



AI-SKILL Gap Analysis Tool

Indracare Lab 2.0

Prepared for

myOnsite Healthcare, LLC.

AUG 2025

Document Control

Rev. No.	Description of Change	Effective Date
1.0		06/18/2025

Authored By

Name	Role	Signature	Date
myOnsite Healthcare			22 nd Aug 2025

Reviewed and approved By

Name	Role	Signature	Date

Document Control	2
1. Project Overview Section — SKILL Gap Analysis Tool	4
Project Brief	4
System Overview	4
2. Data & Requirements Section	5
Document Corpus / Data Set	5
Evaluation Dataset	5
3. Technical Requirements Section	6
Primary Technical Domain — Knowledge Mapping & Diagnosis	6
Secondary Technical Domain — Curriculum Generation & Sequencing	6
Integration/Architecture Domain — Progress Tracking & Prediction	7
4. Advanced/Challenging Requirements	7
Complex Feature 1 — Cross-Graph Skill Alignment	7
Complex Feature 2 — Explainable Predictions & Interventions	8
Production Considerations	8
Security/Compliance	8
Containerization & Dockerization (MANDATORY)	8
5. Output Schema / API Specification	9
Endpoints / Interfaces	9
Validation Rules	10
JSON Schema (representative)	10
6. Evaluation Framework	11
Multi-Dimensional Scoring	11
Advanced Methods	11
Benchmarking Targets	12
7. System Adaptation / Flexibility Requirements	12
8. Evaluation Complexity	13
9. Advanced Challenge Specifications	13
Hot-swap / Migration Requirements	13
Zero-Downtime / High-Availability	13
Multi-Tier Evaluation Suite	13
Challenge-Type Testing	14
Performance Benchmarking	14
10. Deliverables Section	14
Primary Deliverables (System)	14
Secondary Deliverables (Evaluation)	14
Documentation Deliverables	15
Tools & DevEx Artifacts	15
11. Success Criteria	16
Technical Excellence	16
System Performance	16

Production Readiness	16
Innovation & Depth	16
APPENDIX A — Content Guidelines for AI Generation	17
APPENDIX B — Usage Instructions	17

1. Project Overview Section — SKILL Gap Analysis Tool

Project Brief

The SKILL Gap Analysis Tool is an enterprise-grade AI platform that ingests multi-subject student data, evaluates mastery, maps knowledge to standardized skill graphs, and prescribes individualized learning paths. It continuously measures progress and predicts academic performance to help educators intervene earlier and at scale.

The system targets K–12 and higher-ed programs, test-prep providers, and workforce reskilling initiatives operating under data privacy and audit requirements. It must integrate with existing student information systems and learning platforms, operate reliably across peak testing windows, and provide transparent, explainable outputs suitable for academic review boards.

Assumptions:

- Stakeholders include program directors, educators, data/assessment teams, and compliance officers.
- Supported subjects: mathematics, language arts, sciences, and elective domains; skill graphs follow widely used standards.
- Must operate in privacy-regulated environments and support internationalization.

System Overview

You're building a platform that:

- Ingests heterogeneous student/assessment/behavioral data at scale
- Diagnoses per-skill mastery and gaps with uncertainty quantification
- Generates personalized curricula and micro-objectives with pacing
- Tracks longitudinal progress and mastery across subjects

- Predicts performance (e.g., pass likelihood) with fair, explainable models
 - Exposes APIs/UI for educators and interoperates with existing systems
-

2. Data & Requirements Section

Document Corpus / Data Set

- **Data types:** assessment submissions, item-level responses, clickstream/engagement events, attendance, assignment metadata, educator annotations, demographic & program attributes (minimized, privacy-respecting).
- **Volume:** initial pilot 50–200K students; 10–30M item responses; 0.5–1.5B event records/year.
- **Complexity indicators:** multi-lingual content, adaptive tests, rubric-based grading, partial credit, multi-select, constructed responses, open-ended text, and time-on-task.
- **Special characteristics:** skill graph alignment, cross-subject pre/co-requisites, curriculum variants by jurisdiction, accommodations (IEP/504), and test security constraints.

Evaluation Dataset

- **Test cases:** 3,000–5,000 student profiles across 4 subjects with stratified demographics and prior-achievement levels.
- **Difficulty tiers:**
 - Tier 1 (30%): clean, complete records; standard MCQ items.
 - Tier 2 (40%): partial/missing data; mixed item types; multilingual prompts.
 - Tier 3 (20%): adversarial/noisy inputs; atypical pacing; off-graph items.

- Tier 4 (10%): edge accommodations, outlier behaviors, and curriculum shifts mid-term.
 - **Special requirements:** include gold-standard educator judgments for gap identification and curriculum relevance scoring; include fairness slices (e.g., by program, language).
-

3. Technical Requirements Section

Primary Technical Domain — Knowledge Mapping & Diagnosis

- **Content & edge cases:** MCQ, MSQ, numeric, short/long text, code snippets, diagrams; partial credit and multi-skill alignment; embedded hints; adaptive routing artifacts.
- **Interfaces/Contracts:** batch/stream ingestion; item-response schema with scoring metadata; skill graph alignment API; uncertainty-calibrated mastery scores with versioned taxonomies.
- **Performance targets:**
 - Per-student diagnostic generation ≤ 2.0 s P95 (cold cache ≤ 4.0 s)
 - Mastery classification accuracy $\geq 92\%$ vs. educator gold labels
 - Skill alignment F1 ≥ 0.90 on held-out labeled items
- **Observability:** evented diagnostics with spans for parsing \rightarrow scoring \rightarrow mapping; logs include student_id (pseudonymous), assessment_id, item_types, alignment_version, latency_ms, confidence, and fallback path; dashboards for P50/P95 latency, error rate, and calibration plots.

Secondary Technical Domain — Curriculum Generation & Sequencing

- **Scope & edge cases:** prerequisite chaining, spiral review, mastery thresholds, remediation and enrichment branches, time constraints (calendar-aware), and resource

modality preferences.

- **Contracts:** curriculum_plan API (objectives, activities, estimated minutes, resources, skill_refs); sequencing policy registry; educator constraints overlay.
- **Performance & SLAs:**
 - Plan generation ≤ 1.5 s P95 for 10–20 objectives
 - Relevance score ≥ 0.88 vs. educator ratings
 - Constraint satisfaction rate $\geq 98\%$ for hard constraints
- **Observability:** structured logs for rule firing, constraint conflicts, backtracking count; tracing for node expansions; dashboard for coverage vs. gaps over time.

Integration/Architecture Domain — Progress Tracking & Prediction

- **Data/Items:** longitudinal snapshots, rolling mastery deltas, attendance streaks, engagement velocity, assignment completion; privacy-preserving joins.
 - **Interfaces:** metrics registry (mastery%, growth%, predicted score band); risk flags with reasons; webhook for intervention triggers; standards-based import/export.
 - **Performance targets:**
 - Prediction refresh ≤ 10 min for 100K students (batch); on-demand ≤ 800 ms P95
 - AUC ≥ 0.85 ; Brier score ≤ 0.18 ; calibration error ≤ 0.03
 - **Observability:** model lineage tags, feature drift monitors, fairness reports (Δ across slices), and cost/compute dashboards.
-

4. Advanced/Challenging Requirements

Complex Feature 1 — Cross-Graph Skill Alignment

- **What:** map items/activities to multiple jurisdictional skill graphs with versioning.
- **Why hard:** non-isomorphic graphs, ambiguous objectives, and temporal version drift.
- **Acceptance:** alignment $F1 \geq 0.90$ across graphs; drift detection within 7 days of upstream changes; human-in-the-loop review UI with agreement $\kappa \geq 0.75$.

Complex Feature 2 — Explainable Predictions & Interventions

- **What:** per-student risk predictions with natural-language rationales and actionable next steps.
- **Why hard:** balancing fidelity, privacy, and pedagogical usefulness.
- **Acceptance:** global + local explanations validated by educators $\geq 85\%$ “useful” rating; SHAP-like attributions sanitized; no sensitive attributes used directly; fairness gap ≤ 3 pts across key slices.

Production Considerations

- Horizontal scale to 1M+ students; HA $\geq 99.9\%$ monthly; RPO ≤ 1 hour, RTO ≤ 1 hour.
- IaC-managed environments; blue/green or canary rollouts; automated backups with quarterly restore drills.

Security/Compliance

- AuthN/AuthZ with least-privilege roles; encrypted data at rest and in transit; secrets isolated from code; audit logs retained ≥ 1 year.
- Data minimization, data subject rights workflow, and export controls.
- Regular privacy impact assessments; third-party risk reviews.

Containerization & Dockerization (MANDATORY)

- **Multi-stage image** with non-root user, deterministic build, minimal attack surface, healthcheck, and build cache layers.
 - **Runtime orchestration** via compose or equivalent: services (api, worker, ingestion, db, cache, message bus), ports, env vars, secrets, volumes, resource limits.
 - **Supply chain**: image scanning, SBOM generation, provenance attestation.
 - **Operations**: startup/ready probes, graceful shutdown, structured logs, metrics export.
 - **Make targets**: `make build`, `make run`, `make test`, `make lint`, `make sbom`.
-

5. Output Schema

Validation Rules

- All IDs are opaque, non-guessable strings; timestamps ISO 8601; numeric scores [0,1]; arrays length limits enforced; constraints must include at least one hard objective.
- Requests/Responses include `schema_version` for compatibility.

JSON Schema (representative)

```
{
  "primary_id": "string",
  "main_response": {
    "content": {
      "mastery_vector": [{"skill_id": "string", "score": 0.0}],
      "gaps": [{"skill_id": "string", "severity": "low|med|high"}],
      "curriculum_plan": {
        "objectives": [{"id": "string", "skill_id": "string", "target_score": 0.8}],
        "activities": [{"id": "string", "label": "string", "estimated_minutes": 30, "skill_refs": ["string"]}],
        "rationale": "string"
      },
    },
    "predictions": [{"subject": "string", "horizon_days": 30, "prob_pass": 0.0, "band": "A|B|C|D"}],
  },
  "metadata": [{"key": "string", "value": "string"}],
  "confidence_score": 0.0,
  "uncertainty_factors": ["data_missing", "domain_shift", "cold_start"]
}
```

```
},
"supporting_data": [{
  "reference": "string",
  "location": {"offset": 0, "span": 0},
  "id": "string",
  "excerpt": "string",
  "relevance": 0.0,
  "quality": 0.0,
  "method": "heuristic|ml|rule"
}],
"performance_metrics": {
  "latency_ms": 0,
  "processing_ms": 0,
  "cost_units": 0.0
},
"system_metadata": {
  "model_info": "string",
  "strategy_info": "string",
  "timestamp": "2025-08-23T00:00:00Z"
}
}
```

6. Evaluation Framework

Multi-Dimensional Scoring

- **Diagnostic Accuracy:** $\geq 92\%$ skill classification vs. educator labels (macro-avg).
- **Curriculum Relevance:** mean educator rating $\geq 4.4/5$; constraint satisfaction $\geq 98\%$.
- **Prediction Quality:** $AUC \geq 0.85$; calibration error ≤ 0.03 ; Brier ≤ 0.18 .
- **Latency/Throughput:** diagnostics ≤ 2.0 s P95; curricula ≤ 1.5 s P95; ≥ 200 rps sustained at peak.
- **Fairness:** disparity ≤ 3 percentage points across defined slices; significance tested.

Advanced Methods

- **Adversarial inputs:** noisy text, code-switched language, time-on-task anomalies.
- **Statistical tests:** calibration curves, KS tests across cohorts, drift tests.
- **Red-team prompts:** constraint-breaking curricula, out-of-domain skills, privacy probes.

Benchmarking Targets

- P95 latency thresholds as above; error budget $\leq 0.5\%$ monthly; availability $\geq 99.9\%$; on-call MTTR ≤ 30 min.

7. System Adaptation / Flexibility Requirements

- Accepts varied input formats (CSV, JSON events, form-data), multiple locales, and different skill-taxonomy versions.
- Handles missing/partial data via imputation or fallback heuristics with uncertainty propagation.
- Supports educator overrides and accommodations without invalidating analytics lineage.

Flexible payload example

```
{
  "student_profile": {"id":"opaque","locale":"en-US","accommodations":["extended_time"]},
  "context": {"subject":"math","taxonomy_version":"v3","window_days":28},
  "responses": [{"item_id":"i123","type":"msq","answer":["A","C"],"time_sec":45}],
  "constraints":
  {"pace_minutes_per_day":30,"modality":["video","practice"],"hard_objectives":["skill:fractions.add"]},
  "options": {"allow_spiral_review":true,"max_activities":20}
}
```

8. Evaluation Complexity

- **Tier 1 (Basic, 30%):** clean data, standard items, single taxonomy, English.
 - **Tier 2 (Intermediate, 40%):** partial data, mixed items, bilingual, accommodations.
 - **Tier 3 (Advanced, 20%):** adversarial/noisy inputs, off-graph items, abrupt pace changes.
 - **Tier 4 (Expert, 10%):** mid-course taxonomy update, transfer students with short history, overlapping programs.
-

9. Advanced Challenge Specifications

Hot-swap / Migration Requirements

- Swap knowledge-graph versions, diagnostic models, curriculum policies, or prediction components with versioned contracts and A/B orchestration.
- **Validation:** shadow traffic for 2 weeks; no regression on accuracy/relevance; latency within $\pm 10\%$.

Zero-Downtime / High-Availability

- Targets: $\geq 99.9\%$ monthly; single AZ/zone failure tolerance.
- Rollout: canary 5% \rightarrow 25% \rightarrow 100% with auto-rollback on error budget burn.
- Monitoring: SLOs on latency, error rate, and fairness drift with paging thresholds.

Multi-Tier Evaluation Suite

- **Tier 1:** unit/contract tests; schema validation; small gold sets.
- **Tier 2:** offline evaluation on stratified cohorts; calibration checks.
- **Tier 3:** shadow production with teacher-in-the-loop ratings.
- **Tier 4:** canary-in-prod with automated rollback criteria.

Challenge-Type Testing

- Cross-graph alignment under version drift.
- Hard-constrained curriculum when time is scarce (e.g., 2 weeks pre-exam).
- Prediction robustness under missing attendance/engagement.

Performance Benchmarking

- Diagnostics throughput ≥ 200 rps; curricula generation ≥ 150 rps.
 - Storage query P95 ≤ 50 ms for hot paths; background refresh ≤ 10 min per 100K students.
-

10. Deliverables Section

Primary Deliverables (System)

- Knowledge mapping service with uncertainty-aware mastery outputs.
- Curriculum generation engine with constraint solver and rationale.
- Progress tracking & prediction services with fairness reporting.
- Interoperability APIs and educator-facing console.

Secondary Deliverables (Evaluation)

- Labeled gold datasets and scoring harness.
- Load, stress, and chaos test suites with reproducible scenarios.
- Fairness and explainability reports with dashboards.

Documentation Deliverables

- Runbook (SLOs, playbooks, on-call).
- Architecture & dataflow diagrams (C4 Level 2/3 description below).
- API reference & schema registry docs.
- Security & compliance checklist and data protection impact assessment.
- Testing strategy and evaluation methodology.

Tools & DevEx Artifacts

- [README](#), [CONTRIBUTING](#), environment bootstrap script.
- Container artifacts: Dockerfile (multi-stage), compose/k8s manifests, SBOM and scan reports.
- Makefile targets: build/run/test/lint/sbom.

Diagram description (C4-style):

- **System Context (L1):** “Educator” and “Admin” actors interact with the SKILL Gap Analysis Tool; integrations include student information systems and learning platforms; data lake/warehouse provides historical inputs.
- **Container (L2):** API gateway; diagnostics service; curriculum service; prediction service; ingestion service; orchestration/queue; metadata store; analytics store; object storage; observability stack.

- **Component (L3):** within diagnostics, components for response parsing, scoring, alignment, uncertainty calibration; within curriculum, components for objective selection, constraint solver, rationale generator.
-

11. Success Criteria

Technical Excellence

- Alignment F1 ≥ 0.90 ; diagnostic accuracy $\geq 92\%$; curriculum relevance $\geq 4.4/5$; prediction AUC ≥ 0.85 .

System Performance

- P95 latencies: diagnostics ≤ 2.0 s; curricula ≤ 1.5 s; predictions ≤ 800 ms.
- Availability $\geq 99.9\%$; MTTR ≤ 30 min; error budget $\leq 0.5\%/mo$.

Production Readiness

- IaC-defined environments; blue/green or canary; backups verified quarterly; SBOMs and image scans with no high/critical vulns at release.

Innovation & Depth

- Explainable, fair predictions with $\leq 3pt$ disparity across slices; versioned multi-graph support; human-in-the-loop workflows.

During development, demonstrate:

1. Load test sustaining ≥ 200 rps diagnostics with ≤ 2.0 s P95.
2. Failure injection: node loss without breaching SLOs.

3. Hot-swap of a diagnostic or prediction component with shadow validation and zero downtime.
-

APPENDIX A — Content Guidelines for AI Generation

- Professional, senior-engineering tone; structured lists; quantified metrics.
 - Vendor-neutral; multiple deployment patterns allowed; avoid prescribing named products.
 - Security & compliance always in scope; monitoring and SLOs mandatory.
 - Complexity escalates Basic → Advanced → Production → Expert; include zero-downtime path.
 - Dockerization requirements are explicit and testable; deliverables categorized into System, Evaluation, Documentation, Tools.
-

APPENDIX B — Usage Instructions

1. No clarifying questions; assumptions labeled.
2. Vendor-neutral; no specific technology stacks named.
3. Every numeric claim has thresholds/targets.
4. Diagram description included (C4 model).
5. Dockerization artifacts specified.
6. Success criteria and evaluation tiers defined.