
DGIN 5401

Operationalized Machine Learning in Healthcare

Course Project Handbook

AIRO-Sim Research Studio

Fall 2026 | Student handbook

Project in one sentence

The class will build and study a shared, transparent cardiac-surgery risk system using fully synthetic AIRO-Sim data. Each student will lead a distinct research question, lead a shared artifact, and submit an individual research paper; the class will also produce a short shared system dossier and an integrated demonstration.

Individual work	A claimable research question, decisive evidence, an independently written paper, reproducible code, and an oral defence.
Shared work	A versioned T0 pipeline, case viewer, containerized runnable system, monitoring demonstration, and concise system dossier.
Warning	AIRO-Sim is fully synthetic and separate from AIRO’s real-data research. Course results are not clinical evidence and are not for patient care.

This handbook states the course requirements. Release-specific schemas and file details appear in the companion *AIRO-Sim Technical Brief*. All deadlines are in Atlantic Time unless stated otherwise.

Contents

1 Purpose, setting, and evidence boundary	1
2 Shared platform, individual studies	1
3 Choosing the anchor outcome and claiming a question	2
4 The shared system backbone	4
5 Shared artifacts, repository practice, and change control	5
6 Individual paper and shared system dossier	7
7 Schedule and cumulative checkpoints	8
8 Minimum persuasive evidence by project type	10
9 Assessment	11
10 Responsible use, services, and public release	12
11 Final package, and oral defence	13

1 Purpose, setting, and evidence boundary

The course project is a research studio in machine learning and clinical data science. Its purpose is to support original, careful work that joins scientific reasoning with sound engineering practice. Strong work may later contribute to a workshop paper, conference paper, journal article, technical report, teaching resource, or public software release. Publication is not guaranteed and is not required for success in the course.

The course uses **AIRO-Sim**, a fully synthetic cardiac-surgery teaching release. AIRO-Sim is related to **AIRO**, an applied-research program that uses approved real clinical data, but the two are not interchangeable. AIRO-Sim contains no patient-level AIRO or Maritime Heart Centre data. Its fields are schema-grounded, while its frequencies, associations, notes, and benchmark performance are deliberate teaching choices.

Three levels of claim

- **Schema-grounded:** many field names, concepts, and coding conventions correspond to plausible registry materials.
- **Clinically plausible:** temporal relationships and dependencies designed to support learning and analysis.
- **Not empirically representative:** frequencies, associations, note language, and model performance are not estimates from the Maritime Heart Centre population nor any other clinical population.

Non-clinical-use warning

This project will not produce a medical device and must not be used for clinical decision-making. Every record and note is synthetic. Reports, repositories, screenshots, interfaces, and presentations must state this clearly. Course findings may support claims about the analyses performed on AIRO-Sim, but they do not establish clinical validity, safety, effectiveness, fairness, approval, or patient benefit.

A project may concern data quality, missingness, labels, static or dynamic prediction, calibration, uncertainty, naturalistic text, multimodal methods, privacy, fairness, robustness, explainability, user interfaces, human factors, workflow, monitoring, governance, or another operational question with a substantive empirical or computational evaluation. A more complicated model is not automatically a stronger contribution.

2 Shared platform, individual studies

The project has two connected layers. The class will maintain a shared technical platform, while each student will pursue an individual research question that can stand on its own.

Shared	Clinical setting; AIRO-Sim release and docs; selected anchor outcome(s); common terminology; outcome and snapshot contracts; selected baseline infrastructure; case-viewer and monitoring interfaces; peer review; shared system dossier.
Led by each student	Confirmed research question; analytical choices; decisive experiments; interpretation; limitations; reproducibility record; final written and oral presentation.
Stewarded by each student	At least one shared artifact or interface on which the class can rely, with a written contract, acceptance checks, version history, and a backup reviewer.
Final products	Individual research papers and reproducibility packages; one short shared system dossier; one integrated, fully synthetic demonstration.

Two forms of responsibility

Every student will have both a **research lead responsibility** and a **shared-artifact stewardship responsibility**. These may be closely related, but they are not the same. E.g., a ML modelling question might rely on stewardship of the inference wrapper, or a demographic confound question might steward the cohort code, or a workflow question might steward the case-viewer code.

2.1 Flexible workstreams

There are no fixed or necessary roles in this project. Workstreams may include:

- clinical framing, intended use, workflow, and human factors;
- cohort construction, measurement, labels, missingness, and leakage;
- baseline and comparative ML modelling;
- calibration, thresholds, uncertainty, and error analysis;
- explainability, subgroup analysis, safety, and governance;
- case viewer, inference service, packaging, and reproducibility;
- monitoring, drift, audit logs, and operational response;
- T1 dynamic prediction, NLP, or multimodal analysis.

A workstream may have more than one student, or one student may connect more than one workstreams. Critical artifacts must have a primary steward and at least one reviewer.

2.2 Leadership, collaboration, and attribution

Leadership means accountability, not ownership or authority. Shared data, code, interfaces, and scientific ideas remain available to the class. Students may reuse shared artifacts and may collaborate on analyses that genuinely serve more than one question. Reuse must be attributed, and every paper must make the student's independent contribution clear.

3 Choosing the anchor outcome and claiming a question

3.1 Anchor outcome

AIRO-Sim v1.0.0 was designed with postoperative delirium as its primary target, but the course may choose another feasible shared outcome. The anchor outcome will be selected by 18 September 2026.

Candidate	Relevant AIRO-Sim fields	Minimum decision checks
Postoperative delirium	<code>comp_neuro_delirium</code> ; optional <code>delirium_onset_hours</code> for timing studies	Event count and missingness; temporal definition; compatibility with T0 and optional T1 questions; note and case-viewer support.
Sternal surgical-site infection	<code>comp_infect_stern_super</code> , <code>comp_infect_stern_deep</code> , or a justified composite	Event count and uncertainty; missing-value and parent-child rules; whether superficial and deep outcomes are combined; feasibility across the class portfolio.

The class may consider another outcome only if the release supports a clear, reproducible definition and enough evidence for several distinct projects. It must state the population, index operation, prediction time, outcome horizon, positive/negative/missing rules, exclusions, and any composite construction.

The release need not change when the outcome changes

The source release will normally remain frozen. Selecting SSI rather than delirium should produce a derived outcome contract, label file, and outcome-specific case fixtures, not a resynthesized dataset. AIRO-Sim will be patched for a verified blocking defect or an incorrect release artifact, not to improve event rates, simplify a project, or favour a model.

A student may study a secondary outcome with approval, but the shared platform will use one anchor outcome so that interfaces, monitoring, and the system dossier remain coherent.

3.2 Research topic allocation

Research topic allocation is not first come, first served. Interests will be considered together so that the final portfolio is fair, feasible, complementary, and useful to the shared system.

1. **Explore (10–16 September 2026).** Review the project map, clinical pathway, Technical Brief, data dictionary, sample cases, and known dependencies.
2. **Register ranked interests.** By 16 September 2026, enter up to three ranked research interests and up to three stewardship preferences in the pinned Research Topic Registry in MS Teams. Include a short rationale, skills you hope to use or develop, and important dependencies.
3. **Reconcile (17–18 September 2026).** The class and instructor compare scope, overlap, critical-path risk, coverage, and learning value. The anchor outcome and provisional leads will be posted by 18 September 2026.
4. **Submit the question card.** A one-page ‘question card’ is due 21 September 2026.
5. **Confirm.** Questions will be confirmed, or returned with a specific revision request, by 22 September 2026. The proposal then develops the confirmed question and is due 24 September 2026.

Question card

A confirmable question card identifies: (1) the one-sentence research question; (2) why it matters; (3) the population, time, outcome, or object of study; (4) the intended contribution or claim; (5) required data and computational resources; (6) comparator and evaluation plan; (7) shared dependencies; and (8) a foreseeable risk and fallback.

3.3 Examples of claimable research questions

The following examples show an appropriate level of specificity. They are not assignments.

Workstream family	Example question
Data and measurement	How do alternative missing-target and parent-child consistency rules alter the estimated prevalence and downstream calibration of the selected outcome?
Static modelling	Does a transparent generalized additive model match or outperform a nonlinear tree model after calibration, and where do their errors differ?
Dynamic prediction	Does information available at CVICU arrival improve calibration or case-ranking beyond the T0 system without temporal leakage?
NLP or multimodal analysis	Does (synthetic) preoperative note text add robust information beyond structured T0 features after controlling for template and documentation shortcuts?
Evaluation and decisions	Which threshold policy gives the most defensible balance between review workload and (synthetic) events detected, under uncertainty?
Explainability and safety	How stable are patient-level explanations under correlated features, missingness indicators, and alternative model families?
Workflow and interface	Which presentation of risk, uncertainty, provenance, and non-use language best supports accurate interpretation in a structured scenario evaluation?
Monitoring	Which data-quality and model-behaviour signals detect the seeded shadow-period shift earliest without excessive false alarms?

3.4 Rescoping

A justified rescope request is due 26 October 2026. Rescoping is appropriate when the data audit reveals an unreliable outcome, too few events, an unresolvable temporal ambiguity, an unavailable dependency, computational constraints, or a question that proves trivial. The request must state the reason, revised claim, affected shared artifacts, and fallback. Accepted rescopes become the final working scope by 29 October 2026.

4 The shared system backbone

The shared system should be more substantial than a research notebook but much smaller than a clinical deployment. It must run locally, use synthetic cases only, and make its limitations explicit.

4.1 Required minimum system

By the final submission, the class will provide:

1. a versioned outcome, cohort, snapshot, and split contract;
2. a reproducible T0 feature and modelling pipeline with at least one transparent baseline;
3. a saved preprocessing and model artifact with a documented inference interface;
4. a case viewer that shows the synthetic case, prediction time, permitted data, risk estimate, uncertainty or confidence information, provenance, missingness, and non-clinical-use notice;
5. a locally runnable container image or equivalent package tested with [Docker](#) or [Podman](#);
6. a monitoring demonstration using the shadow period or another approved simulation, including at least one data-quality signal and one model-behaviour signal;
7. a concise data statement and model/system card; and
8. a short shared system dossier that explains the architecture, versions, contracts, shared evidence, limitations, and individual contributions.

The system may use a lightweight API, a direct local application backend, or another sensible architecture. The interface need not mimic an electronic health record. Reliability, legibility, provenance, and reproducibility matter more than visual polish or infrastructure complexity.

Deployment language

The final system is a **containerized, deployment-oriented teaching prototype**. It is not a deployable clinical product. It must run in a local or course-approved environment, use only synthetic data, and demonstrate shadow-mode thinking rather than live clinical use.

4.2 Acceptance gate for the T0 backbone

The class should not add major extensions until the following checks pass:

- one row per operation and unique `operation_id`;
- repeated operations from one `patient_id` remain in one partition;
- outcome and missing-outcome rules are explicit and reproducible;
- all T0 predictors have a defensible preoperative availability rule;
- no outcome, discharge, T1, T2, or outcome-only note information enters T0;
- preprocessing is fit on training data only;
- a simple baseline can be trained and evaluated from a documented command;
- the prediction table passes its schema checks;
- the case viewer can render the supplied synthetic sample cases; and
- the container builds and starts on at least two course machines or environments.

4.3 Earned extensions

T1 dynamic prediction, NLP or LLM work, multimodal modelling, T2 landmark analysis, or a richer service architecture are extensions. An extension may begin after the dataset-analysis checkpoint when its upstream T0 contracts and acceptance checks are stable, the extension has a named lead, and the instructor agrees that it will not endanger the required backbone.

Novelty alone is not a reason to expand scope. An extension should answer a distinct, meaningful research question or provide a necessary shared capability.

5 Shared artifacts, repository practice, and change control

A shared artifact is any file, schema, script, model, case fixture, figure, or decision record on which another student's work may depend. Shared artifacts should be small enough to understand and stable enough to use.

5.1 Artifact contract

Every material shared artifact must record:

1. **Steward and backup reviewer;**
2. **Purpose and dependants;**
3. **Inputs**, including paths, schemas, snapshots, and assumptions;
4. **Outputs**, including paths, schemas, and version identifiers;
5. **Guarantees** that downstream work may rely on;
6. **Acceptance checks;**
7. **Known risks and fallback;** and
8. **Version history**, including downstream analyses that must be rerun.

Shared artifact	Typical steward	Minimum documentation
Outcome contract and labels	Measurement or data lead	Definition, horizon, missingness, exclusions, event count, version, generation command.
Cohort and feature tables	Data or modelling lead	Unit of analysis, joins, snapshot, columns, checksums, transformation code, leakage checks.
Reference predictions	Modelling lead	Data/model version, preprocessing, split, output schema, seed, inference command.
Evaluation package	Evaluation lead	Metrics, denominators, uncertainty method, threshold policy, calibration outputs, schema checks.
Case-viewer object	Interface or safety lead	Prediction time, modality, provenance, missing-value behaviour, required warnings, validation rules.
Container and run interface	Systems lead	Build command, run command, ports, mounted inputs, expected outputs, tests, limitations.
Monitoring output	Monitoring lead	Signals, reference windows, trigger rules, delayed-label assumptions, response or rollback owner.

5.2 Repository practice

The class will use one private shared GitHub repository during the term. The exact layout may evolve, but the following rules apply:

- use issues or another visible record for material tasks, decisions, and blockers;
- protect the main branch once shared interfaces begin to stabilize;
- make substantive changes to shared interfaces through a reviewed pull request;
- obtain at least one review from an affected student for a change that can alter another project;
- keep exploratory notebooks separate from scripts or documented notebooks that generate final evidence;
- record environments, seeds, model versions, prompt or schema versions, and sources of nondeterminism;
- never commit instructor-only files, credentials, real patient information, or copied clinical notes;
- tag or archive the exact shared versions used by each final paper.

Minimal change-control rule

A change that can alter another student's results must be reviewed, versioned, announced, and linked to the analyses that may need to be rerun. Small local experiments do not require this process.

5.3 Weekly status updates

A brief project stand-up should answer:

1. What did I learn since the last stand-up?
2. What evidence supports it?
3. What usable artifact changed?
4. What dependency, decision, or risk needs attention?
5. What is the next smallest decisive analysis or integration step?

6 Individual paper and shared system dossier

6.1 One growing individual paper

For each student, each checkpoint revises and extends the **same individual manuscript**. Students may use a common L^AT_EX template and shared bibliography resources, but there is no shared class paper. Each final submission should read as one coherent research paper rather than a set of assignments pasted together.

Every final paper should contain, adapted sensibly to the respective research question:

1. title and abstract;
2. introduction, motivation, intended use or analytical purpose, and precise contribution;
3. literature review or related work as a synthesis rather than a list of summaries;
4. research question and hypotheses, or explicit analytical objectives;
5. data, cohort, outcome, timing, missingness, and synthetic-data limitations;
6. methods, preprocessing, baselines, split or resampling strategy, and reproducibility details;
7. decisive experimental results or analyses, including uncertainty, sensitivity, and error review where appropriate;
8. discussion, failures, alternative explanations, limitations, and what the evidence supports;
9. operational, privacy, equity, safety, and governance implications where relevant;
10. conclusion and responsible next steps;
11. references and appendices.

Purely descriptive essays are not sufficient. A project centred on governance, privacy, workflow, or human factors still requires a substantive method and empirical evaluation, such as a structured audit, simulation, prototype evaluation, threat model, reproducible case analysis, or controlled comparison.

6.2 Shared system dossier

The class will prepare one concise dossier, normally three to five pages plus links to repository documents. It should cover:

- the anchor outcome and intended non-clinical use;
- system architecture and prediction snapshots;
- versions of the cohort, features, model, case-viewer schema, container, and monitoring package;
- the shared acceptance checks and their results;
- a compact account of what the individual studies collectively showed;
- shared limitations, unresolved risks, and the next evidence stage;
- an accurate contribution statement.

The dossier is not a substitute for an individual paper and should not duplicate long passages from those papers.

7 Schedule and cumulative checkpoints

7.1 Allocation, reviews, and scope decisions

Date	Milestone	Purpose
10–16 September 2026	Explore	Review the project map, Technical Brief, data dictionary, sample cases, outcome candidates, and dependencies.
16 September 2026	Ranked interests due	Submit research interests, stewardship preferences, rationale, skill goals, and dependencies.
17–18 September 2026	Reconcile	Compare overlap, feasibility, critical-path risk, and portfolio coverage.
18 September 2026	Anchor outcome and provisional leads posted	Publish the outcome contract draft, provisional research leads, stewards, backups, and dependencies.
21 September 2026	Question card due	Submit the question, intended contribution, evidence plan, risks, dependencies, and fallback.
22 September 2026	Questions confirmed	Questions are confirmed or returned with a specific revision request.
26 October 2026	Rescope requests due	Submit a justified rescope after the data audit, including effects on shared artifacts.
29 October 2026	Design review	Formative review of the claim, assumptions, comparator, acceptance checks, dependencies, and feasibility.
29 October 2026	Scope lock	Accepted rescopes become the final working questions unless exceptional evidence requires another change.
9–13 November 2026	Fall break	No project deliverable.
19 November 2026	Evidence review	Formative review of preliminary results, failures, alternative explanations, and the most decisive remaining analysis.

7.2 Cumulative manuscript and system checkpoints

Stage	Due	Individual paper	Shared system state
1	24 September 2026	1–2 pages: introduction , question, intended use or purpose, hypotheses/objectives, evidence plan, dependencies, risk register, and project plan.	Outcome contract draft; repository and stewardship map; dossier skeleton; initial system boundary.
2	8 October 2026	3–4 pages: literature synthesis , evidence map, strongest comparator, later validation or critique, and unresolved gap.	Shared terminology and literature map; initial intended-use and non-use statement.
3	22 October 2026	5–7 pages: data/cohort , provenance, labels, timing, missingness, event counts, synthetic-data limitations, and data audit.	Frozen outcome/cohort contract; T0 feature policy; split protocol; leakage and missingness checks; outcome-specific case fixtures if needed.
4	5 November 2026	8–10 pages: reproducible methods , preprocessing, baselines, metrics, uncertainty, acceptance checks, and experimental design.	Runnable T0 baseline; prediction schema; case-viewer contract; container build; monitoring plan; approved extension decisions.
5	26 November 2026	12–15 pages: results , uncertainty, calibration or decision consequences where relevant, robustness, subgroup or temporal checks, error analysis, and preliminary discussion.	Integrated demonstration candidate; frozen shared artifacts; monitoring results; dossier draft; final challenge interface ready.
6	9 December 2026	Up to 20 pages: complete paper , abstract, mature discussion, limitations, implications, conclusion, references, appendices, and reproducibility package.	Final dossier; tagged repository; runnable container; case viewer; monitoring demonstration; model/system card and data statement.

Oral presentations and the integrated demonstration are scheduled for 3 December 2026.

7.3 Checkpoint expectations

The **literature** review should normally include at least five recent core scientific publications, ordinarily from 2023–2026, and at least five additional publications that establish broader context, methods, or canonical mechanisms. Older work remains appropriate when foundational.

The **dataset analysis** must define the unit of observation, identifiers, linkages, outcome, prediction cut-off, missingness, event counts, repeated operations, splits, and limits of the synthetic evidence. It should focus on analyses that can affect downstream work. This is the formal opportunity to request a rescue.

The **methods** checkpoint must describe all transformations, baselines, model or analytical procedures, grouping and split rules, hyperparameter strategy, uncertainty method, metrics, and acceptance checks well enough for another researcher to reproduce the planned work.

The **experiments** checkpoint must provide enough evidence to support, weaken, or reject the stated claim(s). A few aggregate metrics are not sufficient. Explain errors, failures, alternative explanations, uncertainty, and whether the evidence changed the claim. Ask follow-up questions, and try to answer them.

8 Minimum persuasive evidence by project type

The exact evidence should follow the question. These examples are not exhaustive:

Project type	Examples of minimum persuasive evidence
Prediction or modelling	Transparent baseline; defensible patient/time split; calibration and uncertainty; threshold or decision consequences where relevant; error analysis; comparison with a stronger method.
Cohort, outcome, or data-quality study	Cohort diagram; provenance; missingness and impossible-value analysis; sensitivity to alternative definitions; leakage audit; documented effect on downstream conclusions.
Dynamic or survival analysis	Time origin; event and censoring definitions; competing-event handling where relevant; landmark or updating design; time-dependent evaluation and calibration.
NLP or clinical text	Corpus characterization; transparent baseline; temporal and template leakage audit; error taxonomy; robustness to note type, time, and documentation practice; factuality or extraction checks where generation is used.
Privacy study	Explicit threat model; attack, disclosure-risk, or privacy-budget evaluation; utility comparison; repeated-release reasoning; subgroup and equity effects.
Fairness or subgroup study	Definition of the relevant harm; denominators and uncertainty; measurement and selection analysis; comparison of mitigations; clear statement that synthetic groups cannot establish real institutional disparities.
Human–AI or workflow study	Named user and task; prototype or scenarios; evaluation protocol; observable outcomes; workload and failure analysis; evidence about interpretation or use, not only model output.
Monitoring or governance study	Operational scenario; monitored signals; reference windows and triggers; incident simulation; named response, authority, rollback, and accountability plan.

9 Assessment

The project mark will be assigned using the following proportions.

Component	Weight	Standard
Literature review	10%	Extensive, current, accurate synthesis of relevant work and evidence.
Dataset analysis	15%	Judicious statistics, provenance, timing, outcome construction, missingness, event counts, and claim-boundary analysis.
Methods	15%	Correct, reproducible technical description and well-justified experimental design.
Experiments and analysis	30%	Decisive experiments, valid procedures, uncertainty, error and robustness analysis, and thoughtful interpretation.
Technical quality and reproducibility	15%	Implementation and testing; reproducible environments and commands; reliable stewardship and integration of shared artifacts.
Written presentation	10%	Coherent research-paper argument, precise writing, effective figures and tables, and correct attribution.
Oral presentation and defence	5%	Clear, accurate, audience-aware presentation and demonstrated understanding of individual and shared work.

The literature review, dataset analysis, methods, and experiments and analysis components will be evaluated at their checkpoint and again in the final paper; the higher mark for each component will be retained. The remaining components are assessed from the final submission, stewardship record, and presentation.

Because final reports are individual, grades are based on the student's own paper, analyses, code, decisions, stewardship, and demonstrated understanding. The shared dossier and demonstration are required class artifacts, but there is no single class-wide mark that can hide unequal contributions.

Question registration, the question card, proposal, design review, evidence review, contract updates, and risk-register updates are required formative stages.

10 Responsible use, services, and public release

10.1 Synthetic data and privacy

Keep the phrase “fully synthetic – not for clinical use” visible in reports, screenshots, demos, logs, and exported cases. Do not describe AIRO-Sim records or notes as de-identified patient data, sampled MHC cases, or examples of MHC documentation style. Do not add real patient information, real clinician names, copied clinical notes, credentials, or restricted material to the project.

The student release and any instructor-only materials must remain separate. Student work must not depend on latent truth, anomaly keys, hidden annotations, benchmark predictions, or an unreleased final challenge set.

10.2 LLMs and hosted services

NLP and LLM work is optional. No student is required or expected to pay for a model, API, cloud account, or other service. Every required deliverable must have a no-cost local or institutionally available path. The course does not reimburse service costs.

A student who uses a hosted or local model must record the provider or model, version or access date, prompt or extraction schema, data sent, outputs retained, cost if any, and reproducibility limits. Unsupported statements, prompt injection, temporal leakage, negation and uncertainty errors, non-determinism, and fluency without evidence must be evaluated as risks. Permission to use synthetic notes in the course does not imply that the same service would be acceptable for real clinical data.

10.3 Clinician input

Clinician review may be useful if it becomes available, but it is optional and not guaranteed. It is not a requirement for any grade or acceptance test. Informal feedback must not be described as clinical validation. Any activity that goes beyond ordinary course or guest feedback and involves people as study participants requires prior instructor approval and applicable institutional processes.

10.4 Public release after the course

Selected work may be prepared for public release after grading, subject to instructor review, university policy, appropriate licences, and accurate attribution. Grades do not depend on a student’s agreement to publish an individual paper or code contribution. Before release, the class must remove grades, peer-review comments, credentials, instructor-only materials, answer keys, unreleased challenge data, and any content that overstates the relationship to AIRO or clinical practice.

A public release should state that it was produced in DGIN 5401 with AIRO-Sim, distinguish AIRO-Sim from AIRO, identify contributors, record software and data versions, and preserve the non-clinical-use boundary.

10.5 Research integrity

- Cite borrowed text, ideas, figures, data, code, prompts, and analytical procedures as appropriate.
- Attribute shared class artifacts, including baseline code, data audits, diagrams, definitions, and reviews.
- Write the final paper independently; do not copy another entity’s prose even when questions are adjacent.
- Never fabricate reported data, citations, experiments, outputs, or statistical results.
- Disclose material use of automated tools in `contributions.txt` and in the methods or acknowledgements when it affected the scientific work.
- Preserve enough of the analytical record to explain how each reported result was obtained.

11 Final package, and oral defence

11.1 Final submission package

Each student submits:

- the final paper as a PDF and editable source;
- all code used to generate analyses, figures, and tables;
- a README with exact run instructions and expected outputs;
- an `environment.yml`, `requirements.txt`, `pyproject.toml`, or equivalent dependency specification;
- fixed seeds or a documented source of nondeterminism;
- frozen versions of shared artifacts used in the paper;
- `contributions.txt`, identifying the student's work and all material shared, human, and automated contributions.

The class submits the shared dossier, tagged repository state, container build and run instructions, model/system card, data statement, case viewer, monitoring demonstration, and contribution record. The instructor must be able to inspect and, within reasonable resource limits, run the work.

11.2 Presentation and defence

The oral presentation should communicate the motivation, question, data, methods, key evidence, limitations, and next evidence stage to a mixed clinical-technical audience. It should not be a compressed reading of the paper.

Students should be prepared to answer:

1. What is your distinct contribution?
2. Which decision in your work was most consequential?
3. What shared assumption or artifact did you inherit?
4. What did another student inherit from your stewardship work?
5. What result would break or materially weaken your conclusion?
6. Which result is least trustworthy, and why?
7. What failed or changed your mind?
8. What can AIRO-Sim not establish?
9. How did your work affect the shared system?
10. What is the next responsible evidence stage?

Final check

Can a careful reader reconstruct the question, outcome, data-generating assumptions, analysis, comparison, uncertainty, limitations, dependencies, shared-artifact versions, and exact claim without guessing?