Darwin Education: Architecture-First Adaptive Learning With Psychometric and Safety Governance

Authors: Demetrios Chiuratto Agourakis (ORCID: https://orcid.org/0009-0001-8671-8878); Isadora Casagrande Amalcaburio (ORCID: https://orcid.org/0009-0005-6883-1809)

Affiliations: [Affiliations to be added]

Corresponding author: [Email to be added]

Abstract

Background. AI tutoring systems in medical education are often evaluated at the interface layer (generation quality) rather than at the control layer (measurement, governance, and safety).

Objective. To document a reproducible, architecture-first adaptive learning system aligned to ENAMED microdata provenance, with explicit psychometric, validation, and human-review gates.

Methods. We audited repository implementations and derived artifacts, requiring every numerical claim to be either repository-anchored (`path:line-range`), INEP-anchored (official URLs), or explicitly labeled `NOT YET COMPUTED`.

Results. Runtime enumeration of the Darwin-MFC exported indexes reports diseases `raw=252`, `unique=215`, and medications `raw=889`, `unique=602`, with duplicate fractions of `14.68%` and `32.28%`, respectively (evidence: `_paperpack/derived/darwin_mfc_runtime_counts.json:13-14`, `_paperpack/derived/darwin_mfc_runtim`
Historical targets `368/690` are retained only as legacy documentation targets, not current corpus truth (evidence: `_paperpack/derived/darwin_mfc_runtime_counts.json:3777-3781`).

Conclusions. This preprint is positioned as an implementation and governance report. It makes no educational efficacy claim and no clinical decision claim. Metrics that require runtime datasets are marked `NOT YET COMPUTED`.

Keywords: adaptive learning, psychometrics, ENAMED, governance, safety instrumentation, reproducibility

---

1. Introduction

This manuscript describes Darwin Education as a governed adaptive-learning architecture, not as an efficacy trial. The focus is implementation transparency: how psychometric inference, safety gating, and human review are wired into the software stack.

Two constraints guide this version (v0.3):

1. numerical claims must be evidence-grounded;
2. unresolved measurements must be explicitly labeled `NOT YET COMPUTED`.

---

2. System Overview (Architecture-First)

Darwin Education combines:

- psychometric inference components (`packages/shared/src/calculators/`),
- adaptive content generation services (`apps/web/lib/qgen/services/`),
- learning-gap routing (`apps/web/lib/ddl/services/`),
- persistence and audit infrastructure (`infrastructure/supabase/`).

The Darwin-MFC submodule provides disease/medication corpora consumed by runtime index exports.

No educational-outcome claim is made in this manuscript.

---

3. Data Provenance and Corpus Accounting

3.1 ENAMED Provenance (Official INEP)

This package references ENAMED microdata provenance from official INEP publication pages, including:

- microdata portal entry published on **19/01/2026**,
- INEP news publication on **20/01/2026**,
- Nota Técnica nº 19/2025 (Angoff + TRI),
- INEP microdata governance/LGPD framing page.

(INEP references listed in Section 8)

3.2 ETL-Documented Values (Repository-Observed)

The ETL README documents the following values:

- total exam items: `100` (evidence: `infrastructure/supabase/seed/enamed-2025-etl/README.md:17-18`
- valid IRT items: `90` (evidence: `infrastructure/supabase/seed/enamed-2025-etl/README.md:18`)
- valid participants: `~900K` (evidence: `infrastructure/supabase/seed/enamed-2025-etl/README.md:20`
- valid/excluded item split (`90` valid, `3` excluded) (evidence: `infrastructure/supabase/seed/enamed-202`

These are treated as ETL-documented repository values, not re-executed claims in this paper cycle.

3.3 Runtime Corpus Counts (Current Source of Truth)

v0.3 uses runtime enumeration from exported index arrays/objects (`todasDoencas`, `todosMedicamentos`) through `_paperpack/scripts/darwin_mfc_runtime_counts.ts` and artifacts in `_paperpack/derived/`.

Table 1. Runtime Governance Table (Darwin-MFC Corpus)

Kind	Raw count	Unique by normalized ID	Duplicate count	Duplicate fraction
Diseases	252	215	37	14.68%
Medications	889	602	287	32.28%

Evidence:

- diseases raw/unique: `_paperpack/derived/darwin_mfc_runtime_counts.json:13-14`
- medications raw/unique: `_paperpack/derived/darwin_mfc_runtime_counts.json:562-563`
- duplicate evidence materialization: `_paperpack/derived/darwin_mfc_duplicates.csv:1`

3.4 Static vs Runtime Counting (Method Distinction)

This repository now maintains two distinct counting paradigms:

1. **Static file-level ID extraction** (regex/source scan): `_paperpack/scripts/derive_darwin_mfc_count`
2. **Runtime exported index enumeration** (actual imported exports): `_paperpack/scripts/darwin_mfc_runtime_counts.ts:334-427`

v0.3 prioritizes runtime counts for manuscript claims. Historical `368/690` values are preserved only as legacy targets from prior documentation metadata (evidence: `_paperpack/derived/darwin_mfc_runtime_counts.json:3777-3781`).

---

4. Safety and Governance Controls

4.1 Validation Decision Gates

The QGen validation service defines weighted decision thresholds:

- auto-approve: `>= 0.85`
- pending review: `>= 0.70`
- needs revision: `>= 0.50`

(evidence: `apps/web/lib/qgen/services/qgen-validation-service.ts:35-37`)

4.2 Medical Pattern Checks and Human Review

The medical verification service implements regex-based dangerous-pattern checks and severity tagging (evidence: `apps/web/lib/qgen/services/medical-verification-service.ts:1`

The review API enforces explicit reviewer decisions (`APPROVE`, `REJECT`, `REVISE`) (evidence: `apps/web/app/api/qgen/review/route.ts:134-149`).

4.3 Citation Verification Controls

Citation verification uses explicit allowlist/blocklist logic and cache TTL (7 days) (evidence: `apps/web/lib/theory-gen/services/citation-verification-service.ts:50-80`).

4.4 Scope Boundary

This manuscript does **not** claim educational efficacy, pass-rate improvement, or clinical decision support performance.

---

5. Reproducibility Capsule

5.1 Commands

```
# Full evidence-pack pipeline
bash _paperpack/scripts/run_all.sh

# Runtime corpus counts
bash _paperpack/scripts/run_darwin_mfc_runtime_counts.sh
```

5.2 Generated Runtime Artifacts

- `_paperpack/derived/darwin_mfc_runtime_counts.json`
- `_paperpack/derived/darwin_mfc_duplicates.csv`
- `_paperpack/derived/darwin_mfc_runtime_log.txt`

5.3 Metrics Status

The following remain `NOT YET COMPUTED` in this manuscript cycle:

- prospective educational impact metrics,
- safety sensitivity/specificity against expert gold labels,
- production latency/throughput distributions.

The Darwin-MFC submodule integration test was previously non-green in this workspace due to missing local dependency installation (`fuse.js` resolution failure); after dependency installation it is green again, and both failure and resolution logs are preserved in `_paperpack/logs/v0.3_test_darwin_mfc.log` and `_paperpack/logs/v0.3.1_test_darwin_mfc.log`.

---

6. Limitations

1. **Runtime-vs-static divergence.** Static extraction and runtime enumeration produce different corpus numbers; only runtime values are treated as current truth in v0.3.
2. **ETL validation metrics are documented but not re-executed here.** Repository docs list target metrics, but this preprint does not claim fresh revalidation.
3. **Outcome studies pending.** No causal or comparative educational efficacy analysis is presented.

---

7. Data and Code Availability

- **Source code:** https://github.com/agourakis82/Darwin-education
- **Repository commit for this manuscript baseline:** `924fee69bac70fdec8baee32c057f0f0f704a79b`
- **Evidence pack:** `_paperpack/` (regenerate with `bash _paperpack/scripts/run_all.sh`)
- **Software DOI (Zenodo, versioned):** https://doi.org/10.5281/zenodo.18592149
- **Software DOI (Zenodo, concept):** https://doi.org/10.5281/zenodo.18487441
- **Release manifest id:** `v0.3.1_release_2026-02-10T09-16-37Z` (`_paperpack/derived/v0.3.1_release_2026-02-10T09-16-37Z_manifest.json`)
- **Darwin-MFC submodule commit (runtime artifact):** `34d5c94f0de814cac45d907352fb580babadd` (evidence: `_paperpack/derived/darwin_mfc_runtime_counts.json:4`)

ENAMED microdata are public INEP data. INEP's microdata governance framing references LGPD-aligned publication practices and suppression/anonymization safeguards for identification risk mitigation (official source cited in Section 8). The archived evidence pack includes a hash-stamped manifest linking manuscript, benchmarks, and runtime corpus enumeration outputs.

---

8. References

1. INEP. Microdados ENAMED (portal, published 19/01/2026). https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enamed
2. INEP. Notícia de publicação dos microdados ENAMED (20/01/2026). https://www.gov.br/inep/pt-br/centrais-de-conteudo/noticias/enamed/inep-publica-microdados-do-enamed-referente-a-primeira-edicao-do-exame
3. INEP. Nota Técnica nº 19/2025 (Angoff + TRI). https://download.inep.gov.br/enamed/nota_tecnica_n_
4. INEP. Microdados: governança e dados abertos (LGPD framing). https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados
5. Reckase, M. D. *Multidimensional Item Response Theory*. Springer, 2009.
6. Holland, P. W., & Thayer, D. T. Differential item performance and the Mantel-Haenszel procedure. In *Test Validity*, 1988.

7. Wozniak, P. A. *Optimization of Learning: SuperMemo Algorithm SM-2*, 1990.

Appendix A. Claim → Evidence Table (Compact)

Claim	Classification	Evidence anchor
Runtime diseases count is `raw=252`, `unique=215`	`repo-anchored`	`_paperpack/derived/darwin_mfc_runtime_counts.j`
Runtime medications count is `raw=889`, `unique=602`	`repo-anchored`	`_paperpack/derived/darwin_mfc_runtime_counts.j`
Historical targets `368/690` are legacy metadata only	`repo-anchored`	`_paperpack/derived/darwin_mfc_runtime_counts.j`
Runtime method imports exported indexes at execution time	`repo-anchored`	`_paperpack/scripts/darwin_mfc_runtime_counts.t`
Static method is source-level derivation	`repo-anchored`	`_paperpack/scripts/derive_darwin_mfc_counts.py`
QGen thresholds `0.85/0.70/0.50`	`repo-anchored`	`apps/web/lib/qgen/services/qgen-validation-ser`
Medical dangerous-pattern checks exist as explicit regex rules	`repo-anchored`	`apps/web/lib/qgen/services/medical-verificatio`
Citation verification cache TTL is `7` days	`repo-anchored`	`apps/web/lib/theory-gen/services/citation-veri`
ENAMED microdata portal publication date `19/01/2026`	`INEP-anchored`	INEP URL #1 in Section 8
ENAMED microdata publication news date `20/01/2026`	`INEP-anchored`	INEP URL #2 in Section 8
ETL documented values (`100` total, `90` valid, `~900K` participants)	`repo-anchored`	`infrastructure/supabase/seed/enamed-2025-etl/R`
Prospective efficacy metrics	`NOT YET COMPUTED`	Not executed in this manuscript cycle

---

Supplementary Materials

- architecture_map.md
- data_audit.md
- safety_pipeline.md
- repro_capsule.md
- benchmarks.md
- derived/darwin_mfc_runtime_counts.json
- derived/darwin_mfc_duplicates.csv
- derived/darwin_mfc_runtime_log.txt

Figure A1. Architecture Diagram Source (Mermaid)

```
flowchart LR
  A[ENAMED Microdata] --> B[Calibration ETL]
  B --> C[(Supabase)]
  C --> D[MIRT/DIF/Unified Model]
  C --> E[DDL Analysis]
  D --> F[Adaptive Question Mapping]
  E --> F
  F --> G[QGen LLM]
  G --> H[Validation + Safety Filters]
  H --> I{Auto-approve?}
  I -->|Yes| J[Serve Content]
  I -->|No| K[Expert Review]
  K --> J
  G --> L[Theory Generation]
  L --> M[Citation/Hallucination Audit]
```