TECHNICAL ANNEX - METHODOLOGY

To begin, we signed up for the free trial as an employer for platforms Manatal and Talenteria. We began by creating an employer account, where we were asked to provide our name ("Privacy International"), location ("London"), company size (selected "1-100" from the drop-down options), email and phone number.

Then, we created 10 identical job openings for a Legal Officer role (Legal Officer 1, Legal Officer 2, etc.) and 10 identical listings for a Technologist role (Technologist 1, Technologist 2, etc.). The job descriptions for each role were pulled from previously published PI job listings.

When creating a job listing, both platforms offered text boxes for inputting the job description, as well as customisable criteria for the recruiter, such as selecting a language-speaking requirement or manually writing in skills.

For Phase 1 of our test we proceeded with the default set-up of just pasting in the job description without any further customisations, and in Phase 2 we went in and adjusted these customisations.

Two PI staff members served as the test candidates. Researcher #1, whom we will call "Test Candidate Legal Officer" (TCLO), would be dedicated to applying to the Legal Officer listings, and Researcher #2, whom we will call "Test Candidate Technologist" (TCT), would be dedicated to applying to the Technologist listings.

Researcher #1 wrote their own CV for the Legal Officer role (we will call this the "Standard CV"), and then they also used ChatGPT to generate a "ChatGPT CV" for the Legal Officer role. Researcher #2 did the same for the Technologist role.

The Standard CV is a bare-bones CV (education and work experience) drafted by the candidates themselves, and the ChatGPT CVs were generated using the following prompt: "My name is [name] and I am applying for this job with Privacy International. My previous employment includes [list of previous jobs]. Please create a CV using the information from this posting", which was accompanied by copy-pasting the entire job listing.

Both types of CVs are based on real and true information of the PI staff researchers.

The goal of these two types of CVs is to test the platforms' ability to discern human text and to find any hypothetical preference the algorithm might have for keyword scanning in an application.

Phase 1

For Phase 1, we tested the CV screening and AI scoring features of both platforms without any additional customisations:

1. Test Candidate Legal Officer (TCLO) applied once to each of the 10 Legal Officer listings, and Test Candidate Technologist (TCT) applied once to each of the 10 Technologist listings, using their Standard CVs. (E.g., TCLO 1 applied to the Legal Officer 1 listing, TCLO 2 applied to the Legal Officer 2 listing - repeated all the way to

TCLO 10 and the Legal Officer 10 listing.) The application process simply involved us uploading our CV on the job application page for each listing.
2. Next, we had all ten Legal Officer test candidates apply to just the one Legal Officer 1 listing and all ten Technologist test candidates apply ten times to the one Technologist 1 listing.
3. The test candidates repeated Step 1 using their ChatGPT CVs.

The purpose of Step 1 is to see how individual applicants would perform for individual job listings, and to see whether they may or may not be given variable AI scores despite using the same CV and applying to listings with identical job descriptions.

The purpose of Step 2 is to see whether scores are variable for an identical candidate applying multiple times to the same job listing.

The purpose of Step 3 is to see whether the algorithm might favour an AI-generated CV that is filled with generic keywords that match the job description over a human-written CV filled with specific examples of work experience.

Phase 2:

Next, we sought to test the customisation features that each platform offered.

For Manatal, we went into each job listing's "AI Recommendations" tab (as the employer), which provided a drop-down view of "recommended" candidates in the overall job pool along with their percentage match score for the current job. In the AI Recommendations tab, we could further customise a range of criteria like adding a language requirement or adding/removing education levels. We added the following United Nations languages criteria: "Chinese, English, Arabic, Russian, French, or Spanish-language proficiency."

Every time a criterion was adjusted, the page would refresh to show the list of recommended candidates based on the newly adjusted criteria, which would also dynamically refresh the candidates' match scores based on the new criteria. Note that at the time of testing, Manatal stated on its website that its AI features are currently in Beta mode and subjected to changes and updates.

We sought to explore whether candidates' AI scores would change significantly with even the slightest modification to the criteria, and whether the changes were reflected consistently across identical candidates.

For Talenteria, we tested the following customisations:

A. Screening Rules
B. Custom Knockout Questions

Note that to test these features, we created the same identical job listings for Legal Officer and Technologist as we did above, as these customisations had to be done at the job creation stage. We then applied to these new customised job listings using a similar procedure to Phase 1:

1. All ten Legal Officer test candidates applied once to each of the 10 Legal Officer "customised" listings, and all ten Technologist test candidates applied once to each of the 10 Technologist "customised" listings.
2. Repeated using ChatGPT CVs.

The "Screening Rules" feature allowed us to set numerical thresholds for what AI Match Scores would automatically filter a candidate into the "accepted", "rejected" or "interview" folders.

We set the Screening Rules for both the Legal Officer and Technologist roles to the following thresholds:

1. A score of 0 to 6 would immediately REJECT a candidate (the candidate would be automatically filtered to the "Reject" folder in the employer dashboard if their application scored below 6).
2. A score of 6 to 9 would send the candidate to the "INTERVIEW" pile.
3. A score of 9 to 10 would send the candidate to the "OFFER" pile.

It was unclear whether the values of 6 and 9 were inclusive or exclusive.

For "Custom Questions", we wrote three generic short-answer questions related to each role. We checked off the "Knockout" box for these questions, which meant that the responses to those questions could immediately disqualify (reject) a candidate if answered "unsatisfactorily".

Note that in their response to our research, Talenteria clarified that knockout questions "are generally intended for strict, objective eligibility requirements rather than nuanced qualitative assessment", such as "authorization to work in a specific country, required licenses or certifications, willingness to work in a required location, or other clear must-have requirements" (see Annex A). This guidance wasn't explicitly provided in the actual setup page for these questions in our free trial environment at our time of testing. Talenteria said "customers who use these features typically receive onboarding, guidance, and support from Talenteria, and they may also refer to our product documentation and manuals" (Annex A).

For the Legal Officer listings, we set the following questions and expected answers, followed by our test candidates' provided responses:

1. Do you have experience using office suite software? [Expected answer: "Yes"]

- "I do - Microsoft Office"

- "Yeah"

- "Not Microsoft Office, but yes other types"

2. Do you have a former non-compete agreement that would impact your work with us? [Expected answer: "No"]

- "None"

- "n/a"

3. Are you willing to travel for work if required? [Expected answer: "Yes"]

- "Potentially"

For Technologist, we set the following questions and expected answers, followed by our test candidates' responses:

1. Do you have a former non-compete agreement that would impact your work with us? [Expected answer: "No"]

- "None"

- "n/a"

2. What coding languages are you familiar with? [Expected answer: "Java, C++, C#, Python, Javascript"]

- "Rust, JS, C++, Python, SQL, Claude Code"

The purpose of the above mix of questions is to see if the algorithm is able to parse nuance in answers like a human reviewer could. At our time of testing, there were no instructions provided to candidates about how to answer the question (the questions were just text boxes on the job application). Naturally, some candidates might assume that providing more information (e.g., listing which office suite they use) would help their case.

AI Video Interviews

Talenteria also offered AI video interviews, where an AI interview bot conducted a video interview with candidates based on pre-established interview questions provided by the recruiter and embellished by follow-up questions generated live during the AI interview.

For the purposes of our test, we used the "generate interview questions with AI" feature to create the following interview questions for each job listing:

1. [TCLO] Can you describe your experience working with international human rights standards and how you have applied them in previous legal roles?
2. [TCLO] How would you approach mitigating legal risks faced by an organization, particularly in the context of technology and human rights?
3. [TCT] Can you describe your experience with automation and orchestration tools like Terraform, Puppet, Chef, or Ansible, and how have you applied them in a project?
4. [TCT] How would you approach analyzing a new device to determine how its data is generated, used, and secured?

Our test candidates scripted an answer to each question which they repeated in each interview. For the sake of consistency in our test, where the AI interviewer bot generated different follow-up questions, we responded with: "That's a great question, I'll have to get back to you on that".

Following the AI video interview, each candidate was assigned an "Average AI Interview Score" out of 10.

A transcript and visual/audio recording of the interview was also generated for each candidate, as well as an AI summary breakdown that scored each interview question individually and provided a summary of the candidate's fitness for the role based on their answer.

At the time of our testing, there was no ability for the recruiter to edit the candidate's AI-generated score (e.g., if the recruiter reviewed the transcript and video and wanted to rescore it themselves). However, Talenteria stated in their response to us that since then, the platform has been updated and the current version now has the ability for the recruiter to override the AI Match Score with their own scores (see Annex A).