverse range of interviewees and job types and the scoring

uses five rating levels, from 'novice' to 'expert'. In **Appendix B** we provide the BARS used for our 'Communication' competency. Additionally, 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.

The models we use to assess candidates through interviews have been trained using expert human evaluations of structured interview responses; the 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 assessments 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 standardised 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 3.5 years, have drawn upon 125,000 interview evaluations, which include over 500,000 applicant video interviews scored and evaluated for bias. More specifically:

# EXPLAINABILITY STATEMENT

## HOW DID WE DESIGN OUR AI?

### How do we use AI in video interviews?

- We collected scoring data from interviews for different levels of roles, type of companies, and geographic locations.
- We trained teams of around 30 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 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 ridge 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 'ridge' regression model helps to ensure that the algorithm generalises well to unseen data, rather than 'overfitting' to the training data. 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 generalise as well (which, in this case, could mean it is unable to recognise different but similar candidate responses).

### How do we use AI in game-based assessments?

We use AI in our game-based assessments to assess a candidate's cognitive ability. 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, and 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 – in this case the candidate's game performance and their cognitive ability. 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 model was trained in a similar way to our ridge regression model 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 behaviour in games, and different types of cognitive ability.

# EXPLAINABILITY STATEMENT

## 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. A typical assessment will consist of 3-6 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 behavioural responses related to a specific competency (e.g., customer service). The games are designed to measure general mental abilities (e.g. numerical reasoning) or five 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): Recommended 60 seconds
- Interview Question Retries (Yes/No): Recommended 1-2
- 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 Top/Middle/Low labels
- Competency and Assessment Percentile Score (On/Off): Recommended Off
- Data Retention: Recommended 2 years, but follows company policy

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 (e.g., Top/Middle/Bottom Tier bands) as compared to all the other candidates that completed the assessment for that position. Additionally, a candidate feedback report can also be generated by the recruiter/manager and sent to the candidate upon request (see **Appendix A** for a sample candidate report).

# EXPLAINABILITY STATEMENT

## WHAT IS THE USER JOURNEY FOR INTERVIEW CANDIDATES USING OUR AI PLATFORM?

### System Configuration Impacts Candidate Experience

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 analysed 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. Rev.ai had a somewhat higher WER (meaning that there were more errors) by between 12-22% depending upon the speaker’s country of origin (which we believe is due to differences in accents) (e.g., 12% WER for Canadian accent, and 22% WER for participants from China). However, the alternative services WERs ranged from 17% (Canada) to 45% (China). Thus, we found, with these tests, that Rev.ai’s system was considerably more accurate than the alternative transcription services, even when tested against speakers from different countries.

# EXPLAINABILITY STATEMENT

## 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. Over the last 3.5 years we have conducted various 'Rater Studies' to build and improve the algorithms and check/mitigate them for bias (adverse impact). In these Rater Studies we have expert ratings on ~30,000 video interviews across 15 competencies which yields ~125,000 total expert human ratings collected. The following table 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 500,000 applicant records to check for and mitigate group differences based on gender, age, and race/ethnicity.

Gender		
Male	13,443	48%
Female	14,380	52%
Age		
Under 40	23,966	86%
Over 40	3,857	14%

Race/Ethnicity		
White	9,936	36%
Black	4,770	17%
Hispanic	9,316	33%
Asian	3,800	14%

### Public AI Models

As noted above, our NLP model, CAKE, is adapted from RoBERTa, which was designed in 2019 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 STATEMENT

## — EXPLAINABILITY OF THE MODELS AND ITS LIMITATIONS —

The scores provided by our ridge regression 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, in addition to individual input features for each word used (Bag of Words feature sets) in the answer, our CAKE model calculates an embedding (i.e. list of numbers, or vector) to each answer. This embedding reflects the weightings that the neural network has attributed to the transcript. The CAKE embedding reflects 768 dimensions or features of the input sentence. We know the bag of words and CAKE features correlate with expert ratings of interview responses (in technical terms, there was an average correlation  $r=0.66$  – with a score of 1 being complete correlation, and 0 being no correlation – in a study of 60,183 people). 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 the Candidate Insights Report which provides individualised explanations of Hirevue’s test scores, for each candidate.

Additionally, to help further explain assessment results beyond the Candidate Insights 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 analysed 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 CAKE model, individual words are not treated independently, and each word is understood in context. Therefore, in

order to explain our models in context, 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 analyse these phrases for patterns and topics that have large effects on model scores. 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 to 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 on the ‘User Journey’.

# EXPLAINABILITY STATEMENT

## 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. Our testing uses a technique called K-Fold cross-validation. Cross validation is a well-known statistical method for evaluating the performance of predictive models, and is commonly used in both the machine learning and psychometric testing fields.

In summary, K-Fold cross-validation involves the following steps:

- We split the data on which the AI system is being trained (i.e. a group of candidates' video interview responses) into various (K) 'batches' (known also as 'folds').
- We then take one batch out of the sample (representing, for example, one hundred candidate responses) and we predict the candidates' scores in that 'out of sample' batch based on the patterns that the system learned from the remaining batches.
- We then compare the predicted score of the 'out of sample' batch of candidates against their respective human scores to get an estimate of the system accuracy.
- We then repeat this process many times until all batches have predictions that were ascertained during their 'out-of-sample' step. Overall, this process can give a good idea of how an algorithm trained on the entire set would be expected to perform on unseen data.

### 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 real 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 significantly 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.

In addition to checking for adverse impact, we permanently remove features in data that are consistently associated with specific protected groups and which are not related to performance at work. For example, the pronunciation of certain words could be correlated with ethnicity. The model is then re-trained without the identified features. 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: <https://www.Hirevue.com/resources/whitepaper/the-next-generation-of-assessments>

# EXPLAINABILITY STATEMENT

## 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.

# EXPLAINABILITY STATEMENT

## HOW DO WE KEEP THE AI SYSTEM UP TO DATE?

### **Dynamic Updates to Customer Systems**

We also monitor customer systems after they have been deployed, on an ongoing basis. This monitoring takes place through: (1) automatic alerts if a particular metric goes 'out of bounds', (2) whenever a code change is made (for example to update a library, or to accommodate hardware updates) this is tested against the old code to check that it does not affect any candidate scores, and (3) our personnel undertake manual checks (assisted by automated reporting) for significant group or bias differences that are unexpected (e.g. female candidates scoring significantly lower than men). Our manual checks include the following:

#### *Distribution of Scores*

When our systems are properly calibrated, we see a mostly unchanging distribution of scores between candidates who do well and badly on our tests. 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 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 then an updated model might be used, but there would be no danger of unfairness based on variations between different interviewers, 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 analysis on the candidates' scores against the protected characteristic data, to check whether candidates with those characteristics score better or worse than average, and if so on what 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.

# EXPLAINABILITY STATEMENT

## HOW DO WE KEEP THE AI SYSTEM UP TO DATE?

### *Thresholding and Completion Rates*

Where our AI system cannot detect a certain minimum level of content in a candidate's interview answer, it is not scored and is flagged for human review. Typically this occurs where fewer than five words are understood and recorded. We call the relative frequency of these events, across all those being interviewed for a particular job the "thresholding rate". A candidate's answer may fall below the threshold 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. We expect a certain level of thresholding issues in any given cohort. The thresholding rate is monitored daily for each customer. If we start to see numbers consistently exceeding the expected thresholding rate, then we investigate and take steps to correct any issues which might be causing this.

Likewise, if we start to see more than expected candidates failing to complete an assessment either through thresholding or other causes, this will also trigger an internal investigation, as it may suggest a technology issue.

### *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.
- Data science team: monitors score drift, validity, and adverse impact when launching new models.
- IO Psychology team: monitors scores of new competency models, and account-level adverse impact and validity concerns.

### *What happens when we spot a potential issue?*

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

# EXPLAINABILITY STATEMENT

---

## 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 groups within them are: 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 avoid or 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 (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 and typically also game-based

# EXPLAINABILITY STATEMENT

## HOW DO WE MANAGE RISKS?

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 tests, if physical or other disabilities prevent them from answering questions by video or completing the game-based assessments (for example, difficulties for a candidate in vision, using their voice or in coordinating their hands). When our AI 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’. Our standard wording is shown below, 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 an exemption from the relevant tests are directed to the relevant department of the employer which is managing the hire process. The employer will then decide whether an exemption will be granted (this process is outside of Hirevue’s control). For interviews that do not include games, 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, 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.

# EXPLAINABILITY STATEMENT

## INTERNAL AND EXTERNAL OVERSIGHT

### 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>.

### Internal Oversight

#### *Different 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 oversights 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,

# EXPLAINABILITY STATEMENT

## INTERNAL AND EXTERNAL OVERSIGHT

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.

#### *External Audits*

Finally, we have obtained external audits from various different expert organisations. 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 analyse our data sets and the procedures we follow to measure and mitigate discrimination or bias. This audit is on-going.

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.

# EXPLAINABILITY STATEMENT

## 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

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

**Candidates are displayed by tier**

The screenshot displays a candidate list on the left and a detailed competency breakdown on the right. The candidate list is organized into tiers: Top Tier (Laura Thompson, John Chen, Jessica Coleman, Eric Gray, Debra Dixon), Middle Tier (Mark Barnes, Nick Roseman, Amanda Hill, Jazmin Jenkins), and Bottom Tier. The competency breakdown for a selected candidate shows scores for Emotional Intelligence, Communication, and Willingness to Learn. A callout box highlights a specific competency level: 'Is likely to be very effective and excel in complex situations that require this competency of ability.'

Competency	Novice	Developing	Intermediate	Advanced	Expert
Emotional Intelligence					
Communication					
Willingness to Learn					

**Sample candidate list by tiered assessment results (Top, Middle, Bottom)**

**When a candidate is selected a breakdown of how they scored in each competency is displayed (in line with BARs)**

# APPENDIX A

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.

1 Invited 2 In Progress 3 Processing 4 Insights Available

## Your Insights

Thank you for your interest in BB Data and the time you took to complete your responses for Cashier.

We worked with a team of IO Psychologists to develop the interview questions and games you just completed to discover the unique competencies you possess. Scroll through to see your individual results.

< Working with People Working Style and Personality Working with Information >

*"Working with People" describes how you approach building relationships and interacting with others in a work context.*

**You may find yourself being brief in your interactions with others.**  
*At times, others may need more information and more context, so be sure to check for understanding when speaking with others.*

**Where did these results come from?**

Our team of experts develops scientifically validated assessments that evaluate skills needed for the job and help reduce bias. We generated text from the audio in your interview and evaluated it. This, in combination with the games you played, were used to discover your unique qualifications.

Read through your report to see insights and results for each skill that would lead to success in this role.

# APPENDIX A

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.



## Your Insights

Thank you for your interest in BB Data and the time you took to complete your responses for Cashier.

We worked with a team of IO Psychologists to develop the interview questions and games you just completed to discover the unique competencies you possess. Scroll through to see your individual results.



*"Working with information" describes how you approach new problems, make sense of the world around you, and adapt to new challenges.*

**You are able to quickly switch between tasks.**  
*Make use of your ability to adapt to changing demands and environments by engaging in diverse projects.*

**You easily process and manipulate numerical information.**  
*This is particularly useful when quickly making data-driven decisions. When communicating with others, include storytelling and visualization to explain numerical concepts.*

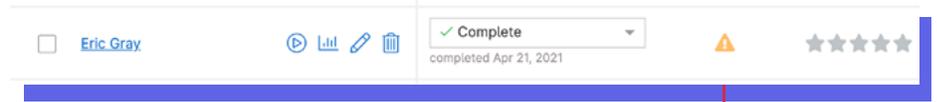


**You are quickly able to problem solve your way through most situations.**  
*You may enjoy taking on new challenges and giving yourself a chance to shine!*

**You have the capacity to readily process complex relationships in your mind.**  
*Others may struggle to keep up, so make sure you take the time to explain and draw out your ideas for them.*

# APPENDIX A

Thresholding: If Hirevue's system is unable to detect a minimum level of understandable content for a candidate's interview answer, then the customer employees reviewing the interview scores will be notified.



**The "warning" signal indicates there is an Insufficient Data Error**

# APPENDIX B

## Behaviorally-Anchored Rating Scale for Communication Competency and Example of Rated Statements

Communication					
This competency refers to the ability to express ideas or a message in a clear and convincing manner. Those ranking high in this competency 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.
Behavioural Examples at Novice, Intermediate, and Expert Proficiency Levels:					
Delivers Clear & Concise message	Message delivered is disorganised, lacks a clear explanation of purpose and importance, and is not delivered in a logical sequence.		Message delivered is reasonably organised, has a clear purpose and importance, and is delivered in a logical sequence.		Message delivered is well organised, has clear explanation of purpose and importance, and is delivered in a logical sequence.
Uses Proper Grammar	Uses improper grammar, syntax, or fails to adhere to other accepted communication conventions (pace, volume, diction, and mechanics appropriate to the media being used).		Uses mostly proper grammar and syntax, and mostly adheres to other accepted communication conventions (pace, volume, diction, and mechanics appropriate to the media being used).		Effectively uses proper grammar and syntax, and skilfully uses other accepted communication conventions (pace, volume, diction, and mechanics 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.
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.

# APPENDIX B

## 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.

### Communication: 4/5

Uses mostly proper grammar and syntax, and mostly adheres to other accepted communication conventions (pace, volume, diction, and mechanics appropriate to the media being used).

Message delivered is well organized, has clear explanation of purpose and importance, and is delivered in a logical sequence.

Demonstrates ability to express ideas or a message in a clear and convincing manner.

### Dependability: 5/5

Delivers beyond what is agreed on when recognizing opportunities that weren't initially considered

Maintains high standards while following through on obligations by allocating sufficient time and focus to ensure high quality work.

Takes pride in delivering a high quality output of work

Carries an innate sense of honor and is able to monitor self-activities appropriately to achieve goals.

Is attentive in how they operate, delivering a thorough and high-quality job.

0:00

0:30

0:45

1:00

1:15

1:45

2:15

2:20

2:35

# APPENDIX C

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

Gender		
Male	13,443	48%
Female	14,380	52%
Age		
Under 40	23,966	86%
Over 40	3,857	14%
Race/Ethnicity		
White	9,936	36%
Black	4,770	17%
Hispanic	9,316	33%
Asian	3,800	14%
Level		
Non-Manager	20,333	76%
Manager	6,577	24%

Industry (Top 10)		
Healthcare	7,008	17%
Retail	6,232	15%
Hospitality, Recreation, & Leisure	4,142	10%
Insurance	3,972	10%
RPO & Sourcing	3,207	8%
Transportation	3,177	8%
Banking & Finance	2,759	7%
Consulting Services	2,445	6%
Food & Beverage	1,369	3%
Technology	1,362	3%
Geographical Regions		
Africa	400	2%
Asia	1,174	5%
Australia & New Zealand	1,419	6%
Europe	2,331	9%
Latin America & the Caribbean	102	< 1%
Northern America	19,420	78%