

APPLIED AI / MATERNAL HEALTH / RELIABLE SYSTEMS

Designing safer maternal-risk monitoring

From a leakage-controlled model to a verified local-and-hosted monitoring, alerting and recovery prototype.

Winner Olokede MSc Systems Engineering • University of Lagos Research case study • 2026

A maternal-risk prediction is not the same as a care response. I designed and evaluated a research prototype that treats model inference as **one component of an auditable monitoring and escalation workflow**—including input checks, uncertainty handling, alert states and simulated connectivity failure.

451 CLEANED ANALYSIS RECORDS	90 HELD-OUT TEST RECORDS	86.4% HELD-OUT HIGH-RISK SENSITIVITY	90/90 LOCAL / HOSTED CLASSIFICATION MATCHES
--	------------------------------------	--	---

Evidence boundary: The system is a reproducible software research prototype using secondary data associated with Bangladesh. It is **not** a clinically validated diagnostic tool, does not establish Nigerian patient performance and has not been deployed in BRIMMHQ care operations.

What I set out to establish

Can we build a defensible model?

Audit the underlying data, limit feature-profile leakage, evaluate multiple model families, and choose a risk alert rule using training-only evidence.

Can we build a safer surrounding system?

Reject unusable measurements, abstain outside training support, persist assessment and alert states, and verify the hosted build matches the frozen research model.

The engineering question was not only “Can it predict?” but also “What does the system do when prediction, communication or follow-through becomes uncertain?”

RESEARCH DESIGN • DATA ENGINEERING • MODEL EVALUATION • API IMPLEMENTATION • RELIABILITY TESTING

OFFICIAL MSc RESEARCH PROJECT TITLE

Design and Evaluation of an AI-Enabled Remote Monitoring System for Maternal Health Risk Assessment in Resource-Constrained Healthcare Settings

01 / SYSTEM ARCHITECTURE

The model sat inside a monitored workflow.

Two implementations shared one frozen decision logic: a reproducible Python reference and a hosted browser-to-TypeScript system.

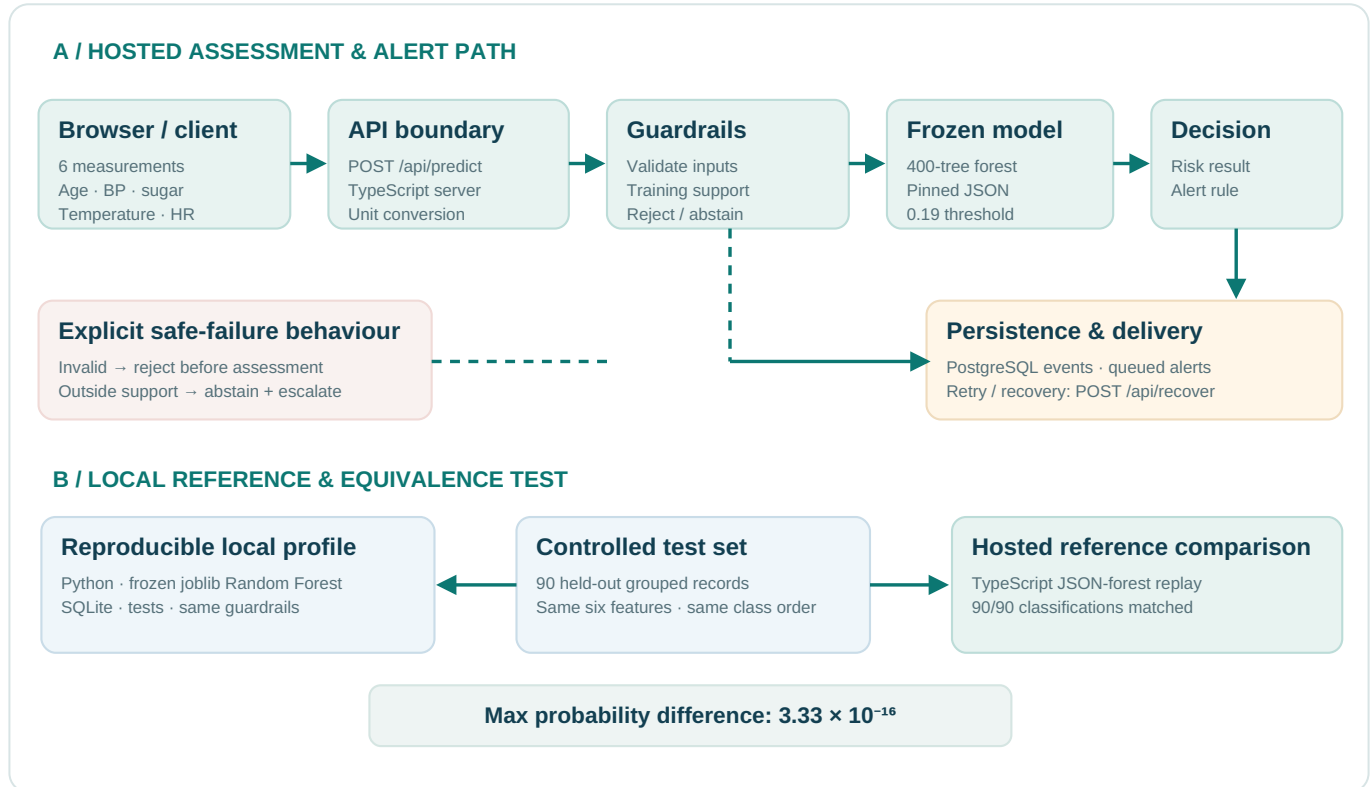


Figure 1. Redrawn from the dissertation's logical architecture (Fig. 3.2) and documented hosted implementation. Solid routes show implemented components; dashed paths summarise guardrail outcomes. Diagram is explanatory, not a network-security certification.

Six measurement inputs

Age; systolic and diastolic blood pressure; blood sugar; body temperature; heart rate. The UI accepts Celsius and converts it at the model inference boundary.

What the server records

Assessment, rejection/abstention status, model output when supported, alert state, queue/retry information and an auditable trail of events.

The hosted interface calls `POST /api/predict` and uses `POST /api/recover` for queued alert recovery. The hosted TypeScript implementation executes a pinned JSON representation of the frozen 400-tree forest; it does not depend on Python running on the researcher's laptop.

Source: MSc dissertation, Sections 3.12–3.15. Platform-assisted portions of the browser/TypeScript build are credited on the final page.

02 / RISK & ALERT LOGIC

A safe failure matters as much as a successful prediction.

The prototype separates input validity, applicability of the trained model, inference, and alert delivery.

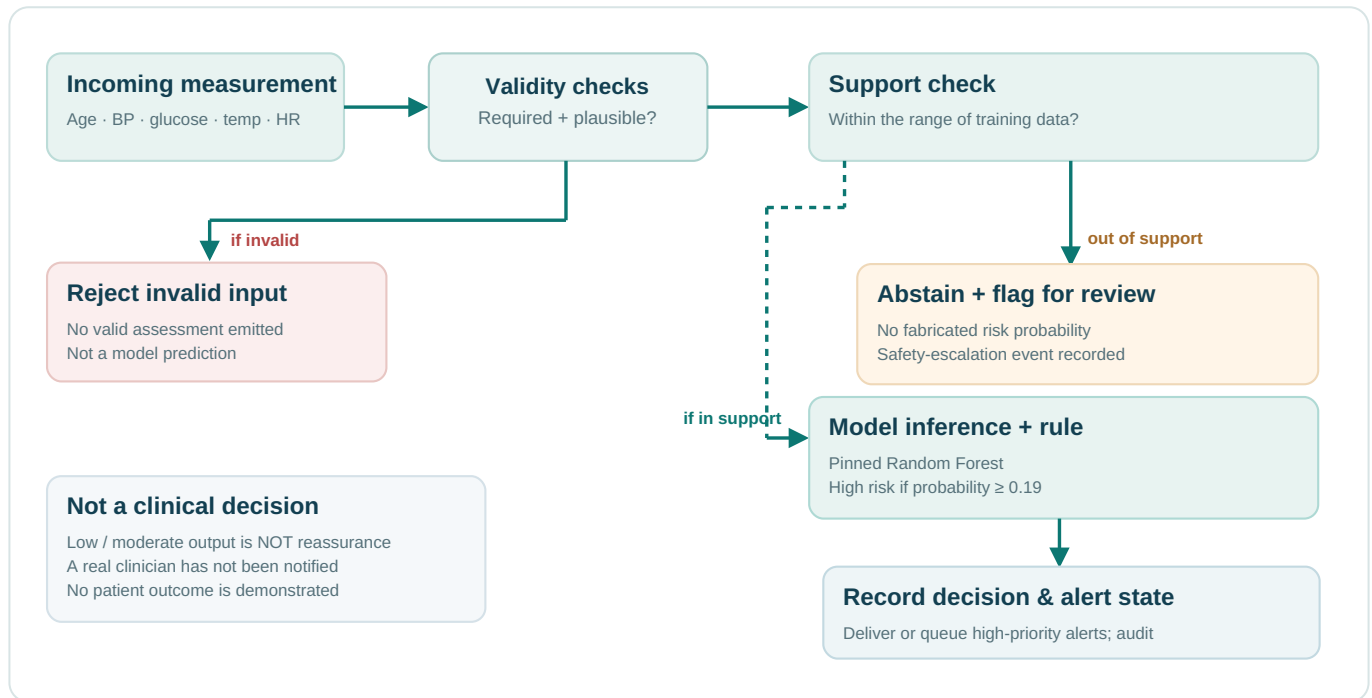


Figure 2. Safety decision logic reconstructed from the documented implementation. “Abstain” means no model probability is manufactured; it is a different outcome from ordinary low-risk prediction.

Delivery is modelled as an explicit state transition

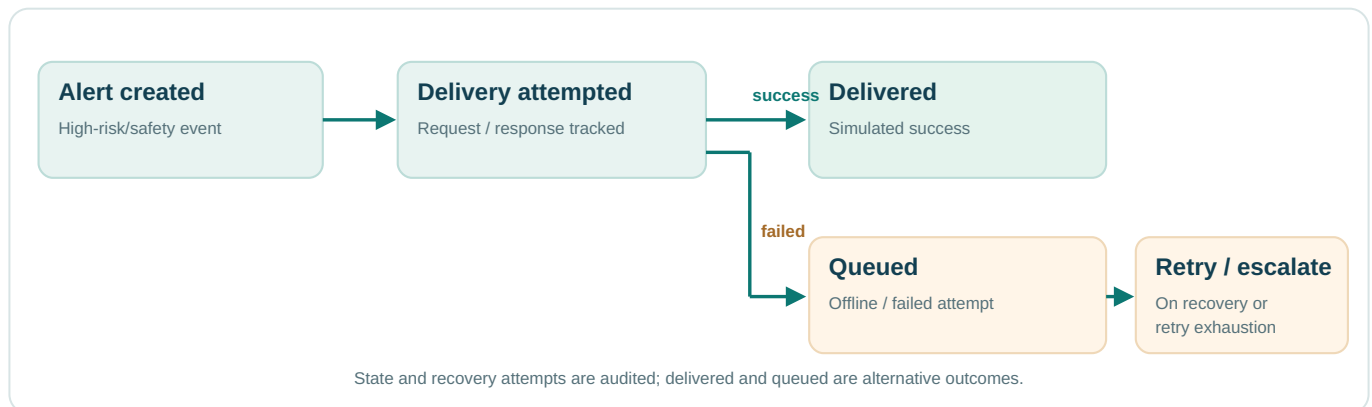


Figure 3. Simplified alert-delivery states based on the dissertation’s finite-state alert model (Fig. 3.2a). Queuing, retries and recovery were tested under simulated network profiles; no real clinical follow-up is implied.

When higher-risk results or safety-abstention conditions require attention, the system records an event and attempts, queues or retries an alert. That is testable software behaviour—not evidence that a health professional saw or acted on an alert.

03 / DATA & MODELLING

Make the modelling pipeline defensible before reporting accuracy.

The downloaded file was audited before the data were split, and identical six-feature profiles were kept out of both partitions simultaneously.

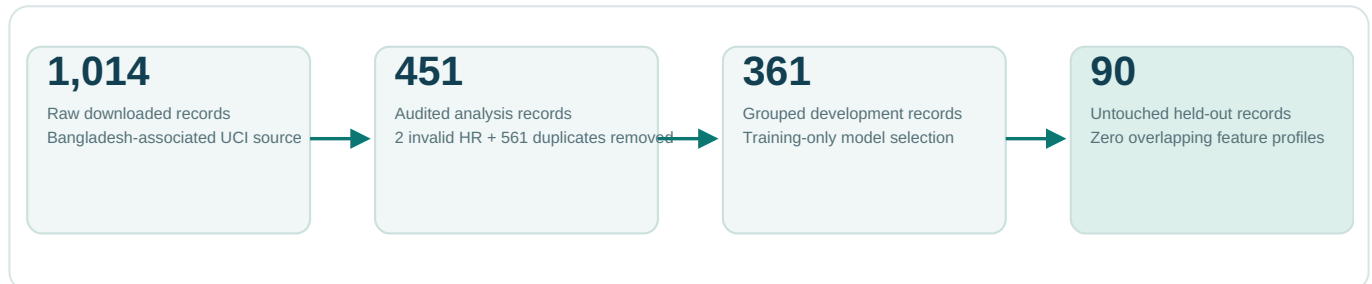


Figure 4. Data-flow and partitioning summary. The locally downloaded file contained 1,014 records; the current UCI landing-page count can differ. Removing two implausible heart-rate rows and 561 exact duplicates left 451 analysis records.

Data quality and leakage

Grouped splitting protected against identical measurement profiles appearing in both development and held-out partitions. Thirty-five conflicting-label profiles were retained as ambiguity, not silently “corrected.”

Model families compared

Dummy baseline; logistic regression; decision tree; Random Forest; support vector machine; histogram gradient boosting. Candidate selection used training-only grouped validation.

An explicit operational trade-off

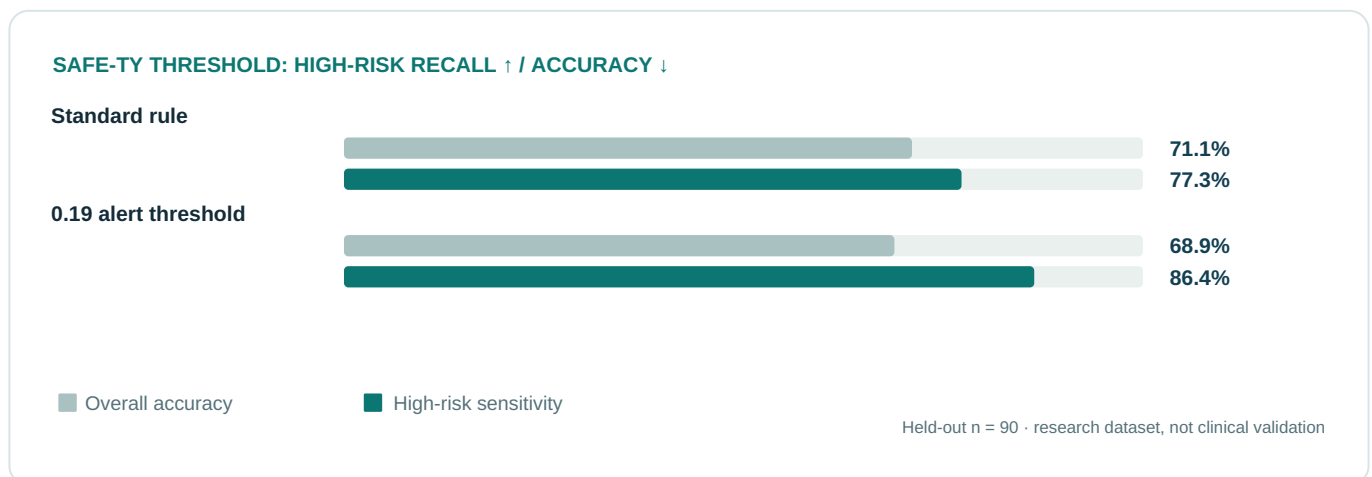


Figure 5. Held-out comparison for the selected Random Forest under its ordinary decision rule versus a training-selected 0.19 high-risk alert threshold. Alert sensitivity increased as overall accuracy declined.

Applying the safety-oriented threshold increased high-risk sensitivity from **77.3% to 86.4%** but lowered overall accuracy from **71.1% to 68.9%**—a deliberately documented trade-off, not evidence that the threshold is appropriate for real clinical use.

04 / EVALUATION & LIMITATIONS

The numbers are informative—and imperfect.

Held-out performance was assessed after the model-selection rule and high-risk threshold were chosen using development data only.

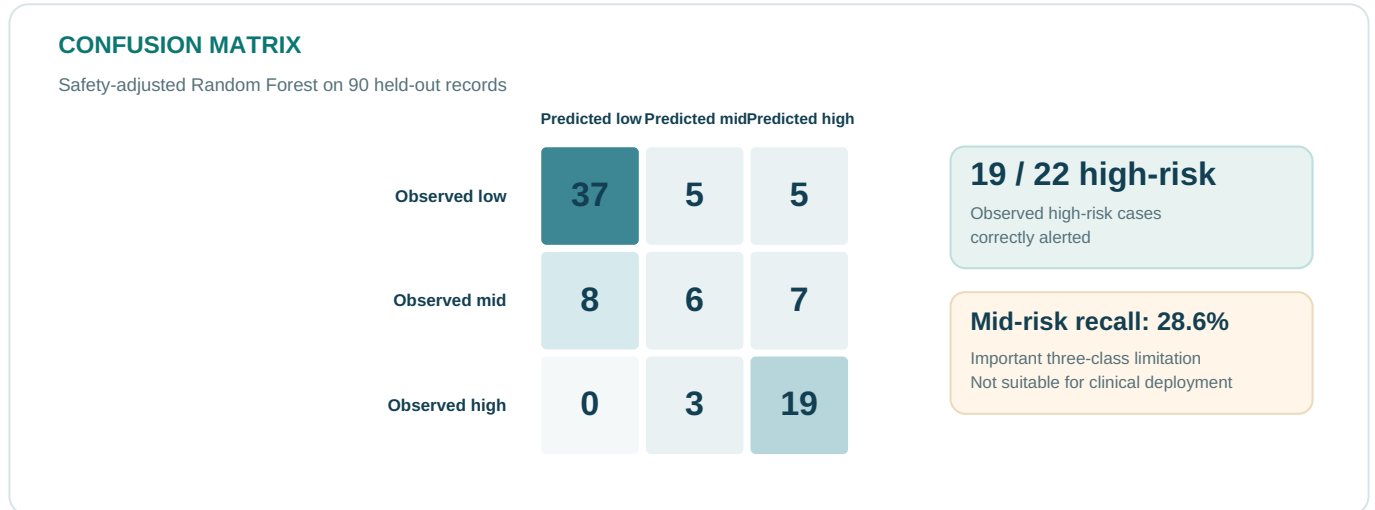
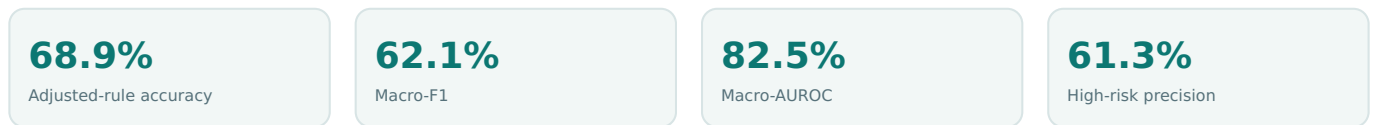


Figure 6. Recreated from dissertation Fig. 4.4. Rows = observed class, columns = predicted class; numbers sum to 90 test records. The weak mid-risk detection is an important limitation.



Read these as research metrics. Only 22 observed high-risk records were in this held-out partition. High-risk sensitivity was 19/22; 3 high-risk records were still missed. A retrospective split of a secondary dataset cannot establish patient safety, clinical calibration, fairness or deployment suitability in Nigeria.

The safety rule generated more alerts to capture additional high-risk records. This made mid-risk performance especially weak (6 of 21 correctly classified) and increased false alerts from other classes. The trade-off underscores why prospective, clinician-led evaluation is necessary.

Source: dissertation Figures 4.3–4.4 and Tables 4.8–4.9. Percentages are rounded from the recorded metrics.

05 / IMPLEMENTED PROTOTYPE

A research interface made the model behaviour inspectable.

The dissertation includes an illustrative Celsius-input screen showing synthetic measurements, a high-risk output and a caregiver alert state.

Maternal Risk Monitor

Research prototype — not validated for clinical use

Age: 40 years

Systolic BP: 140 mmHg

Diastolic BP: 100 mmHg

Blood sugar: 15.0 mmol/L

Body temperature: 36.7 °C

Heart rate: 90 bpm

Assess risk and generate alert

HIGH-RISK SCREEN

High-risk probability: 98.2%

Safety threshold: 19.0%

Alert status: **CAREGIVER ALERT**

Largest local probability changes

- Systolic BP: +18.5 percentage points
- Blood sugar: +12.6 percentage points
- Diastolic BP: +3.8 percentage points

Explanation is diagnostic and non-causal.

Figure 7. Cropped from the actual dissertation's Figure 3.3 (Illustrative Celsius prototype input and high-risk output). Values shown are **synthetic**, not a patient record. The interface and underlying prototype have not been clinically validated.

Implementation and verification evidence

90/90

Same held-out classes in Python and hosted TypeScript

11/11

Local automated prototype tests passed

13

Hosted regression tests reported in dissertation

The frozen Python joblib reference and hosted JSON-forest representation produced identical classes for all 90 held-out records, with maximum class-probability difference 3.33×10^{-16} against a specified 10^{-12} tolerance. Checks covered unit conversion, invalid values, outside-support abstention, alert creation and outage recovery.

Seven seeded communication profiles covered normal and degraded connectivity, complete outage, and subsequent recovery. These were controlled simulations, not tests of hospital workflows, a real SMS carrier or field connectivity.

The implementation used the Lovable development environment to assist with portions of the browser/TypeScript application and deployment, as disclosed in the dissertation. Research specification, design decisions and evaluation evidence were led by the author.

TRY THE LIVE RESEARCH PROTOTYPE

maternal-risk-monitor.lovable.app

Use fictional or prefilled measurements only. Submitted inputs and outputs are stored in the research audit database; do not enter real patient information.

Try the assessment, abstention rules, and simulated connectivity / alert recovery.

Not clinically validated. The simulated alert does not contact a healthcare provider.

06 / WHAT IT PROVES—AND WHAT IT DOES NOT

Technical proof is necessary. Clinical readiness is a different bar.

The work demonstrates a testable systems-engineering approach, not a deployable maternal-risk diagnosis product.

What the prototype establishes

Traceable requirements; reproducible data preparation; grouped evaluation; an explicit safety-oriented threshold; invalid-input and abstention states; two matching execution profiles; and reproducible software and network simulations.

What remains unproven

Prospective Nigerian patient performance, clinical safety and effectiveness, regulatory approval, security certification, wearable integration, live hospital connections, or confirmed clinician response to alerts.

What I would require before real-world use

Appropriate research and data permissions; prospective Nigerian validation under clinician oversight; review of calibration, decision thresholds and potential subgroup disparities; secure operations, human escalation responsibilities, external interoperability testing and applicable regulatory assessment.

How this connects to BRIMMHQ

BRIMMHQ is a **separate human-led maternal-support venture** working on continuity between formal healthcare visits. It does not deploy this research classifier in clinical care. My transferable learning is about structured follow-up, reliable escalation, auditability and knowing when technology should hand responsibility back to people.

Takeaway: A prediction only becomes useful within a system that can handle incomplete evidence, constrained connectivity, human accountability and what happens next.

Research provenance and attribution

1. **Research:** Oloke, Winner Abiola (2026). *Design and Evaluation of an AI-Enabled Remote Monitoring System for Maternal Health Risk Assessment in Resource-Constrained Healthcare Settings*. MSc Systems Engineering research project, University of Lagos.
2. **Dataset:** Ahmed, M. (2020). *Maternal Health Risk* [dataset]. UCI Machine Learning Repository. <https://doi.org/10.24432/C5DP5D>. CC BY 4.0. Source rows were audited, deduplicated and grouped; the charts and summary figures are research-derived analyses.
3. **Guidance:** Professor O. S. Asaolu (academic supervision); Dr John Ogbemhe (technical redevelopment guidance). Acknowledgement does not imply endorsement of this independent portfolio presentation.
4. **Visuals:** Architecture, flow and charts redrawn from the dissertation's documented methods/results. Figure 7 reproduces an illustrative prototype interface from the submitted dissertation; values are synthetic.
5. **Build assistance:** Lovable assisted with portions of hosted browser/TypeScript implementation and deployment; research specifications, model/evaluation decisions and verification claims remain attributable to the author.

This is a professional research case study, not the full dissertation. No participant records, stakeholder quotations or identifiable patient information are reproduced. Any public release remains subject to applicable university publication and IP requirements.