Maki

MakiPeople - Response to additional questions

THE AUDITS

1. The audit report discloses the entire dataset was test data: As the AEDT, why was Mochi's historic data insufficient for this audit cycle? Will Mochi have enough production volume that the March 2027 audit will run on historic data, and will you commit to that publicly?

Test data was used in this audit cycle because we operate under strict data usage agreements with our customers. While we do run validation studies on production data for our customers' own purposes, we are not permitted to use that data for our own evaluation or audit purposes. This is a contractual constraint tied to how customer data may be used, not a reflection of production volume.

That said, the use of test data carries a genuine methodological advantage: it allows auditors to systematically vary transcript content, including inserting specific words or linguistic patterns, to stress test how the scoring prompt responds under controlled conditions. This level of systematic evaluation is not always possible with production data, where coverage of edge cases cannot be guaranteed. Importantly, the test data was independently created by the auditors, not by Maki. Our role was to run the transcripts through the system and return the results. Such an approach maintains the integrity and independence of the evaluation.

As Mochi continues to scale, we are taking a deliberate approach to data collection across the multiple dimensions it assesses. For the March 2027 audit cycle, our expectation is that we will explore mechanisms, such as updated customer agreements or de-identified/aggregated data arrangements, that would allow us to run the audit primarily on historic data, potentially augmented with specific linguistic or behavioural scenarios that require controlled evaluation. This is dependent on what arrangements can be reached with customers, and we will confirm the approach formally ahead of the next audit cycle.

2. Holistic AI authored this audit as your independent Local Law 144 auditor and, per your response, is also the external partner finalizing your EU AI Act conformity assessment. How do you and Holistic AI evaluate auditor independence given the concurrent commercial engagements?

When the same third party runs multiple engagements, questions of independence can legitimately arise. One key point here is that both engagements are structured as formal audits - not consulting or advisory work - which provides an important structural safeguard against the most common form of independence conflict, where a firm audits work it

helped design. Holistic AI runs both workstreams through separate teams, maintaining operational separation across the two audits.

We also remain aware that the external bias audit methodology itself could benefit from a fresh set of eyes - a second third party reviewing the bias audit approach is something we recognize as good practice and are open to as our AI governance matures.

DISABILITY AND VETERAN EXCLUSION

3. Your response says clients can collect voluntary self-identification data to support applicant-flow recordkeeping. Per MakiPeople's privacy policy and help section, demographic data including disability status is also collected by MakiPeople to improve fairness and unbiased in the handling of assessments.
Have any adverse impact analysis, internal or external, ever been run on Mochi outcomes by disability status or protected veteran status, using voluntary self-identification data or any other method? If yes, what did it show? If no, is one planned, and on what timeline? Why have these demographic groups, which are of significant size in the United States, been omitted?

As of this audit cycle, we do not yet have sufficient voluntary self-identification data on disability status or protected veteran status to support statistically defensible adverse impact analyses on Mochi outcomes for these groups. This is not an omission by design. It reflects the reality of where we are in our US market entry and the voluntary nature of demographic disclosure.

Protected veteran status represents approximately 6% of the US adult civilian population. Given that we have only recently entered the US market and are in the early stages of client partnerships, the proportion of candidates disclosing veteran status in our dataset remains too small to draw meaningful conclusions. Publishing results from an underpowered analysis would risk producing misleading findings, which serves neither candidates nor clients.

Disability status presents an additional layer of complexity. Disability is not a single, homogeneous category. The CDC report meaningfully different prevalence rates across cognitive (13.9%), mobility (12.2%), hearing (6.2%), and vision (5.5%) impairments, each of which may interact differently with a conversational AI assessment. Collapsing these into a single group for analysis would obscure more than it reveals. We are committed to doing this analysis properly, which requires both sufficient sample sizes within disability subgroups and a methodologically appropriate framework.

We intend to pursue both analyses once we have the evidence base to support them. We will not release findings until we are confident they are well-supported and defensible.

4. At what point is a candidate affirmatively told, before the conversation begins, that they may request an accommodation or an alternative format? Is the option presented to every candidate? Can you provide an example?

The accommodation request process is defined by each customer - they determine how and when candidates are informed of their right to request an accommodation, as this sits within their broader hiring process and candidate communication flows.

What Maki provides is the capability and features to support accommodation requests effectively. This includes the ability to extend assessment time, allowing candidates to complete assessments under conditions suited to their needs.

5. Your response states that disability information can be collected on a voluntary, consent-based basis and surfaced to recruiters. From my perspective, this may benefit candidates with disabilities when governed by concrete protections. How does the platform distinguish accommodation data, which process administrators need, from self-identification data, which must be firewalled from decision-makers?

We have identified that our response on this specific point was inaccurate. To clarify our earlier response, it is important to distinguish between two separate processes:

- Accommodation request data is operational; it exists solely to allow process administrators to action adjustments such as extended assessment time. We do not collect sensitive disability data in this process.
- Where disability data is collected on a voluntary, consent-based basis, this is strictly for fairness and adverse impact study purposes. This data is never surfaced to recruiters or hiring decision-makers - it is a completely distinct process from the accommodation request process. It exists to analyze whether assessment outcomes are equitable across candidate groups, not to inform any individual hiring decision.

SCORING AND MEASUREMENT

6. Your response states scoring is content-only and does not evaluate delivery, pace, or speech patterns. Your product page described real-time evaluation of "linguistic performance," your FAQ says the system evaluates "communication," and the quote in the Recruitics release references "skills, communication, and fit." Please define what "linguistic performance" and "communication" measure in the scoring rubric, and reconcile those terms with content-only scoring.

Linguistic performance refers specifically to language proficiency as defined by the Common European Framework of Reference (CEFR). Scoring is conducted exclusively on transcript content. The system assesses the richness, accuracy, and complexity of the language a candidate produces in response to the questions during the assessment. To support this, candidates are actively encouraged to share as much information as possible, sufficient time is built into the experience to enable a meaningful response, and guardrails are applied where transcript content is insufficient for a reliable evaluation. Scoring outputs are continuously benchmarked against human expert ratings, and where discrepancies arise, they are analysed and addressed through targeted guardrail refinements.

Communication, as referenced in our product and marketing materials, refers to a distinct behavioural competency. This means we evaluate a candidate's ability to recognise information, synthesise it, and convey it clearly to a target audience. This is assessed using Behaviourally Anchored Rating Scales (BARS), evaluating the content and structure of what a candidate communicates, not how they sound when communicating it.

These are two separate constructs, assessed through two separate methodologies.

7. The audit report's system description (page 17) states that Mochi captures candidate responses as audio and converts them to text using ASR before scoring. The audit methodology (page 18) states the synthetic transcripts, already text, were submitted directly to the scoring system. Which ASR provider or model does MakiPeople/Mochi use? Will you provide documentation on testing and results of the model's accuracy?

This is a relevant remark that we have also identified internally. The ASR bias risk is managed at multiple levels:
- First, through ASR technology selection - benchmarks and model cards for ASR systems include bias information, and this is a factor in our evaluation and selection process.
- Second, through internal testing before deployment in production - we do not rely solely on independent bias audits to assess ASR performance.
- Third, through regular requests for and reviews of ASR bias documentation from our providers. As AI technology providers, ASR providers are themselves subject to bias mitigation obligations under AI laws, and we actively monitor their compliance as part of our vendor governance.
- Fourth, through continuous monitoring of bias on an ongoing basis - not only through periodic independent audits - which allows us to identify and mitigate risks as they arise rather than at fixed audit intervals.
We generally do not disclose which ASR provider we use for intellectual property reasons.

THRESHOLDS AND HUMAN REVIEW

8. Your product page workflow states that candidates who pass the threshold flow to the next stage, and your EU AI Act Instructions for Use require human review of every adverse decision. Across live deployments, what percentage of candidates scored below threshold are subsequently advanced by human reviewers? What is the median time a reviewer spends on a below-threshold candidate's record?

Yes, human review of every adverse decision is clearly mandated in our EU AI Act Instructions for Use - this is a requirement, not a recommendation.

Regarding below-threshold pass rates: we do not currently generate platform-wide insights on this metric, as doing so cross-client proves complex. Customers own their review processes, and many action them at the ATS level or through other platforms outside of Maki, which limits our centralised visibility. In practice, we regularly have customers who contact us to discuss and adjust their thresholds, as they observe strong candidates falling below their initial threshold settings - which is itself an indication that human review is actively taking place.

Regarding time spent per record: similarly, tracking reviewer time is not easily feasible given that review processes are frequently conducted at least partially outside of our platform - for instance at the ATS level.

We recognise that providing customers with statistics and insights on how they conduct reviews would be valuable, particularly in the context of post-market monitoring, and it is something we consider as part of our EU AI Act program. We are however inherently limited by the fact that our customers are the ones with full visibility and control over their review process, typically at their ATS level.

9. Your response says you are preparing customer guidance on the applicant-definition implications of pre-apply screening. Will that guidance be published publicly, and when? Would you share it with Job Board Doctor on release?

As with our NYC Local Law 144 guidance, we do not publish this type of guidance publicly: it is an internal information resource prepared for our customers, not a public-facing document.

We see this kind of guidance as a facilitator: it reflects our understanding of the relevant legal landscape and helps customers think through the question. However, we are deliberate about not positioning it as a substitute for customers' own legal counsel, nor do we want to give customers a false impression of compliance. Customers remain responsible for their own legal and regulatory obligations, and our guidance is intended to inform and prompt their own analysis rather than replace it.

Maki

10. Confirming this is your final version (attached) and that you intend it for publication in full (with the exception of your contact information) on JobBoardDoctor.com. If so, I'll run it as your initial response, labeled as such.

Cf: Updated text in second PDF attached