



ANONYMISED REFERENCE ENGAGEMENT

# Annex III Classification & Annex IV Gap Memo

---

SUBJECT	TalentFit AI · Series B · EU mid-market HR-tech provider
ENGAGEMENT	Diagnostic (extended) · 5 working days
PREPARED BY	Chiekh Alloul, Partner · Tenth Partner

---

*This is an anonymised sample. Names, model architectures, scale figures, and dates have been altered to remove identifying information from a real engagement. Structure, methodology, and the analysis itself are representative of the work product delivered to clients.*

## SECTION 1 **Executive summary**

---

### **BOTTOM LINE**

- 01 All three of TalentFit AI's production systems — candidate matching, screening summarisation, structured interview scoring — are classified as high-risk AI under Annex III §4(a) of the EU AI Act.
- 02 The Article 6(3) low-risk exemption is not available for any of the three. The screening LLM has a superficial case for §6(3)(a) (narrow procedural task) but is defeated by the downstream coupling to the matcher.
- 03 Of the eighteen Annex IV requirements analysed, eleven are satisfied or near-satisfied by artefacts that already live in the engineering stack. Seven require new documentation work.
- 04 Three material gaps drive the remediation effort: no Article 72 post-market monitoring plan; no documented Article 9 risk-management system; subgroup metrics computed in the training pipeline but not exported per shipped system.
- 05 Procurement exposure is real today. Two of TalentFit AI's named target enterprise accounts have already added an AI Act due-diligence section to their vendor security review.

Recommendation: a two-week Readiness Sprint produces version one of the technical file from existing artefacts, plus the Article 9 RMS template and a 26-week roadmap to operational alignment before 2 August 2026. Estimated engagement value: €9,500.

## SECTION 2 **Annex III classification per system**

---

Each system in scope is assessed against Annex III §4 (employment, workers management, and access to self-employment) and tested against the four exemptions in Article 6(3).

---

### **Candidate Matching (System 1)**

**HIGH-RISK**

<b>FUNCTION</b>	Ranks candidates against open requisitions using a two-tower retrieval stage followed by a gradient-boosted reranker.
<b>I/O</b>	Input: candidate features and job-spec embeddings. Output: ordered list of candidates with relevance score in [0, 1] per (candidate, requisition) pair.
<b>SCALE</b>	Deployed across approximately 150 enterprise customers; in the order of 600,000 ranking calls per month.
<b>ANNEX III PARAGRAPH</b>	Annex III §4(a) — recruitment and selection of natural persons.
<b>ARTICLE 6(3) ANALYSIS</b>	Not eligible. The ranking materially shapes recruiter attention; relevance scores are conditioned on candidate profile features (profiling within the meaning of GDPR Article 4(4)); ranking precedes any human selection. None of §6(3)(a)–(d) apply.

---

### **Screening Summarisation (System 2)**

**HIGH-RISK**

<b>FUNCTION</b>	Extracts structured signals from CV and résumé text using a fine-tuned mid-size LLM.
<b>I/O</b>	Input: free-text CV. Output: structured profile — skills inventory, years of experience by domain, predicted seniority band, and a one-paragraph role-fit summary.
<b>SCALE</b>	Approximately 4 million CVs processed in the trailing 12 months across the same enterprise base.
<b>ANNEX III PARAGRAPH</b>	Annex III §4(a) — recruitment and selection.
<b>ARTICLE 6(3) ANALYSIS</b>	There is a superficial case for §6(3)(a) — narrow procedural task — because the system performs structured extraction rather than substantive scoring. The case is defeated because the structured output feeds directly into System 1 (the matcher), so the cumulative effect is ranking influence. Article 6(3)(d) — deviation-from-prior-decision-making patterns — also bites because the LLM-produced seniority prediction can override prior human triage. Not eligible.

---

## Structured Interview Scoring (System 3)

**HIGH-RISK**

<b>FUNCTION</b>	Scores recorded video-interview answers against a defined competency rubric using a fine-tuned multi-label classifier built on a speech-to-text frontend.
<b>I/O</b>	Input: video plus transcript. Output: dimension scores (communication, role-fit, structured thinking) and an overall job-fit score in [0, 100].
<b>SCALE</b>	Currently in pilot with twelve enterprise customers; full production launch planned for Q3 2026.
<b>ANNEX III PARAGRAPH</b>	Annex III §4(a) — recruitment and selection.
<b>ARTICLE 6(3) ANALYSIS</b>	Not eligible. Scoring is substantive (the model directly assesses individual candidates); no procedural-task exemption available. The fact that a human reviews the score before action is taken does not move the analysis — the scoring itself is the regulated act.

## SECTION 3 **Annex IV gap analysis**

---

Eight representative requirements from Annex IV §1, §2 and the operating Articles, mapped against TalentFit AI's existing engineering artefacts. The full mapping covers thirty rows and is delivered as part of the Readiness Sprint.

REFERENCE	REQUIREMENT	STATUS	SEVERITY	EVIDENCE / GAP NOTE
Annex IV §1(a)	System identifier and version	✓ <b>Satisfied</b>	–	MLflow experiment_id plus git SHA per shipped model. No work required.
Annex IV §1(c)	Description of how the system interacts with hardware and software	◐ <b>Partial</b>	<b>MEDIUM</b>	Service architecture diagrams exist for each system in the engineering wiki; no Annex-IV-shaped interaction document. One day of writing closes the gap.
Annex IV §1(g)	Validation and testing logs, signed and dated	◐ <b>Partial</b>	<b>MEDIUM</b>	MLflow run history and W&B reports are comprehensive; signing-and-dating discipline is missing. Resolved by a release-gate template that signs and dates the run on tag.

---

REFERENCE	REQUIREMENT	STATUS	SEVERITY	EVIDENCE / GAP NOTE
Annex IV §1(h)	Validation procedures, metrics by demographic subgroup	o <b>Partial</b>	<b>HIGH</b>	Subgroup metrics are computed in the training pipeline (fairlearn) but are not exported per shipped system. Material remediation: build an export job that writes per-system subgroup tables on every release.
Annex IV §2(b)	Design choices, assumptions, rationale	x <b>Missing</b>	<b>HIGH</b>	No model card or design-decision log exists for any of the three systems. New-build work: one model card per system, ~1 day of writing each from existing engineering memory.
Annex IV §2(g)	Test logs validating performance on representative inputs	o <b>Partial</b>	<b>MEDIUM</b>	Test logs exist in W&B; not signed-and-dated as Annex IV requires. Same release-gate template fix as §1(g).

REFERENCE	REQUIREMENT	STATUS	SEVERITY	EVIDENCE / GAP NOTE
Annex IV §9 / Art. 72	Post-market monitoring plan	× <b>Missing</b>	<b>CRITICAL</b>	No PMM plan exists. Drift monitoring is ad-hoc and not connected to subgroup metrics. Largest single piece of new work in the engagement.
Article 73	Serious incident reporting workflow (15 / 10 / 2-day clocks)	◦ <b>Partial</b>	<b>HIGH</b>	On-call rota and incident process exist; no detection-to-report SLA defined for AI-Act-classified incidents. Eight hours of process design closes the gap.

#### STATUS

- ✓ Satisfied — existing artefact meets the requirement.
- Partial — artefact exists; shape, signing, or scope misses the requirement.
- × Missing — no artefact today; new build required.

#### SEVERITY

- CRITICAL** Critical — blocks the technical file; remediate first.
- HIGH** High — material gap; remediate within the Sprint window.
- MEDIUM** Medium — closeable within four hours of work.
- LOW** Low — cosmetic; close at the next release.

## SECTION 4 Article 9 risk register – initial seed

---

The Article 9 risk-management system is a documented, lifecycle-long process. The five rows below are the seed risks identified during the Diagnostic; the full register is built out during the Sprint and maintained as part of the engagement.

RISK	LIKELIHOOD	SEVERITY	MITIGATION
Subgroup performance drift in the matcher post-deployment	MEDIUM	HIGH	Quarterly subgroup metric review; alerting thresholds wired to the on-call rota.
Hallucinated CV signals from the screening LLM (fabricated employers, mis-extracted dates)	LOW	MEDIUM	Output schema validation; LLM-judge consistency check on a 1% rolling sample; recruiter spot-check UI affordance.
Bias amplification across the matcher → screener → interview-scoring pipeline	MEDIUM	HIGH	Per-stage slice tests; coupling-aware monitoring; quarterly end-to-end fairness evaluation on a held-out cohort.
Re-purposing of system outputs for hiring decisions outside the documented scope	LOW	HIGH	Contractual scope clause in customer terms; UI affordance constraints; quarterly customer-side compliance attestation.
Data drift after a customer integrates a new ATS	MEDIUM	MEDIUM	Rolling evaluation set per customer; retraining cadence defined; per-customer onboarding gate.

---

## SECTION 5 **Article 72 post-market monitoring plan – outline**

---

The PMM plan is the largest single piece of new documentation work. Outline below; full plan delivered as a Sprint artefact.

### **TELEMETRY COLLECTED PER INFERENCE**

Input distribution snapshot (categorical and numerical features); prediction confidence; downstream action observed (clicked, contacted, interview-requested, hired); customer identifier and system version.

---

### **SUBGROUP TRACKING**

Per-system, per-30-days, per-protected-attribute (where data is available and lawful to track). Results written to a versioned subgroup-metrics dashboard and exported to the technical file appendix on the quarterly review.

---

### **DRIFT THRESHOLDS**

Population shift threshold: ten per cent KL-divergence on the input feature distribution triggers a review. Subgroup performance gap threshold: five per cent absolute gap on any tracked subgroup triggers escalation.

---

### **REPORTING CADENCE**

Monthly internal review by the engineering lead. Quarterly written report co-signed by the engineering lead and the Partner. Annual external review (optional, recommended at Programme tier).

---

### **LINKAGE TO ARTICLE 73**

Any PMM-detected event meeting the Article 3(49) serious-incident definition triggers the incident playbook (Section 6) and starts the 15 / 10 / 2-day clocks.

## SECTION 6 **Article 73 incident reporting playbook**

---

Article 73 imposes the 15 / 10 / 2-day reporting clocks. The playbook converts on-call detection into a regulator-ready report.

### **TRIGGER CONDITIONS**

Article 3(49) serious-incident definition operationalised against the three systems. Examples: a hiring outcome materially altered by a model failure; a subgroup performance breach beyond mitigation; a data-leak incident exposing CV content.

---

### **TIME-TO-CLASSIFY SLA**

T+24 hours after on-call detection: an engineering-and-Partner classification call decides whether the event is an Article 73 serious incident.

---

### **REPORTING CLOCKS**

15 days for general serious incidents; 10 days where serious harm occurred or is likely; 2 days for widespread infringement of Union law. Templates for each clock are part of the deliverable.

---

### **ESCALATION CHAIN**

On-call engineer → engineering lead → Partner → external counsel (if required) → competent national authority report.

## SECTION 7 26-week remediation roadmap

---

Calendar to 2 August 2026 enforcement of Annex III high-risk obligations. Owners shown are illustrative role labels; named owners are agreed at Sprint kickoff.

WEEKS	WORKSTREAM	OWNER	DELIVERABLE
<b>1-4</b>	Documentation gap closure	Engineering Lead + Partner	Three model cards · signed validation logs (release-gate template) · design-rationale log · §1(c) interaction document.
<b>5-10</b>	Post-market monitoring plan operational	Engineering Lead	PMM plan v1.0 · subgroup metric exports per system · drift alarms wired to on-call · monthly review cadence in production.
<b>11-16</b>	Article 9 risk-management system documented	Partner + Compliance	Documented RMS · risk register synchronised with PMM alarms · quarterly review process.
<b>17-22</b>	Article 73 incident reporting workflow	On-call Lead + Partner	Trigger detection rules · classification SLA wired into on-call · reporting templates for each of the 15 / 10 / 2-day clocks.
<b>23-26</b>	Annex IV technical file v1.0	Partner	Unified Annex IV technical file with signature pages · hand-off pack · readiness assessment letter for procurement use.

---

## SECTION 8 **Engagement next step**

---

TalentFit AI is in scope for high-risk obligations on all three production systems. The artefacts already in the engineering stack cover roughly two-thirds of Annex IV; the remaining third is concentrated in three workstreams (post-market monitoring, model cards, subgroup-metric exports) that a two-week Sprint addresses end-to-end. The recommended path is a Readiness Sprint kicking off within ten working days of acceptance, followed by the 26-week remediation roadmap above.

*Chiekh Alloul, Partner · Tenth Partner · 6 May 2026*



Tenth Partner · A specialist practice for EU AI Act readiness ·  
hello@tenthpartner.com · tenthpartner.com