EU AI Act Annex IV — AcmeScreen v2.4.0 (PowerQuant sample)

EU AI Act Annex IV — Technical Documentation

PowerQuant ApS — Confidential

Field	Value
Customer	Acme HR-tech ApS (CVR 99999999)
Product / system	AcmeScreen — Automated CV-screening module
Version	v2.4.0
Document date	2026-05-07
Classification	High-risk AI system per Annex III §4(a) (employment, workers management, access to self-employment)
Classification reference	Reg. (EU) 2024/1689, Annex III point 4(a)
CE-marking status	Pending — internal-control conformity assessment (Annex VI, Art. 43(2)); CE self-declared per Art. 48 - no notified body required for an Annex III point 4 system
EU declaration ID	EU-DOC-AcmeScreen-v2.4.0-2026-05-07
Review cycle	Quarterly (next: 2026-08-07); full re-issue on minor version bump per Art. 11(2)

> **FICTIONAL EXAMPLE — Acme HR-tech ApS er en konstrueret kunde til demonstration; alle data, bias-statistikker, og bias-metrics er illustrative.** Skabelon udleveret til virkelige Modul 3-kunder; tal, navne og certifikater er fiktive.

Annex IV Technical Documentation — AcmeScreen v2.4.0
§1 — General description of the AI system
1.1 Intended purpose
AcmeScreen is a SaaS module embedded in the Acme HR-tech recruitment platform. Its **intended purpose** is to perform an *automated initial CV screening* — i.e. ranking inbound CVs against a job description supplied by the employer-customer and surfacing the top N candidates (default N=20) for human recruiter review. The system is **not** intended to make final hiring decisions, reject candidates without human review, or assess personality traits. Rejection of a candidate always requires an explicit human action by a trained recruiter (see §6).
1.2 Markets and end-users
- **Geographic markets:** Denmark (primary) and Sweden (secondary, since 2025-Q4). Roll-out to Norway is on the 2027 roadmap. - **Sectors:** medium-sized employers (50–500 FTE) in IT-services, professional services, retail-HQ, and light manufacturing. Sold B2B per recruiter-seat; employers are customers, applicants are data subjects. - **Categories of natural persons affected:** job applicants (data subjects), recruiters

Annex IV Technical Documentation — AcmeScreen v2.4.0
(operators), and HR managers (oversight).
1.3 Version and release identification
- Product version: **v2.4.0** (released 2026-04-22, semantic versioning) - Model artifact: `acmescreen-bert-da-sv-2.4.0` (SHA-256: `b7a1...c4e2`, full digest in `/canonical/legal/MODEL_DIGEST_2026-05-07.txt`) - Previous version superseded: v2.3.4 (decommissioned 2026-04-29, retained for re-training audit per Art. 12(1))
1.4 Hardware and software architecture
- **Inference runtime:** Python 3.12, PyTorch 2.4.1, HuggingFace `transformers` 4.46.2, `sentence-transformers` 3.1.1, served via TorchServe 0.12 behind FastAPI 0.115. - **Container orchestration:** Kubernetes 1.30 on AWS EKS (eu-central-1, Frankfurt), node pool m6i.2xlarge for inference, g5.xlarge (NVIDIA A10G) for the embedding stage. - **Storage:** Postgres 16 (RDS, encrypted at rest with KMS CMK), S3 for model artifacts (versioned, MFA-delete), CloudWatch + Grafana Cloud for telemetry. - **Network boundary:** VPC-private, accessed only through the Acme application API gateway. No direct public ingress to inference pods. - **Third-country data transfer:** none — all processing in EU (Frankfurt). Model training was performed in eu-central-1 only; no transfer outside the EEA.

§2 — Detailed description of elements and development process

2.1 Training data

- **Source corpus:** anonymised CVs collected through Acme's own platform 2020-01-01 → 2024-12-31, n = 420 000 unique CVs across 11 800 job postings.
- **Anonymisation pipeline:** strips name, address, email, phone, photo, DOB, CPR/personnummer *before* the training warehouse. Original PII stays in the operational DB only. DPO-reviewed 2024-11; minute at `legal/DPIA_2026-05-07.md §A.3`.
- **Languages:** Danish (~63%), Swedish (~24%), English (~13%).
- **Splits:** 80/10/10 train/val/test; test set frozen 2024-12-31, never used during hyperparameter sweep.
- **Labels:** binary `passed_screen` derived from anonymised recruiter actions (advanced-to-interview vs. archived). De-confounded for recruiter-id and calendar week to reduce reviewer-specific bias.

2.2 Feature engineering

CV text is parsed by a deterministic Danish/Swedish NLP pipeline (spaCy 3.7 with `da_core_news_lg` and `sv_core_news_lg`) into structured fields: years of experience, education level (ISCED-mapped), declared skills, language proficiencies, and free-text narrative. The transformer model consumes the narrative + a structured-features prompt; **no inference** is made on prohibited attributes (gender, ethnicity, age beyond legal minimum) — these fields are dropped before tokenisation.

2.3 Model architecture and training

- **Architecture:** BERT-base, 12 layers, 768 hidden, multilingual checkpoint `xlm-roberta-base` fine-tuned on the Acme training set with a binary classification head.
- **Hyperparameters:** lr=2e-5, batch=32, 3 epochs, warmup 10%, weight decay 0.01, AdamW.
- **Compute:** 4×A10G for ~14 hours per training run; full reproducibility via locked `requirements.lock` and pinned dataset hash.

- **Tooling:** Weights & Biases for experiment tracking (project `acmescreen-prod`, all runs retained ≥10 years per Art. 12), MLflow 2.16 for model registry.

2.4 Third-party libraries (key dependencies, frozen versions)

Library	Version	Licence	Purpose
`torch`	2.4.1	BSD-3	Tensor / autograd
`transformers`	4.46.2	Apache-2.0	Model loading
`sentence-transformers`	3.1.1	Apache-2.0	Embeddings
`spacy`	3.7.6	MIT	NLP parsing
`fairlearn`	0.11.0	MIT	Bias diagnostics
`evidently`	0.4.33	Apache-2.0	Drift detection
`fastapi`	0.115.0	MIT	Serving API

Full SBOM in CycloneDX format at `legal/SBOM_acmescreen_2.4.0.cdx.json`; renewed on every release per ISO/IEC 5230 SCA practice.

2.5 Data labelling protocol

Recruiter “advance-to-interview” events are the ground-truth proxy. Acme’s labelling protocol — documented in `internal/labelling_protocol_v2.pdf` — requires recruiters to log a structured reason code when archiving a candidate; reason codes that map to non-merit grounds (e.g. “position withdrawn”) are excluded from training. Inter-rater reliability sampled monthly (κ = 0.71 on a 200-CV double-labelled subset, target ≥ 0.65).

2.6 Validation methodology and performance metrics

Held-out test set (n = 42 000):

- Precision@20 = **0.84**
- Recall@20 = **0.79**
- F1 = **0.81**
- AUC-ROC = 0.88
- Calibration ECE = 0.04 (Platt scaling applied post-hoc)

Subgroup analysis required by Art. 10(2)(g) and Art. 15 reported in §5 and §7.

§3 — Monitoring, functioning and control
§4 — Risk management system (Article 9)
A continuous, documented risk-management process per Art. 9(2) is operated as part of the Acme ISMS. The system is reviewed **quarterly** and on every minor version bump. Below is the live risk register (extract; full register in `legal/RISK_REGISTER_2026-05-07.xlsx`).
4.1 Identified risks
1. **R-01 Bias on protected characteristics** — disparate true-positive rates across gender, age band, or ethnicity proxy. 2. **R-02 Hallucinated qualifications** — model attributing skills/qualifications not present in CV text. 3. **R-03 GDPR Art. 22 over-reach** — the screening becoming, in practice, a “decision based solely on automated processing” without meaningful human review. 4. **R-04 Security breach** — exfiltration of CV data, model weights, or inference logs. 5. **R-05 Data poisoning / training-set tampering** — adversarial CV submissions designed to degrade or bias the model in retraining.
4.2 Risk-mitigation table
| ID | Risk | Inherent | Mitigation | Residual | Owner | Re-eval | |—|—|—|—|—|—|—| | R-01 | Bias on protected attributes | High | Pre-processing strips protected attrs; quarterly Fairlearn audit (demographic parity, equal opportunity); subgroup TPR delta gate at ≤ 0.05 (block release if breached) | Medium | Head of ML | Quarterly | | R-02 | Hallucinated qualifications | Medium | Output is a *ranking over the input CV*, not generative text; SHAP traces verify all surfaced skills exist in CV; nightly grounding test on 500-CV regression set | Low | ML Lead | Monthly | | R-03 | GDPR Art. 22 over-reach | High

§3 — Monitoring, functioning and control
| Mandatory human action before rejection (UI + API enforced); contractual flow-down to employer-customers in DPA §7.3; quarterly UI usage audit verifying ≥ 99% rejections show ≥ 5 s recruiter dwell time | Medium | DPO | Quarterly | | R-04 | Security breach | High | SOC 2 Type 1 in place; pen-test annually + on major version; KMS-encrypted at rest; least-privilege IAM; vulnerability disclosure programme `security@acmehr.dk` | Low | CISO | Continuous | | R-05 | Data poisoning | Medium | Training data drawn only from authenticated employer accounts (no scraping); anomaly detection on CV submission velocity per IP; manual approval gate for retraining datasets | Low | Head of ML | Per retrain |
4.3 Continuous review schedule
Quarterly Risk Council (CTO, DPO, Head of ML, CISO, external counsel) reviews the register. Material changes feed §11 Post-Market Monitoring. Incident response is governed by `legal/INCIDENT_RESPONSE_2026-05-07.md` and connects to Art. 73 reporting (see §11).

§5 — Data and data governance (Article 10)

5.1 Training-set composition

- 420 000 CVs, deduplicated by content hash.
- Job categories (top 5): IT/software (31%), sales (14%), administration (12%), customer service (9%), engineering (8%).
- Time coverage: 2020–2024 (5 years); class balance at training time 41% positive / 59% negative after stratified sampling.

5.2 Bias-relevant attribute statistics (illustrative)

Where attributes were inferable from anonymised proxies (forename pattern → gender proxy; education-graduation-year → age band; first-language metadata → broad ethnicity proxy), distributions on the training data were:

- **Gender proxy:** 48% female / 51% male / 1% non-binary / unclear. Demographic parity difference (Fairlearn) on test set: 0.03 (target ≤ 0.05).
- **Age band:** 18–24: 19%, 25–34: 41%, 35–44: 24%, 45–54: 11%, 55+: 5%. TPR delta for 45+ band vs. 25–34: 0.04 (within tolerance).
- **Language-of-CV:** Danish-first 63%, Swedish-first 24%, English-first 13%. Recall delta English-first vs. Danish-first: 0.06 (above tolerance — flagged in §11 monthly bias re-check; mitigation: targeted upsampling in next retrain).

These are diagnostic proxies only; the live system never receives the protected attribute as a feature per Art. 10(5).

5.3 Data-quality controls

Pre-ingestion checks: schema validation, encoding normalisation (UTF-8 NFC), language detection, duplicate detection, malformed-PDF rejection. Post-ingestion: spot-checks of 100 random CVs per week by a data-quality analyst; results logged in `dataops/qc_log.md`.

5.4 Data-cleaning protocol

PII-stripping is *additive-deny*: a regex/NER union strips name, email, phone, address, photo metadata, CPR, personnummer, IBAN. Any CV whose stripping flags >5 PII tokens after the first pass is quarantined and reviewed manually before either deletion or inclusion in the training corpus.

5.5 DPIA reference

A Data Protection Impact Assessment per GDPR Art. 35 was completed 2024-11-18, refreshed 2026-04-30. Stored at `legal/DPIA_2026-05-07.md`. The DPIA covers: necessity & proportionality, risks to data-subject rights, safeguards (incl. those listed here), and

the consultation outcome (Datatilsynet was not formally consulted — risk assessed as not “high” residual risk per Art. 36 GDPR after mitigations).

5.6 Lawful basis (GDPR)

For Acme as **processor** to its employer-customers (controllers):

- **Art. 6(1)(b) GDPR** — performance of the contract between the data subject (applicant) and the prospective employer (job-application context), as far as the customer relies on it.
- **Art. 6(1)(f) GDPR** — legitimate interest of the employer in efficient candidate triage; balancing test documented in `legal/LIA_2026-05-07.md`. The balancing test concludes that the interest is legitimate, processing is necessary, and the impact is proportionate given the human-review safeguard (§6) and the right to object.

For Acme as **controller** of the training dataset (separate processing activity):

- **Art. 6(1)(f) GDPR** — legitimate interest in product improvement, with anonymisation as the principal safeguard. Data-subject opt-out honoured at submission time (UI checkbox; default *opt-in to anonymised improvement*; opt-out propagated to training warehouse within 30 days).

5.7 Special-category data (GDPR Art. 9)

Special categories (health, religion, trade-union membership, ethnic origin, sexual orientation, biometric, genetic) are **blocked at preprocessing** by an NER + keyword denylist. Any CV in which special-category content cannot be redacted with high confidence is excluded from training. The live inference path also filters these tokens before they reach the model. No Art. 9(2) condition is relied upon — the design intent is to *not process* special-category data.

§6 — Human oversight (Article 14)
§7 — Accuracy, robustness and cybersecurity (Article 15)
7.1 Accuracy metrics
Reported on the frozen test set (see §2.6) and re-

§6 — Human oversight (Article 14)
evaluated quarterly on a rolling holdout drawn from production traffic with explicit data-subject consent for evaluation use:
- Precision@20 = **0.84** (95% CI 0.83–0.85) - Recall@20 = **0.79** (95% CI 0.78–0.80) - F1 = **0.81** - Subgroup: gender-proxy parity 0.03, language-proxy gap 0.06 (flagged).
7.2 Robustness — adversarial testing
Three attack vectors tested 2026-Q1:
1. **Keyword-stuffing attack** — CVs padded with high-frequency positive tokens. Detection rate 96% via length-normalised TF-IDF anomaly. 2. **Paraphrase attack** — same content rewritten to game tokenisation. Ranking variance < 5 percentile-points (acceptable). 3. **White-text injection** (PDF zero-opacity tokens) — 100% caught by the PDF text-layer hash check at ingest.
Results in `security/RED_TEAM_REPORT_2026-Q1.pdf`. Next red-team window: 2026-Q3.
7.3 Cybersecurity
- **Threat model:** STRIDE, last refresh 2026-03; abuse cases include credential stuffing, model extraction, training-data exfil, prompt injection via CV. - **Pen-test:** annual external (next: 2026-09 by independent firm); plus internal quarterly. - **SOC 2:** Type 1 issued 2026-02; Type 2 expected 2026-Q4. - **Vulnerability disclosure:** `security@acmehr.dk`, 90-day policy, public on `acmehr.dk/.well-known/security.txt`. - **Patching SLA:** critical CVE within 7

§6 — Human oversight (Article 14)
days, high within 30, medium within 90.
7.4 Failure-modes log
Maintained at `legal/FAILURE_MODES_2026-05-07.md`. Each entry: failure mode, detection mechanism, recovery procedure, last observed date. Top three currently:
- F-01: tokenizer OOM on extreme-length CVs (>50 pages) — mitigated by 32 k-token cap at ingest. - F-02: language detector mis-classifies bilingual CVs as English — mitigated by per-paragraph detection. - F-03: confidence calibration drift after major retrain — mitigated by post-hoc Platt re-fit gated in CI.

§8 — Risk management system summary table (executive view)

This re-states §4 in compact form for board / customer-CTO review.

ID	Risk	Severity (inherent)	Likelihood	Mitigation summary	Residual	Re-eval cadence
R-01	Disparate-impact bias	High	Medium	Pre-processing + quarterly Fairlearn gate	Medium	Quarterly
R-02	Hallucinated qualifications	Medium	Low	Ranking-only design + SHAP grounding test	Low	Monthly
R-03	GDPR Art. 22 over-reach	High	Medium	Hard human-action gate + UI dwell audit	Medium	Quarterly
R-04	Security breach	High	Low	SOC 2 + pen-test + KMS + IAM	Low	Continuous
R-05	Data poisoning	Medium	Low	Authenticated source + retrain gate	Low	Per retrain

Residual-risk acceptance signed by CTO + DPO at each quarterly Risk Council. Last sign-off: 2026-04-29.

§9 — Harmonised standards applied
§10 — EU declaration of conformity (Article 47)
Issued in accordance with Article 47 of Regulation (EU) 2024/1689. To be signed on completion of the internal-control conformity assessment (Annex VI, Art. 43(2)). No notified body is required for this Annex III point 4 (employment) system.
`` EU DECLARATION OF CONFORMITY
1. AI system (product name + type + version): AcmeScreen — automated CV-screening high-risk AI system, v2.4.0
2. Name and address of provider: Acme HR-tech ApS, [street], [city], Denmark, CVR 99999999
3. This declaration is issued under the sole responsibility of the provider.
4. Object of the declaration: AcmeScreen v2.4.0, model artifact acmescreen-bert-da-sv-2.4.0 (SHA-256 b7a1…c4e2)
5. The object described above is in conformity with: - Regulation (EU) 2024/1689 (AI Act) - Regulation (EU) 2016/679 (GDPR) - Applicable harmonised standards listed in §9 above
6. References to relevant harmonised standards used: ISO/

§9 — Harmonised standards applied
IEC 42001:2023, ISO/IEC 23894:2023, ISO/IEC 27001:2022.
7. Notified body: Not applicable. Conformity assessment for this Annex III point 4 (employment) system is based on internal control (Annex VI) per Art. 43(2); no notified body is involved.
8. Where applicable, additional information: Annex IV technical documentation reference: EU-DOC-AcmeScreen-v2.4.0-2026-05-07.
9. Signed for and on behalf of: Acme HR-tech ApS
Place and date of issue: Copenhagen, [date] Name, function: [CEO name], Chief Executive Officer Signature: _____ ```
The signed declaration shall be kept at the provider’s disposal for 10 years after the AI system is placed on the market or put into service per Art. 47(2), and a copy shall be submitted to the relevant national competent authority on request.

§11 — Post-market monitoring (Article 72)

A documented post-market monitoring plan is maintained per Art. 72(1) and the implementing-act template (when published). Until the implementing act is published, the plan follows the structure below.

11.1 Monitoring plan

Cadence	Activity	Trigger thresholds	Owner
Continuous			SRE

Cadence	Activity	Trigger thresholds	Owner
	Latency, error rate, confidence histogram	5xx > 1% / 10 min, mean latency > 500 ms / 5 min	
Weekly	KS / PSI feature-drift batch	KS p<0.01 ∧ effect>0.10	ML Lead
Monthly	Bias re-check (Fairlearn) on rolling 30-day inference	Any subgroup TPR delta > 0.05	Head of ML
Quarterly	Accuracy re-evaluation on consented holdout	F1 drop > 0.03 vs. baseline	Head of ML
Annually	External audit (AIMS + technical)	Any non-conformity	DPO + CTO

Findings are aggregated into a quarterly Post-Market Monitoring Report, archived alongside this Annex IV file.

11.2 Incident reporting (Article 73)

Serious incidents per Art. 3(49) and Art. 73(1) — including any incident leading to harm to a person's health, fundamental rights, property or environment, and any malfunction breaching Union law protecting fundamental rights — are reported by the provider to the market-surveillance authority of the Member State concerned.

For Denmark, the relevant authority is **Datatilsynet** (data-protection-related serious incidents) and the future designated AI market-surveillance authority once formally appointed under Art. 70. Reporting timeline: **immediately upon establishing a causal link, and not later than 15 days** after the provider becomes aware of the serious incident, in line with Art. 73(2). Where the incident results in death, this is reduced to **10 days**; widespread infringements or critical-infrastructure incidents follow the **2-day** path under Art. 73(3).

For GDPR-side personal-data breaches, the existing Art. 33 GDPR **72-hour** notification path to Datatilsynet remains the applicable rule — the AI Act incident-report regime is *additive*, not a substitute.

11.3 Escalation thresholds

- **L1 — internal investigation:** any single bias gate breach, any P2 alert.

- **L2 — DPO + CTO notified within 4 h:** repeated L1 in same week, any P1 alert, any data-subject complaint citing automated decision-making.
- **L3 — authority notification path opens:** confirmed disparate-impact harm, confirmed personal-data breach, confirmed security incident with applicant data exposure, or any event matching Art. 73(1).

Each escalation has a named on-call owner; rota in `ops/ONCALL_2026-05.md`.

11.4 Sanctions awareness

Non-compliance with the obligations on providers of high-risk AI systems is subject to administrative fines up to **EUR 15 million or 3% of total worldwide annual turnover** under Art. 99(4). Acme management is briefed on this exposure annually.

How PowerQuant generated this report
This document is the output deliverable of two PowerQuant modules. It was not authored by a generic legal template; it is grounded in artefacts produced during the engagement.
- **Module 1 — AI Inventory & Art. 4 Register & Gap-Analysis (5 working days, fixed-fee).** Produced the AI-system inventory, the Art. 4 AI-literacy register, and a gap-analysis against AI Act Arts. 9–15, 26, 50 and GDPR Arts. 22, 35. - **Module 3 — Annex IV Technical File & RMS & PMM (~6 weeks).** Produced (i) this Annex IV file, (ii) the Art. 9 RMS in §4 and the live risk register, (iii) the §11 PMM plan, (iv) the §10 EU declaration template. - **Cross-references:** companion artefacts at `canonical/legal/DPIA_2026-05-07.md`, `legal/LIA_2026-05-07.md`, `legal/RISK_REGISTER_2026-05-07.xlsx`, `legal/INCIDENT_RESPONSE_2026-05-07.md`, `security/RED_TEAM_REPORT_2026-Q1.pdf`. Module 1 and Module 3 are listed on `powerquant.dk/priser` with current scope and

pricing. - **What this is not:** a legal opinion. PowerQuant is not a law firm. The output is reviewed by the customer’s own external counsel. Conformity assessment for this system is internal control under Annex VI (Art. 43(2)); no notified body is involved. - **Maintenance:** this document is versioned alongside the AcmeScreen software release. A minor model release triggers a §2 + §7 refresh. A material change triggers a full re-issue and a fresh EU declaration per Art. 47.

End of Annex IV technical documentation — AcmeScreen v2.4.0 — 2026-05-07.