Instructions for companies.

How to use testee lawfully and fairly: what it does, where it falls short and what stays your job.

Updated 15 November 2026

Arcolabs OÜ, registry code 17587452, Sepapaja tn 6, 15551 Tallinn, Estonia · hello@testee.ai

For owners, admins and reviewers at companies that interview candidates with testee. These are the instructions for use from Arcolabs OÜ, the provider of testee: what it's for, how its results come about, their limits and what you need to do as the employer.

Your own duties under the employment and data protection law that applies to you stay yours. These instructions aren't legal advice; check your process with your legal adviser.

What testee is for.

- **Intended purpose.** Structured first-round interviews for a role. An AI interviewer asks the questions your team set, in blocks (Conversation, Quiz, Coding, Whiteboard, Role-play, Review and Written response), and asks follow-ups. The analysis then reports how the candidate's answers covered the expectations you wrote, with quotes as evidence. Your team reviews every result and decides.
- **Who's who.** Arcolabs OÜ develops testee and offers it under its own name, so it is the provider of the AI system (EU AI Act Article 3(3)); it also processes candidates' data on your behalf, under the Data processing agreement. Your company uses testee under its authority, so it is the deployer (Article 3(4)) and the controller of that data. If you put your own name or trademark on testee, change it substantially or use it for another purpose, you become its provider yourself (Article 25(1)).
- **A high-risk AI system.** Under the EU AI Act, AI used to evaluate job candidates is high-risk (Annex III, point 4(a)), and we treat testee as such. Your duties as a deployer are summed up below.

- **Languages.** Interviews can be held in 13 languages. Arabic, Dutch, English, German, Hebrew, Italian, Spanish and Urdu have passed our interviewer and analysis evaluations; Brazilian Portuguese, French, Hindi, Indonesian and Simplified Chinese are in beta.
- **Training data.** testee runs on OpenAI's general-purpose models, which OpenAI trained on its own data. We don't train or fine-tune any model, and never with your data or candidates' data; we write the prompts and test every version on made-up interviews labelled by people.

What a result means.

- **A status for each expectation.** Covered, partial, covered after a prompt, incorrect or not covered, each with the quotes from the interview behind it.
- **The coverage score.** A number from 0 to 100, computed in code from those statuses, never by the AI: covered counts in full, partial and covered after a prompt count half, incorrect and not covered count nothing, and a must-have expectation weighs twice as much as a nice-to-have. It's labeled "for review, not a hiring decision".
- **Must-haves.** The report also shows whether every must-have expectation was covered.
- **Summary and follow-ups.** The AI summary lists strengths, gaps and suggested questions for a live follow-up interview.
- **What the AI never judges.** Emotion, accent, fluency, tone, confidence, personality, culture fit, appearance or any characteristic protected by law. There is no face or voice analysis and no biometrics, and no AI processes camera or screen recordings.

Limits you need to know.

- **Speech recognition makes mistakes.** Transcripts can contain errors, more so with strong accents, speech impairments, background noise or a poor microphone. The analysis is told to read such errors charitably; when a status depends on a few words, listen to the audio.
- **The analysis can be wrong.** A status can be too generous or too strict. Every status links to its evidence, and reviewers can change it; the score follows.
- **Your expectations shape the result.** Vague or unrelated expectations give vague results. Write each one as something a good answer shows, related to the job.

- **Beta languages.** In Brazilian Portuguese, French, Hindi, Indonesian and Simplified Chinese the interviewer and the analysis haven't yet met our evaluation targets. Review those results with extra care.
- **Generated interviews are drafts.** An interview generated from a job description or a role title opens in the builder for your team to check. Nothing is published without a person.

Measured accuracy.

We test every version of the analysis and the interviewer before it ships, on made-up interviews labelled by people (never real candidates' data). The figures below are for the current version, with 95% confidence intervals; a change to a model or prompt ships only when it meets the targets.

- **Coverage analysis.** On 139 interview blocks in 13 languages (752 expectations judged by the AI, 21 November 2026): the AI's status matched the human label for 90% of expectations (87–93%; target at least 80%), Cohen's kappa 0.86 (0.82–0.9; target 0.70). It gave more credit than the label for 5% (3–8%; target at most 5%) and less for 1% (1–2%; target at most 15%). 98% of evidence quotes were exact (94–100%; target 95%). The block's coverage score was off by 2.5 points on average out of 100 (1.6–3.5; target at most 8).
- **Interviewer.** In simulated conversations with scripted candidates, the interviewer asked one question at a time, never revealed an expectation or answer, logged hints correctly, completed blocks and wrapped up on time in every conversation (target at least 95%, and 100% for never revealing). Not praising or judging answers and not narrating its own steps are measured too; they're the weakest checks.
- **Robustness.** The tests include speech-recognition errors, paraphrases, keyword stuffing, confidently wrong answers, and instructions hidden in answers, code and whiteboard labels; the analysis didn't follow any hidden instruction.
- **Cybersecurity.** testee runs on Google Cloud in the EU, encrypted in transit and at rest. Every request is checked against the company it belongs to, keys live in Google Secret Manager, and our staff work in a separate admin zone with two-factor sign-in and an audit log. Candidates' answers, code and whiteboard labels reach the AI as data, never as instructions (tested above). More on the Trust page.

- **What can lower it.** Real interviews differ from test cases: poor audio, strong accents, speech impairments and beta languages make transcripts and statuses less reliable; vague expectations lower agreement. Accuracy on real candidates across accents and groups isn't measured yet; we will publish those figures here when it is.

What testee needs from you.

- Expectations and questions related to the job. Don't ask about health, family, religion, origin, age or other protected characteristics, and don't write expectations that stand in for them.
- Correct quiz options and realistic time limits for each block. The interviewer starts wrapping up at 80% of a block's time.
- Only the candidate data you need: a name and an email address are enough to invite someone.
- Interviews you've checked before you publish them, including the ones generated or written by the assistant.

Human review.

People make every hiring decision. To keep it that way:

- Choose reviewers who understand the role and know how testee works and where it falls short (the EU AI Act asks for this AI literacy). Share these instructions with them.
- Read the evidence, not only the score. Listen to the recording where a status matters, and change statuses that are wrong.
- Record the team decision (advance, hold or reject) with a note. Don't reject candidates on the score alone or automatically, including through the API, webhooks or a connected applicant tracking system.
- When a candidate asks for a person to review their result, do so and tell them the outcome (GDPR Article 22).
- Use the suggested follow-up questions to check gaps and doubts in a live interview.

Lifetime, resources and maintenance.

- **What it runs on.** testee is a hosted service: nothing to install. Candidates need a current browser, a stable internet connection and a microphone (or they type their answers); reviewers need a current browser.
- **Lifetime.** testee is in service for as long as your subscription runs. Each version of the models and prompts stays in service until we replace it; we replace a model before its maker retires it, and only after the new version meets the accuracy targets above.
- **Updates.** We deploy fixes and improvements continuously, with nothing for you to do. Changes that affect results or these instructions are announced as described under Versions and changes.
- **Your maintenance.** Review an interview's expectations when the role changes, check a sample of results against the recordings every month, keep reviewers trained, and check your retention settings once a year.

Integrity signals.

- **What they are.** When a company turns the setting on (Settings → Interview monitoring, off by default), the browser notes tab switches, focus changes, copy, cut and paste (where and how much, never what) and offline periods; the server adds reconnects, reloads, a second connection and a browser change, and, when they're recorded, the camera turning off or screen sharing stopping. Candidates are told before they consent, and the room shows "Activity notes on".
- **Signals, not proof.** Assistive technology, other apps a candidate needs and connection problems cause the same signals. They have no score and never change the coverage score. Never reject a candidate because of them: ask about them, and mark each one explained or dismissed.
- **Where you hire.** In the EU, the EEA and the UK, reading these browser events needs the candidate's consent (ePrivacy Directive Article 5(3), PECR regulation 6): they aren't needed for the interview the candidate asked for, and a candidate can rarely consent freely to an employer. Leave the setting off there; elsewhere, turn it on only where you need it and your local rules allow it.

Recording and consent.

- **Every interview.** Audio and transcript are recorded. Before anything is recorded, candidates see the AI notice and accept it; testee stores what they accepted, its version and when, and shows it on the results page and in the PDF.
- **Your own consent text.** Add yours in the interview's settings (Candidate consent), for example your retention period or a link to your privacy notice. It's shown next to the standard notice.
- **Camera and screen.** Off by default (Settings → Interview monitoring). Turn it on only when you need it. Candidates see what is recorded, why, who can watch it, how long it's kept and whom to contact, and agree before they start. Only Reviewers, Company admins and organization owners and admins can watch the video, and every viewing is in the audit log. For a candidate who asks for an interview without recording, turn it off on their invitation (Recording) before they start.
- **Other people.** Ask candidates to take the interview somewhere private, so no one else is recorded.

Accommodations.

- **Built in.** Candidates can type every answer instead of speaking, a choice on the interview's landing page. Captions are always on, everything works from the keyboard, and motion stops when the system asks for less.
- **Known limits.** Whiteboard blocks are hard to use with a screen reader, and the code editor's screen reader support is limited. Invite a candidate who needs it to a copy of the interview without them, or offer another way to answer.
- **More time.** When a candidate asks for more time, set Extra time (+25%, +50% or +100%) when you invite them, or on their invitation until they start. Every time limit grows by it and the candidate sees the longer times, with no label. An interview runs at most 60 minutes, extra time included, so only the shares that keep it within 60 minutes are offered; an interview that already takes 60 minutes can't get extra time, so invite a candidate who needs it to a shorter copy of the interview. The results say "Extra time" so reviewers read durations fairly; changes are in the audit log.

- **Tell candidates how to ask.** Say in your notice whom to contact (the accommodation offer template does this). Don't ask candidates whether they use assistive technology or about their health; ask what would help.

Telling candidates.

- Tell candidates before they're invited that an AI conducts and analyzes the interview, what it assesses and that people decide: GDPR Articles 13 and 14, and the EU AI Act for high-risk systems used on people. testee's own notice before the interview doesn't replace yours.
- New York City: under Local Law 144, a notice at least 10 business days before use and a published bias audit. On your careers page or in the job ad, or by email through testee: write your notice under Invitation emails and mark the invitation (or the import, or the ATS job) for a candidate who lives in New York City; the notice goes with the invitation, and the interview opens only after 10 business days.
- Illinois: under the Artificial Intelligence Video Interview Act, notice, an explanation and consent before the interview.
- The notice templates cover each case, with gaps for your details.

Records, retention and requests.

- **Retention.** Settings → Data retention: by default, candidates with no activity for 365 days are erased, and audio and recordings are deleted 180 days after the interview. Set what your policy needs, up to 10 years. Some laws ask for longer than the defaults: California's rules on automated-decision systems keep hiring records, including the system's data, for 4 years (1,461 days), and Colorado's for 3 years from 2027. Erasing a candidate deletes their data whatever the setting.
- **Logs.** testee records, for each interview, an AI log: the interview and version ids, the language, the models and prompt versions used, status, end reason, times and duration, the analysis' status, model and prompt version, and when people decided; never names, answers, scores or the decision. We keep it 190 days after its last entry. Your company's audit log (who did what in the dashboard) is kept 365 days, and results until the candidate is erased.

- **Reading and keeping the logs.** The EU AI Act asks you to keep the logs under your control for at least six months. Download the AI log as a CSV file (Settings → AI log → Download AI log), one row per interview with a header naming each field; times are in UTC. Download it at least every six months, and before you erase candidates or shorten retention, and keep the files for six months. Use it to check which version ran an interview, spot failed or unfinished interviews and analyses, and show when a person reviewed each result.
- **Access and erasure.** Candidates → row menu → Export data gives a candidate their data as a file, including each interview's analysis (the statuses with their explanations and quotes, the score and the summary) and your team's decision, never reviewers' comments or quiz answer keys; Erase candidate deletes it everywhere, recordings included, normally within minutes. That also meets Illinois' 30-day deletion deadline.
- **Sharing.** Share results only with people involved in the hiring decision. A shared PDF is outside testee and isn't deleted with the candidate.
- **Impact assessment.** Most companies in the EU need a data protection impact assessment for AI interviews. Start from the DPIA template.

Where not to use testee.

- As the only basis for rejecting a candidate, or to reject anyone automatically.
- To assess emotion, personality or any protected characteristic, or to monitor people's behavior beyond the interview.
- To evaluate your current employees for promotion, discipline or dismissal: testee is built and evaluated for hiring.
- With candidates under 16, or under the age of employment where you hire.
- For jobs in New York City before your bias audit summary is published.

The US and the UK.

- **New York City (Local Law 144).** If the coverage score substantially assists your decisions for a job in the city, testee counts as an automated employment decision tool. You need an independent bias audit within the year before use, its summary on your website, and

the notice 10 business days before (testee can send it with the invitation, as described under Telling candidates). Our results export and the API give your auditor the data.

- **Illinois (AI Video Interview Act).** Written for video interviews. Give the notice and get consent when you record the camera, and to be safe for every interview. Share recordings only with people needed to evaluate the candidate, and delete them within 30 days of a request. If AI analysis alone decides who gets an in-person interview, you also report demographic data each year; with human review, it doesn't.
- **Illinois (Human Rights Act).** Since 1 January 2026, tell candidates when you use AI in recruitment and hiring decisions, for every job in the state, not only video interviews. The AI interview notice template does this.
- **Disability.** The ADA in the US and the Equality Act in the UK require reasonable adjustments for applicants. Offer them as described above.
- **Colorado and California.** From 1 January 2027, Colorado asks for a notice before use, an explanation within 30 days after an adverse decision, and correction and human review on request; California's privacy rules on automated decision-making apply when a tool replaces or substantially replaces human decision-making, so the human review described here matters there too. Keep records as above.
- **Other places.** More states are adding rules on AI in hiring. Check the ones where you hire.

Your duties under the EU AI Act.

These apply in full from 2 December 2027; following them now keeps your process ready. What to do under Article 26:

- **Use it as instructed (26(1)).** Follow these instructions, including where not to use testee.
- **Human oversight (26(2)).** Name the reviewers for each interview: people who understand the role, know these instructions and have the authority to change a status or overrule the result. Give them time to read the evidence.
- **Input data (26(4)).** Keep what you control relevant to the role: the expectations, questions and quiz answers, and whom you invite.
- **Monitoring (26(5)).** Check a sample of results against the recordings every month, and watch for patterns such as one group of candidates scoring lower. If you think testee presents a risk to candidates' health, safety or rights, stop using the interview and tell us

at hello@testee.ai without delay; report a serious incident to us and to your market surveillance authority at once.

- **Logs (26(6)).** Keep the logs under your control for at least six months: download the AI log as described under Records, retention and requests.
- **Workers' representatives (26(7)).** Before you use testee at the workplace, for example to interview your own employees for an internal role, inform the workers' representatives and the employees concerned.
- **Impact assessment (26(9)).** Use these instructions and the DPIA template for your data protection impact assessment under GDPR Article 35.
- **Telling candidates (26(11)).** Tell candidates they'll be subject to a high-risk AI system that assists hiring decisions (the notice templates do this), and give a candidate who asks a clear explanation of the AI's role in a decision about them (Article 86): download it from their result with Candidate explanation (a PDF in the interview's language: the expectations, the AI's and your team's statuses with the explanations and their quotes, the scores, the summary and the decision).
- **AI literacy (Article 4).** Make sure the people who use testee know how it works and where it falls short, for example by having them read these instructions.
- **Authorities (26(12)).** Cooperate with the authorities; we help with the information we hold.

Problems.

Write to hello@testee.ai when a result looks wrong or unfair, a candidate complains, or something didn't work. Report anything that could have harmed a candidate at once. We answer every report and use them to improve testee.

Versions and changes.

- **The current version.** Interviewer: the model gpt-realtime-2.1 with the prompt interviewer-2026-11-21-quiz-note, transcription by gpt-transcribe. Analysis: the model gpt-6.1-sol with the prompt analysis-2026-11-21-workspace-data. Each interview's AI log names the versions that ran it.

- **Material changes.** A change that we know or should know is likely to change results, how testee performs for its purpose, its intended use, its limits or the safeguards described here (a new model, a prompt change that moves the measured accuracy, a new block type or a changed score) is a material change. We announce it by email to every organization owner and admin before it takes effect where we can, and at the latest when it does, and record it below with its date. Routine fixes and cosmetic changes aren't.
- **Records.** We keep each published version of these instructions (the dated PDF), the version identifiers, this list of changes and the notices we sent for at least three years.
- **Changes.** 21 November 2026: the interviewer gets the company's interview content and the candidate's own work (code, whiteboard labels, comments) in marked sections it treats as data, never as instructions; in a role-play the character reacts to what the candidate says and does, never to how they sound or seem; and when a topic's time runs out mid-answer, the interviewer acknowledges it before the next question (prompt interviewer-2026-11-21-marked-data). The analysis reads the session's automatic notes as quoting the candidate's work and judges words and actions, never the candidate's emotional state (prompt analysis-2026-11-21-workspace-data; the measured accuracy above is its run). Generated interviews no longer get expectations about how the candidate sounds or seems. Not a material change: the measured accuracy and the interviewer's checks stay within the noise. 20 November 2026: the interviewer ends an interview early only after asking the candidate whether they want to stop and hearing them confirm; a candidate who says they are panicking or facing an emergency is offered a short pause or the end, and is never asked why; a block the candidate asked to skip is recorded as skipped, not as finished by the interviewer; and conduct (insults, attempts to take the interviewer out of its role, long off-topic detours) is recorded by kind for each block (prompt interviewer-2026-11-20-distress-2). A material change: it changes how the interviewer runs and ends an interview. 20 November 2026: when the interviewer review finds that the interviewer gave away an expected answer, that item counts as covered after prompt (half credit) if the candidate addressed it only afterwards, and the score follows; the analysis no longer weighs the role's seniority (the expectations are written for it), and code removes any sentence of its own text that judges the person and any evidence that doesn't point at the candidate's real work (prompt analysis-2026-11-20-decided-items). A material change: it changes scores. 18 November 2026: an interview runs at most 60 minutes, extra time included, so only the extra time that keeps it within 60 minutes is offered; when the interviewer's connection is renewed or restored during an interview, the new connection gets a recap of the whole interview so far (its plan and the transcript) instead of only the current block's last lines (prompt interviewer-2026-11-18-

recap-cap). Both are material changes: the first limits an accommodation described here, the second changes how the interviewer continues an interview. 16 November 2026: extra time per candidate (+25%, +50% or +100% on every time limit) as an accommodation, set by your team on the invitation. 15 November 2026: integrity signals are off by default (they were on); the New York City AEDT notice can go with invitations, with a 10-business-day notice period; the AI log and its download; the measured accuracy, lifetime and maintenance, and the deployer duties added to these instructions. No change to how interviews are run or analyzed.