iMocha Skills Intelligence AI Inference Engine

Technical Overview for Data Privacy & AI Governance Teams

Prepared By:
Nilesh Pethani
AI Architect, iMocha
nilesh@imocha.io

Table of Content

Executive Summary

iMocha Skills Intelligence uses an AI inference engine to automatically detect, validate, and score workforce skills from multiple data sources. This document provides a technical overview of how the inference engine works, what data it processes, and the controls in place to ensure responsible, auditable, and privacy-compliant operation.

This document is intended for data privacy officers, AI governance teams, legal counsel, and security reviewers evaluating iMocha for enterprise deployment.

Purpose: Automatically infer verified skill profiles from employee data — covering resumes, certifications, learning activity, project history, and AI-assisted interviews — without requiring manual self-assessment.

Outcome: A continuously updated, multi-source validated skills graph for each employee, used to power workforce planning, skill gap analysis, and career development.

Governance: All inferences are explainable, auditable, and configurable. No black-box scoring. Employees can view and contest inferred skills.

Sincerely,

Nilesh Pethani
AI Architect, iMocha
nilesh@imocha.io

1. Data Sources Used for AI Inference

The inference engine processes five primary data source types. Each source contributes independently to a skill's inferred confidence score. iMocha never requires all sources to be active — organizations configure which sources are enabled.

1.1 Resume and Profile Data

What is processed: Structured and unstructured text from employee resumes, LinkedIn profiles, or internal HR profile systems (via integration).
How inference works: Natural language processing (NLP) extracts job titles, responsibilities, technologies, tools, and tenure. Skills are mapped to iMocha's Skills Taxonomy using semantic similarity models. Years of experience per skill are estimated from role duration and context.

Data Field	Inference Output	Confidence Weight
Job title and role history	Skill category and proficiency range	Medium
Technologies and tools listed	Specific skills with version context	High
Responsibilities described	Applied skills and domain expertise	Medium
Education and degree fields	Foundation skills and domain knowledge	Low–Medium

1.2 Certifications & Credentials

What is processed: Completed certifications from integrated platforms (Udemy, Coursera, Pluralsight, LinkedIn Learning, internal LMS, vendor programs) or manually submitted certificates.
How inference works: Each certification is mapped to iMocha's Skills Taxonomy via a pre-built certification-to-skill dictionary maintained by iMocha's team. Certifications carry a higher confidence weight than self-assessed data because they represent third-party validation.

Certification Type	Handling
Industry-standard certs (AWS, Azure, PMP, CPA)	Direct high-confidence skill mapping from curated dictionary
University or academic credentials	Mapped to foundation and domain skills; proficiency set to intermediate
Vendor/platform-specific certs	Mapped via platform partnership dictionary or NLP fallback
Expired certifications	Inferred skill retained but confidence score is time-decayed
Unrecognized certificates	Flagged for manual review; not auto-inferred

1.3 Learning & Course Activity

What is processed: Completion records, progress data, quiz/assessment scores, and course metadata from integrated LMS or learning platforms.

How inference works: Completed courses are mapped to skills using course metadata and content tags. Partial completions are scored proportionally. Assessment scores within courses feed into the proficiency estimate for the associated skill.

Key inference signals from learning data:

- Course completion (after course completions contribute fractional weight)
- Assessment or quiz scores within courses (adjusted for difficulty calibration)
- Course recency (more recent activity scores higher in confidence)
- Repeat engagement with the same skill area (increases confidence over time)

1.4 Projects & Work History

What is processed: Project records from integrated tools (Jira, GitHub, internal project systems) or manually entered project summaries.

How inference works: Project titles, descriptions, and metadata are parsed to extract applied skills. Skills inferred from project data are weighted more heavily than declarative sources (resumes, self-assessments) because they represent demonstrated application.

Key Principle: Skills inferred from project activity are treated as applied evidence and receive a higher confidence multiplier than skills inferred from profile text alone.

Project Signal	Inference Value
Repo languages and frameworks (GitHub)	Technical skills with high confidence
Ticket categories and resolution patterns (Jira)	Domain skills and methodology indicators
Project role and contribution scope	Seniority signals and leadership skills
Cross-functional project involvement	Collaboration and soft skill indicators

1.5 AI Interview

What is processed: Responses from structured AI-conducted interviews, including spoken or written answers, problem-solving exercises, and scenario-based questions.

How inference works: AI Interview is iMocha's highest-signal inference source. Interview sessions are structured around specific skill domains. Natural language understanding models evaluate response quality against validated rubrics for each skill area. Scores are generated per skill.

AI Interview inference components:

- Speech-to-text transcription (for voice interviews) using industry-standard ASR models
- NLU evaluation of answer quality, depth, and accuracy per question
- Competency rubric scoring — each question maps to a validated competency framework
- Confidence calibration — responses are scored relative to a validated population distribution
- Human review option — all AI Interview scores can be flagged for recruiter or manager review

Bias Mitigation: AI Interview models are regularly evaluated for demographic bias across gender, language background, and accent. Bias audit reports are available upon request for enterprise customers.

2. How the Inference Engine Works

2.1 Skills Taxonomy Foundation

All inferences are anchored to iMocha's Skills Taxonomy — a structured, hierarchical library of over 25,000+ skills organized by domain, sub-domain, and proficiency level. The taxonomy is the controlled vocabulary against which all inferred skills are normalized.

Note: For an organization, all inferences are anchored to their skills taxonomy.

Taxonomy Component	Description
Skill nodes	Individual skills with canonical names, aliases, and related terms
Domain hierarchy	Groupings by function (e.g., Data Engineering > Cloud > AWS)
Proficiency levels	Standardized levels: Beginner, Intermediate, Experienced , Proficient
Skill adjacency mappings	Related skills to enable gap analysis and career path modelling
Version and deprecation tracking	Skills are versioned to handle technology evolution

2.2 Multi-Source Confidence Scoring

Each inferred skill is assigned a confidence score (0–100) computed from the weighted combination of all available source signals. The scoring model is additive — more confirming sources increase confidence; conflicting signals reduce it.

Default source confidence weights (Weights are configurable. Customers can increase weight for internally-proctored assessments or reduce weight for self-reported resume data to match their governance requirements)

Data Source	Default Weight	Rationale
Certifications	25%	Third-party validated; objective evidence
Learning & Course Completion	10%	Demonstrated learning intent with partial performance signal
Projects / Work Activity	25%	Applied evidence of skill use
AI Interview/ Assessments	20%	Direct, structured assessment — highest signal quality
Managers Rating	10%	Provides observed, contextual evidence of skill application
Resume / Profile/ Self-Rating	10%	Declarative; useful for breadth, lower precision

2.3 Proficiency Level Assignment

Once a confidence score is established, the engine maps it to a proficiency level using a calibrated scale. The Proficiency Level/ Confidence Range below are defaults and can be adjusted per organization.

Proficiency Level	Confidence Range	Description
Beginner	20–39	Has encountered the skill; limited practical application
Intermediate	40–59	Can apply the skill with guidance; developing independence
Experienced	60–79	Consistently applies the skill independently; solid depth
Proficient	80–100	Advanced mastery; can lead, teach, and advance the skill area

2.4 Confidence Decay Over Time

Skill inferences are time-sensitive. The engine applies a decay function to reduce confidence for skills that have not been recently validated. Decay rates vary by skill volatility category:

Skill Category	Decay Rate	Example
Rapidly evolving (e.g., AI/ML frameworks)	High — 6-month half-life	PyTorch, Kubernetes
Moderately evolving (e.g., cloud platforms)	Medium — 12-month half-life	AWS Solutions Architect
Stable technical (e.g., SQL, Java core)	Low — 24-month half-life	SQL querying, OOP principles
Domain knowledge (e.g., finance, HR)	Very low — 36-month half-life	IFRS accounting, labor law

3. Data Privacy & Security Controls

3.1 Data Minimization

The inference engine is designed on a data minimization principle. It extracts only the signals needed to produce a skill inference and does not retain raw unstructured text (e.g., full resume content or interview transcripts) beyond the inference processing window, unless the customer explicitly enables archival storage with appropriate consent controls.

Data Element	Retention Policy
Raw resume text	Processed in-session; not stored unless archival enabled
AI Interview transcripts	Configurable: 30/60/90-day retention or immediate deletion post-scoring
Course completion records	Retained as structured metadata (course ID, score, date); not full content
Project activity data	Structured signals only (skill tags, dates); no raw ticket content retained
Inferred skill scores	Retained in skills profile with full audit trail of source contributions

3.2 AI Model Data Usage

iMocha's inference models are trained on aggregate, anonymized skill-signal datasets (PII removed prior to any training process). Individual employee data is never used to train or fine-tune shared inference models. The models are static during inference — they do not learn or update from individual employee interactions.

3.3 Employee Transparency & Rights

iMocha is designed to support employee data rights consistent with GDPR, CCPA, and equivalent frameworks:

Right	How It Is Supported
Right to access	Employees can view their complete skills profile and the source of each inferred skill
Right to correction	Employees can dispute inferred skills; disputes trigger human review workflow
Right to erasure	Employee skill profiles can be fully deleted on request
Right to explanation	Each inferred skill shows contributing sources and confidence breakdown
Right to opt-out	Organizations can configure which inference sources are active

4. AI Governance & Responsible AI Principles

4.1 Explainability

Every skill inference produced by the engine is fully explainable. The system provides a breakdown of which sources contributed to the inferred skill, how much each contributed, and what the raw score from each source was. No skill is inferred from a black-box process.

Explainability artifacts available per inferred skill:

- Source contribution breakdown (% weight per data source)
- Raw confidence score per source
- Taxonomy path (how the inferred skill maps to the skills taxonomy)
- Recency-adjusted score (showing impact of time decay)
- Validation status (whether any source was third-party verified)

4.2 Human Oversight

The inference engine is positioned as a decision-support tool, not a decision-making system. Inferred skill data is used to inform — not automate — talent decisions. The following human oversight mechanisms are in place:

- Manager review: Managers can review, endorse, or flag any inferred skill for their direct reports
- HR override: HR administrators can manually adjust inferred proficiency levels with a documented rationale
- Dispute resolution: Employee-initiated disputes route to a defined human review process
- Talent decision disclaimer: iMocha outputs are advisory; customers are contractually responsible for any employment decisions

4.3 Bias Monitoring

iMocha conducts regular bias audits across all inference models, with particular focus on:

Bias Risk Area	Monitoring Approach	Frequency
Gender bias in skill inference	Differential analysis across gender groups per skill domain	Quarterly
Language and accent bias (AI Interview)	Accuracy parity testing across language backgrounds	Quarterly
Recency bias favoring younger employees	Tenure-adjusted scoring validation	Semi-annual
Credential bias (favoring formal education)	Source weight calibration review	Annual

Enterprise customers can request bias audit reports for their organization's inference outputs upon written request.

4.4 Model Governance

Governance Dimension	iMocha Commitment
Model versioning	All inference model versions are tracked; customers are notified before any model update
Rollback capability	Previous model versions retained for 12 months; rollback available on request
Change management	Significant model changes require 30-day advance notice to enterprise customers
Third-party audit	Annual third-party AI model audits conducted; summaries available on request
Incident response	AI inference errors or anomalies trigger defined incident response with customer notification SLA

5. Integration Architecture & Data Flow

5.1 Integration Methods

iMocha Skills Intelligence connects to enterprise data sources through the following integration patterns:

Integration Type	Details
HRIS / HR System (Workday, SAP SuccessFactors, Oracle HCM)	Bidirectional via REST API or SFTP batch; employee profiles and org structure synced
LMS / Learning Platforms (Cornerstone, Degreed, LinkedIn Learning, Coursera)	OAuth-based API integration; completion events and scores pulled in near real-time
Project Tools (GitHub, Jira, Azure DevOps)	Read-only API access; activity signals extracted without storing full content
Assessment Platforms (iMocha Assessments, third-party)	Native integration; assessment scores fed directly into inference pipeline
Certification Platforms (Credly, Acclaim, vendor portals)	API or badge import; certifications verified against issuer records

5.2 Data Flow Overview

The inference pipeline processes data in the following sequence:

1. Ingestion: Data is pulled from connected sources via secure API calls or batch imports. Raw data is processed in an isolated inference environment.
2. Normalization: Unstructured text is parsed and normalized. Skills terminology is mapped to the Skills Taxonomy using NLP and semantic matching.
3. Signal extraction: Relevant signals (skills mentioned, scores, completion status, dates) are extracted and structured.
4. Confidence scoring: Each signal is scored and weighted by source type and recency. Composite confidence scores are computed per skill.
5. Conflict detection: Cross-source conflicts are identified and flagged or resolved per the resolution logic.
6. Profile update: Inferred skill scores are written to the employee's skills profile with a full audit record.
7. Decay scheduling: A scheduled process re-evaluates confidence scores over time and applies decay adjustments.

6. Summary: Key Assurances for Governance Review

The table below summarizes the key questions typically raised by data privacy and AI governance teams, and iMocha's response to each.

Governance Question	iMocha Position
Is employee data used to train AI models?	No. Employee data is never used to train or fine-tune shared inference models without explicit contractual consent.
Are inferences explainable?	Yes. Every inferred skill shows a full breakdown of contributing sources and confidence scores.
Can employees see and contest their skill inferences?	Yes. Employees have full visibility into their inferred skills profile and can initiate a dispute for human review.
What happens if an inference is wrong?	HR administrators can override inferences. Disputes trigger a documented human review process.
Is AI used to make employment decisions?	No. iMocha outputs are advisory. Employment decisions remain with human managers and HR teams.
Is facial recognition or biometric data used?	No. AI Interview uses NLP on spoken/written responses only — no biometric or facial analysis.
What data is retained after inference?	Configurable. Raw text is not retained by default. Structured skill scores and audit logs are retained per customer configuration.
Is the system compliant with GDPR/CCPA?	Yes. Data subject rights (access, correction, erasure, explanation) are supported. DPA available on request.
Are bias audits conducted?	Yes. Quarterly bias audits across gender, language, and demographic factors. Audit summaries available on request.
What certifications does iMocha hold?	SOC 2 Type II certified. ISO 27001 in progress. Full compliance documentation available on request.

Brewing A *Skills-First* Planet

Call us **+1-408-915-5185** or write to us at **support@imocha.io**

About iMocha

iMocha is a global AI-powered Skills Intelligence platform that enables organizations to adopt and scale a skills-first strategy across hiring, workforce readiness, internal mobility, and continuous learning.

At its core, iMocha leverages AI-driven skills agents to create a closed-loop system - continuously inferring skills from resumes, certifications, and work data; validating them through assessments and conversational AI interviews; and translating them into structured, actionable insights for better talent decisions.

Built on a robust skills taxonomy and ontology, iMocha provides organizations with a dynamic and continuously evolving view of their workforce capabilities and talent landscape.

Designed as an invisible intelligence layer, iMocha seamlessly integrates with existing enterprise systems - HCM platforms such as SAP SuccessFactors and Workday; learning platforms like Degreed and Cornerstone OnDemand; e-learning providers including Udemy, Coursera, LinkedIn Learning, and Pluralsight; as well as ATS solutions like ICIMS and SmartRecruiters - enabling rapid adoption without disrupting existing workflows.

To know more, visit us at www.imocha.io.