# Gaurav Tuli – Applied ML Portfolio (Projects 01–08, Concise 2025)

## Preface

* Data & AI thought leader with 10+ years of experience turning ML innovations into real-world health solutions.
* PhD-trained in applied machine learning and network science at Harvard Medical School and Virginia Tech.
* Proven track record delivering end-to-end machine learning solutions in healthcare, from concept to deployment.
* Nationally recognized innovator: won a DHS bio-threat detection challenge and served as a White House panelist.
* Led large-scale AI collaborations (with Meta, Google, CDC, and top universities) to advance global pandemic surveillance and privacy-preserving health research.
* Published 30+ research papers (PNAS, Lancet Digital Health, etc.) and featured by major media for impactful AI-driven healthcare projects.

## Project 01: Bio-chemical Threat Detection (Pandemic Pulse – DHS Hidden Signal Challenge Winner)

**1. At a Glance**  
- Developed a real-time bio-threat surveillance tool integrating social, search, and health data.  
- Won a national innovation challenge, gaining White House recognition for bio-threat detection.  
- Enabled outbreak alerts days ahead of official reports, showcasing AI’s public health impact.

**2. The Problem**  
- Traditional outbreak surveillance lags weeks behind, missing early signals of new threats.  
- Critical signals spread across social media, web searches, and news are siloed and hard to monitor in real time.  
- Novel bio-threats start with sparse, noisy digital indicators that conventional systems often overlook.

**3. The Solution**  
- Unified data pipeline continuously ingests multi-source feeds (Twitter, Google Trends, news) for anomaly detection.  
- Tiered filtering by pathogen priority and source credibility isolates the most credible weak signals.  
- Automated NLP and statistical algorithms flag unusual symptom clusters or location-based spikes in real time.

**4. Architecture Overview**  
- Modular cloud pipeline on AWS with engines for data extraction, enrichment, event detection, visualization, evaluation, and alert sharing.  
- Real-time ingestion from APIs (Twitter, Google) and RSS feeds funnels into a processing queue with feedback loops to refine signal quality.  
- Custom NLP enrichment and geo-clustering components clean data and add context (e.g., mapping tweets to locations) for accurate detection.  
- An interactive dashboard displays mapped alerts and timelines, updating continuously for early situational awareness.  
- A feedback mechanism updates keywords and data sources based on detected events, improving sensitivity over time.

**5. Results and Impacts**  
- Successfully flagged real events (e.g., Boston Marathon bombing, Hepatitis A outbreak) in retrospective tests, proving earlier detection than official reports.  
- Outperformed 69 other solutions to win DHS’s Hidden Signal Challenge, earning national recognition and stakeholder buy-in (including a White House panel invitation).  
- Delivered a functional prototype dashboard tested by public health officials, validating the tool’s practical use for early outbreak warning.

**6. Skills and Tools Used** (technical skills table)

| **Skill/Tool Category** | **Application in Project 01** |
| --- | --- |
| Data Engineering & APIs | Python pipelines for real-time ingestion (Twitter API, Google Trends, RSS feeds) |
| Natural Language Processing | Regex filtering and keyword extraction to identify health threat signals |
| Cloud Infrastructure | AWS cloud deployment for scalable processing of high-volume streaming data |
| Algorithms & Analytics | Statistical anomaly detection, clustering algorithms, and custom social-media geolocation |
| Visualization & Dashboard | D3.js and map libraries to build an interactive outbreak monitoring dashboard |
| Collaboration & Management | Cross-sector coordination (federal agencies, health officials) and agile project management (Trello, GitHub) with strict data governance |

**7. Cross-Project Capabilities**  
- Real-time integration of diverse, noisy data sources – a skill later applied to other domains (e.g., fusing IoT sensor feeds or hospital data streams).  
- Applying AI to public health: experience aligning technical solutions with epidemiological needs, a theme continued in pandemic monitoring and clinical ML projects.  
- Cross-sector collaboration: honed ability to unite government, healthcare, and tech partners, replicated in later projects like COVID-19 surveys and hospital safety models.

**8. Published Papers/Tools**  
- **Challenge White Paper:** *“Pandemic Pulse: Using Digital Exhaust to Detect Bio-threats”* – internal paper outlining the winning methodology.  
- **Prototype Dashboard:** Interactive web dashboard (internal) demonstrated multi-source outbreak alerts in one view for agencies.  
- **Related Research:** Prior social-media surveillance studies (e.g., detecting foodborne illness from Twitter) provided the foundational methods used in this project.

## Project 02: Clinical Diagnostic ML – Food Allergy Risk Stratification

**1. At a Glance**  
- Created a machine learning–based composite biomarker panel for accurate food allergy risk assessment.  
- Outperformed any single test, reducing unnecessary oral food challenges and associated risks.  
- Improved patient safety by identifying which children can safely undergo allergen exposure versus those at high risk.

**2. The Problem**  
- Standard allergy tests (IgE, skin prick) often yield ambiguous results and can’t predict true allergic reactions reliably.  
- The gold-standard oral food challenge is risky, resource-intensive, and used sparingly, leaving many patients undiagnosed or in limbo.  
- Rising food allergy rates heighten the need for better diagnostics to distinguish truly allergic patients without putting them through dangerous procedures.

**3. The Solution**  
- Built an ML model that combines multiple inputs (specific IgE levels, skin test results, patient history) to predict if a patient is truly allergic or likely tolerant.  
- Trained on real hospital data (patients with known oral challenge outcomes) so the model “learns” the patterns of feature combinations that signal a true allergy.  
- Designed a prototype pipeline (“ML-Allergy”) integrated with the EHR: it pulls a patient’s lab results and history, runs the ML risk algorithm, and outputs a risk score with guidance to clinicians.

**4. Architecture Overview**  
- Six-component ML pipeline from data extraction to decision support. It automatically pulls relevant lab results and clinical history from the EHR data warehouse via ETL processes.  
- Feature engineering handles ~20+ predictors (IgE, ratios, symptoms, etc.), accommodates missing data, and uses regularization to rank important features and avoid overfit.  
- Iterative model training with cross-validation tests various algorithms (logistic regression, random forest, gradient boosting) to maximize predictive AUC while maintaining generalizability.  
- The final model achieved high performance (AUC ~0.96) and was configured to emphasize safety (e.g., tuning thresholds to minimize false negatives for allergies).  
- The output is presented in a clinician-friendly format within workflow: a risk score (“High risk – 85% chance of reaction”) along with key contributing factors, seamlessly delivered alongside existing lab reports.

**5. Results and Impacts**  
- The ML risk model reached ~96% accuracy (AUC ~0.96) in validation, far exceeding the predictive power of any single test, which translates to much more confident clinical decision-making.  
- Combining multiple test results proved its value: the model clearly separated allergic vs. tolerant patients, confirming that multiple weak indicators together create a strong signal.  
- In practice, this enables doctors to avoid unnecessary and risky food challenges for low-risk patients and to focus resources on those truly at risk, improving safety and reducing patient anxiety.

**6. Skills and Tools Used** (technical skills table)

| **Skill/Tool Category** | **Application in Project 02** |
| --- | --- |
| Data Collection (EHR) | SQL and Python (Pandas) to extract and merge allergy test results and clinical notes from hospital databases (Epic EHR) |
| Machine Learning (Python) | Scikit-learn & LightGBM for model development; cross-validation, grid search, and regularization (L1/L2) for feature selection and tuning |
| Statistical Analysis | ROC/AUC analysis, bootstrapped confidence intervals, and custom threshold tuning to maximize negative predictive value (safety first) |
| Clinical Domain Integration | Incorporated medical expertise (e.g., weighting “history of anaphylaxis” appropriately); close collaboration with allergists to embed domain logic in the model |
| Healthcare Data Standards | Used ICD-10 and LOINC codes to identify data in EHR; leveraged FHIR resources to integrate the model output into clinical systems (ensuring compatibility and privacy) |
| Communication & Visualization | Presented results in clinician-friendly terms (e.g., “avoid X% of unnecessary challenges”) and used clear visual aids (calibration plots, decision curves) to gain physician buy-in |

**7. Cross-Project Capabilities**  
- **Clinical Decision Support Development:** End-to-end experience building a clinical ML tool (data ingestion, model, workflow integration) carried into later projects like ICU decision support and maternal-infant care tools.  
- **Interdisciplinary Collaboration:** Skill in partnering with clinicians (allergists in this case) to ensure ML solutions meet real-world needs – later applied with ophthalmologists (Vision project) and intensivists (ICU project).  
- **Regulatory/Ethical ML Practice:** Navigated patient data privacy (HIPAA, IRB approvals) and clearly communicated model limitations to stakeholders – a critical skill in all healthcare AI projects dealing with sensitive data.

**8. Published Papers/Tools**  
- **Internal White Paper:** *“ML Composite Biomarker Panel for Food Allergy Diagnosis”* – detailed proposal circulated within the hospital to outline the project’s design and rationale.  
- **Hospital Knowledge Sharing:** Findings presented at Boston Children’s Hospital grand rounds and an innovation showcase, raising awareness of ML’s potential in diagnostics.  
- **Prototype Tool:** Developed a pilot “Allergy ML Risk Calculator” (Excel interface + Python backend) for clinicians to input patient data and get a risk estimate, now undergoing evaluation for integration into the EHR’s decision support system.

## Project 03: Pediatric Vision Assessment (Pre-verbal Visual Acuity “Vision Ninja”)

**1. At a Glance**  
- Created an iPad-based gaze-tracking app (“Vision Ninja”) to objectively measure visual acuity in infants and toddlers.  
- Cut infant eye exam time by ~30%, replacing slow manual methods with a quick, engaging digital test.  
- Leveraged machine vision (OpenCV) to assess eyesight in pre-verbal children, enabling detection of vision issues at ages where standard eye charts fail.

**2. The Problem**  
- Standard infant vision exams (like observing if a baby follows objects or using cards) are slow, subjective, and often imprecise.  
- Subtle vision problems (e.g., mild lazy eye or focus issues) can go undetected until a child is older and able to cooperate with traditional tests, missing a crucial early treatment window.  
- Exams for toddlers can be stressful and inefficient – keeping a young child’s attention is hard, and lengthy appointments often yield incomplete or unreliable results due to fatigue or fussiness.

**3. The Solution**  
- Developed a tablet “game” that displays high-contrast shapes (optotypes) and uses the device’s front camera with OpenCV to track the infant’s eye movements in real time.  
- The app employs an adaptive algorithm: it shows a shape and detects if the baby’s gaze fixes on it. If the child looks, the shape gets smaller next round; if not, it gets bigger – homing in on the smallest visible size (like a digital eye chart).  
- The testing process is mostly automated – the examiner just starts the test and the system adjusts stimulus size, logs each response, and determines the infant’s visual acuity threshold without requiring verbal input from the child.

**4. Architecture Overview**  
- **Device & Vision Components:** An iPad with its camera runs the app; OpenCV-based gaze detection identifies whether the infant looks at each shown shape within a set time window.  
- **Adaptive Staircase Algorithm:** The app’s logic dynamically adjusts the optotype size up or down based on the infant’s responses, following a staircase strategy to pinpoint the smallest discernible shape.  
- **Minimal Manual Interaction:** The interface is simple for clinicians – enter basic patient info and tap start. The system handles presenting stimuli and deciding when to stop (once acuity threshold is found), while logging all gaze responses.  
- **Data Logging:** Every stimulus presentation and outcome is recorded internally. These logs can be analyzed (in Python) to fine-tune parameters (e.g., how long to display shapes, how long a “meaningful gaze” must last) and improve future test versions.  
- **Prototype Platform:** Built as an iOS app (Swift) with integration of OpenCV libraries. The design emphasizes engaging visuals (simple shapes, sounds) to maintain the baby’s attention and uses on-device processing for immediate feedback.

**5. Results and Impacts**  
- Early prototype testing showed exam times reduced by roughly one-third, making the vision test much faster and less taxing for infants and parents.  
- Provided quantitative eye exam results for infants (e.g., exact smallest shape recognized), replacing subjective judgments with objective metrics that can be tracked over time.  
- Promises earlier identification of vision issues (like amblyopia or high refractive errors) in babies, allowing interventions (glasses, patching) to begin during a crucial developmental window rather than waiting years.

**6. Skills and Tools Used** (technical skills table)

| **Skill/Tool Category** | **Application in Project 03** |
| --- | --- |
| Computer Vision (OpenCV) | Real-time face and eye detection to track infant gaze on an iPad camera feed; customized for infant face proportions |
| Mobile App Development | Built an iOS app in Swift, integrating camera APIs and OpenCV frameworks for on-device processing |
| UX/UI Design (Toddlers) | Created a baby-friendly interface with high-contrast shapes and engaging feedback; timed presentations to align with infant attention spans |
| Data Logging & Analysis | Implemented in-app logging of stimuli and responses; used Python (pandas) offline to analyze gaze response patterns and adjust algorithm thresholds |
| Usability Testing (Agile) | Conducted iterative tests with infants and gathered feedback from pediatric ophthalmologists, rapidly refining the prototype in short development cycles |
| Clinician Collaboration | Worked closely with eye specialists to translate gaze-tracking results into familiar clinical terms (approximate Snellen equivalents) and ensure trust in the new method |

**7. Cross-Project Capabilities**  
- Human-centered design of AI tools: this project shows the importance of tailoring technology to end-users (babies and clinicians), a user-first approach that Dr. Tuli applies when designing clinician dashboards or patient-facing AI tools in other projects.  
- Real-time signal processing: experience gained in processing video feedback within seconds parallels his work in streaming data analysis for outbreak detection and ICU monitoring, demonstrating proficiency in developing time-sensitive analytics.  
- Rapid prototyping & stakeholder buy-in: by building a compelling prototype with clear “wow factor,” he secured stakeholder support – a tactic he repeated in other initiatives (e.g., quick demos for the mother-infant dashboard) to accelerate funding and adoption.

**8. Published Papers/Tools**  
- **Prototype Documentation:** *“Vision Ninja – Validation Prototype Definition Notes (Oct 2022)”* – an internal design and results report used to brief the project team and hospital innovation leaders.  
- **Conference Abstract:** Prepared *“Gaze-Tracking for Infant Visual Acuity: Feasibility Study”* for submission to a vision science conference (ARVO), aiming to present the methodology and early results to the research community.  
- **Prototype App & Demo:** The Vision Ninja iPad app (prototype) was demonstrated internally via a video and live demos, helping the team win a **FastTrack Innovation in Pediatric Healthcare 2022** award. (Not yet a public app, but recognized in-house for its innovative approach.)

## Project 04: ICU Clinical Decision Support – Continuous Hypoxemia Monitoring

**1. At a Glance**  
- Developed a continuous, noninvasive method to estimate arterial blood oxygen (PaO₂) using standard ICU monitor data.  
- Enabled real-time tracking of patient oxygenation in the ICU, outperforming traditional static formulas especially at high oxygen saturation levels.  
- Published in a peer-reviewed study, demonstrating improved guidance for managing ventilator settings and oxygen therapy in critical care.

**2. The Problem**  
- Dangerous drops in blood oxygen (hypoxemia) can occur between intermittent arterial blood gas tests, since these invasive checks are done only a few times a day.  
- Pulse oximeter readings (SpO₂) are continuous but become unreliable at the top end (97–100% SpO₂) – they can’t distinguish a safe patient from one with dangerously high oxygen levels.  
- Existing surrogate formulas for PaO₂ are often inaccurate, especially at saturation extremes, and don’t adapt to individual patient differences, making clinicians hesitant to trust them.

**3. The Solution**  
- Collected a large dataset of paired ABG measurements and bedside monitor readings (~52,000 samples) from ICU patients, and used it to develop better estimation models.  
- Explored machine learning approaches (including neural networks using inputs like recent SpO₂ trends and heart rate) and also optimized a conventional oxygenation equation by tuning physiological parameters to fit real patient data.  
- Selected a new empirically tuned equation (“Sauthier ePaO₂”) that provided the best accuracy. Implemented it for continuous use: the ICU monitor (or a linked computer) takes in streaming SpO₂ and heart rate and computes an estimated PaO₂ every minute, displaying it to clinicians and triggering alerts if oxygen drops.  
- Built in real-time quality control: the algorithm watches the difference between pulse oximeter heart rate and EKG heart rate as a noise indicator – if they diverge, it treats the SpO₂ data as unreliable and adjusts or flags accordingly.

**4. Architecture Overview**  
- **Data Pipeline:** Retrieved high-frequency vital signs data around each ABG draw from two hospitals’ systems (merging data logged every 5 seconds from one system and every 1 second from another, standardized to 5-second intervals). Stored all combined data in a PostgreSQL database for analysis.  
- **Analysis & Modeling Environment:** Used Python and R extensively (Pandas, NumPy, SciPy, scikit-learn, Keras/TensorFlow) to clean data, engineer features (e.g., recent SpO₂ averages, slopes, heart-rate disparities), and train/test multiple candidate models.  
- **Model Evaluation:** Employed rigorous statistical evaluation – generated Bland-Altman plots to assess agreement between estimated and actual PaO₂, calculated intraclass correlation coefficients (agreement), and examined error across different SpO₂ ranges to ensure reliability even at high saturations.  
- **Deployment Design:** Specified how the final algorithm would run on existing ICU monitors: a lightweight script or plugin continuously ingests real-time vitals, applies the estimation formula, and outputs a refreshed PaO₂ estimate every few seconds. This can feed into the bedside monitor display (trend graphs, ARDS index calculations) and alarm systems for immediate clinical use.  
- **Safety Considerations:** Built into the design fail-safes like capping displayed values beyond validated ranges (e.g., instead of an implausible number at extreme highs, show an indicator like “PaO₂ very high”) and ignoring data during sensor disconnections or extreme noise, ensuring the tool would be trusted by clinicians.

**5. Results and Impacts**  
- The optimized estimator provided sensible, stable PaO₂ readings even when SpO₂ was 100%, overcoming the “infinite reading” problem of traditional equations and giving clinicians meaningful data at the high end of oxygenation.  
- Continuous monitoring enabled earlier detection of hypoxemia trends – for example, the system could alert if a patient’s oxygen was quietly deteriorating between blood gas tests, allowing interventions (like ventilator adjustments) sooner and improving ARDS management by tracking severity in real time.  
- The work was published in *Critical Care Explorations* (2021) and has since been cited in multiple studies, influencing the field of digital critical care. It provided a blueprint for integrating predictive algorithms into ICU workflows, with discussions underway to pilot the estimator in a real ICU setting.

**6. Skills and Tools Used** (technical skills table)

| **Skill/Tool Category** | **Application in Project 04** |
| --- | --- |
| Data Engineering (ICU) | SQL (PostgreSQL) and Python to manage ~65k blood gas records with high-frequency vital sign data; automated data extraction around events |
| Statistical Analysis | Advanced stats: Bland-Altman agreement analysis, bootstrapped confidence intervals, hypothesis tests to validate model accuracy and bias |
| Machine Learning & Modeling | Keras/TensorFlow to prototype neural networks for regression; exhaustive parameter search to optimize a physiological equation for best fit |
| Domain Knowledge (Critical Care) | Integrated understanding of oxyhemoglobin dissociation curves, ARDS criteria, and physiological factors (P50 shifts) to guide modeling and interpret results |
| Software Development | Python scripting for simulating real-time monitoring (feeding sequential data to the model); implemented the final formula in an efficient form suitable for bedside device integration |
| Collaboration & Publication | Coordinated a multi-center team (Boston & Montreal) under data-sharing agreements; co-authored a peer-reviewed journal paper and produced publication-quality figures (e.g., error distribution charts) |

**7. Cross-Project Capabilities**  
- Translating data science research into a clinical tool: experience in going from retrospective analysis to a validated model ready for real-world use (as done here) is directly applicable to other projects where ML models must be deployment-ready (e.g., the allergy risk model or mother-infant dashboard logic).  
- Time-series data expertise: skills in handling and analyzing high-frequency time-series (ICU monitor streams) are transferable to other domains, including real-time social media feeds in outbreak surveillance or sensor data in IoT health projects.  
- Multi-site data integration: learned to standardize and combine data across different hospital systems, a capability that proved valuable in later projects focused on data interoperability (such as linking records in the mother-infant health information exchange project).

**8. Published Papers/Tools**  
- **Journal Publication:** *Critical Care Explorations* (Oct 2021) – “Estimated PaO₂: A Continuous and Noninvasive Method to Estimate PaO₂ and Oxygenation Index.” This open-access paper details the methodology and validation results, serving as the primary reference for this work.  
- **Open Science Resources:** The team shared portions of the de-identified dataset and analysis code via an online repository referenced in the publication, enabling other researchers to reproduce or build upon the findings.  
- **Community Impact:** The method has been discussed in critical care forums and was highlighted in a critical care innovation newsletter, indicating its relevance. Its successful publication and citations (including by a 2024 digital medicine article) underscore its influence on subsequent research in noninvasive monitoring.

## Project 05: Maternal-Infant Health Data Integration – National Standards Mapping

**1. At a Glance**  
- Mapped 180 key data elements for substance-exposed mother–infant pairs to national health data standards to identify interoperability gaps.  
- Found over 91% of required mother-baby health data can be represented with existing standards (USCDI v4 and FHIR R4).  
- Published in JAMIA (2025), providing the first framework to link maternal and infant electronic records for coordinated care in the opioid crisis context.

**2. The Problem**  
- Mother and baby medical records are usually kept separate (often in different systems), making it hard for caregivers to see the full picture for substance-exposed mother–infant pairs.  
- Important information specific to these dyads (e.g., maternal opioid treatment plans, infant withdrawal scores, social services involvement) may not be recorded in standardized fields, hindering data sharing and care coordination.  
- Privacy laws around substance use (like 42 CFR Part 2) and reliance on slow, after-the-fact data (e.g., insurance claims) further complicate timely information exchange, leaving providers “flying blind” when trying to help these vulnerable mothers and babies.

**3. The Solution**  
- Conducted a comprehensive crosswalk of critical clinical and social data elements needed for managing substance-exposed mother–infant pairs, mapping each one to fields in national interoperability standards (USCDI and FHIR). This identified which elements are already covered and which are missing.  
- Surveyed Health Information Exchange (HIE) networks and hospitals to assess if and how they currently link mother and infant records, and whether they use available standards to do so.  
- Developed recommendations based on the findings: for instance, suggesting additions to data standards for any gaps (ensuring future versions include fields for things like “maternal opioid treatment status”), and urging adoption of existing standards to support automatic mother-infant record linkage in healthcare systems.

**4. Architecture Overview**  
- **Cross-mapping Methodology:** Employed a blend of manual expert review and script-assisted search to match 180 specialized data needs to standard data definitions. Used spreadsheets as a catalog and wrote simple Python scripts to query FHIR/USCDI specifications for relevant terms, ensuring nothing was overlooked.  
- **HIE Survey & Analysis:** Designed a targeted questionnaire (mix of multiple-choice and short answer) for HIE organizations and hospital IT staff, asking about their capability to link mother-baby records and their use of FHIR/USCDI. Analyzed responses qualitatively and quantitatively (using Python for aggregating answers where applicable) to gauge the real-world interoperability landscape.  
- **Visualization of Results:** Created clear visual summaries of the mapping outcome – for example, a Sankey diagram to illustrate how many of the 180 elements map cleanly to standards versus how many fall outside. These visuals helped communicate where current standards suffice and where gaps lie.  
- **Expert Validation:** Worked iteratively with clinical experts (neonatologists, OB/GYN, addiction specialists) and informaticians to verify that all crucial data elements were captured and that mappings were semantically correct. This ensured the framework’s accuracy and credibility.  
- **Integration Points:** Outlined how the findings can be used in practice: e.g., showing that because 91% of needed data has standard fields, a dashboard (as in Project 6) can reliably pull most data via FHIR APIs today, while highlighting which remaining data might require custom handling or advocacy for standard updates.

**5. Results and Impacts**  
- Discovered that an overwhelming majority of required mother-infant data points (over 91%) are already representable within current USCDI and FHIR standards, a very encouraging finding – it means technical barriers to linking these records are relatively low if systems use the standards.  
- Identified the specific gaps (the ~9% of elements not covered) and provided concrete suggestions to standards bodies. This work was published in *JAMIA* (2025), immediately establishing a knowledge base for others trying to connect maternal and child health data.  
- The findings have directly influenced policy and practice: for example, the Office of the National Coordinator (ONC) cited this study in a health equity data guidance, using it as an example of crucial data exchange for at-risk populations. Within HHS, the mapped element list became the foundation for a “Dyad Data Tracker” initiative, essentially turning the research into a blueprint for a real system to monitor outcomes for substance-exposed families.

**6. Skills and Tools Used** (technical skills table)

| **Skill/Tool Category** | **Application in Project 05** |
| --- | --- |
| Healthcare Data Standards | Deep expertise in USCDI v4 and FHIR R4; navigated standard terminologies to map custom data needs to standard fields (ensuring semantic alignment) |
| Data Mapping & Scripting | Utilized spreadsheet tools and Python scripts to systematically catalog and match 180 data elements to standard data dictionaries; tracked coverage vs gaps with precision |
| Survey Design & Analysis | Crafted clear, targeted survey questions for HIE stakeholders; employed survey methodology best practices and analyzed responses (via pandas and manual review) to extract meaningful patterns |
| Stakeholder Collaboration | Coordinated with government (ONC, HHS) and clinical partners, communicating technical findings in accessible terms and incorporating their feedback to validate results |
| Data Visualization | Created explanatory visualizations (e.g., Sankey diagrams, coverage charts) using Python (Plotly) and other tools to convey complex mapping outcomes simply for decision-makers |
| Domain & Regulatory Knowledge | Quickly absorbed domain specifics (NAS scoring, maternal OUD care) and privacy regulations (HIPAA, 42 CFR Part 2) to ensure the mapping and recommendations accounted for real-world clinical and legal contexts |

**7. Cross-Project Capabilities**  
- **Standards-Based Integration:** Mastery of data standards (USCDI, FHIR) obtained here is directly applied in other projects that integrate data (e.g., using FHIR APIs for the dyad dashboard or pulling EHR data for the ICU alert system). Knowing the standards accelerates development and ensures interoperability in his solutions.  
- **Policy-Tech Bridging:** Demonstrated ability to operate at the intersection of health policy and technical implementation – a rare skill that means he designs solutions (across projects) that not only work technically but also align with and leverage national health IT initiatives, increasing their impact and adoption potential.  
- **Thorough Problem Scoping:** Showcased a strategic approach of surveying and understanding the broader ecosystem before building a tool. This research-first mindset (seen here in mapping the landscape, similarly in reviewing existing methods before new model development in other projects) ensures that each project addresses the right problem the right way and produces generalizable insights, not just a one-off fix.

**8. Published Papers/Tools**  
- **Peer-Reviewed Study:** Published in *Journal of the American Medical Informatics Association (JAMIA)*, 2025 – *“The substance-exposed birthing person–infant dyad and health information exchange in the United States.”* This paper provides the detailed mapping results and is a key reference for improving data sharing in this area.  
- **Supplementary Artifact:** A comprehensive supplementary spreadsheet listing all 180 data elements and their mapping to standard codes was released with the publication. This deliverable serves as a ready-to-use reference for health IT implementers looking to configure their systems for mother-infant data exchange.  
- **Knowledge Dissemination:** Findings were presented to the ONC’s interoperability working group in 2024 (timed with USCDI v4 discussions) and at the AMIA 2025 conference, spreading awareness in both policy and informatics communities. The project’s influence is further evidenced by ONC guidance documents and an internal HHS project (Dyad Data Tracker) that have incorporated its recommendations.

## Project 06: Maternal-Infant Dyad Dashboard – Coordinated Care Interface Prototype

**1. At a Glance**  
- Built a prototype “Dyad Dashboard” that links a mother’s and her infant’s electronic health records into one combined view.  
- Displays critical maternal and infant health data side-by-side, allowing clinicians to coordinate care for substance-exposed mother–baby pairs at a glance.  
- Implemented using real standards (FHIR/USCDI) on synthetic data, demonstrating a user-friendly solution for integrated mother-infant care within existing hospital workflows.

**2. The Problem**  
- Often a mother’s health record (in an adult or OB system) is completely separate from her baby’s record (in a pediatric system), forcing providers to flip between systems or request records and risking that important information (e.g., mother’s infectious disease status or treatment history) might be missed in the baby’s care.  
- Even if data sharing is technically possible, without a dedicated interface, relevant details can be buried in lengthy documents. For instance, a pediatrician might overlook the mother’s medication-assisted therapy details if it’s hidden in discharge paperwork, which could be crucial for interpreting the newborn’s condition.  
- Care coordination among OB, neonatal, and social work teams relies on manual communication (meetings, calls). Without a unified view, critical context can fall through the cracks. At the same time, simply dumping both charts together could overwhelm clinicians with information, so a carefully designed solution is needed to present combined data clearly.

**3. The Solution**  
- Designed and prototyped a web-based dashboard that securely pulls in data from both mother’s and infant’s health records and presents them in one synchronized interface (e.g., mother on the left, baby on the right).  
- Followed user-centered design: consulted with clinicians to determine what information is most vital for dyad care (e.g., maternal medications, recent toxicology results, infant NAS scores, feeding plans) and how to layout the interface for clarity (parallel timelines, matching sections for mother and baby).  
- Built the prototype as a SMART-on-FHIR app, meaning it can plug into an EHR system using standard FHIR APIs to fetch patient data if mother-infant linkage exists. It uses existing authentication (OAuth) and respects privacy settings while drawing the data together. The use of FHIR ensures the solution can work across different EHR systems with minimal custom integration.  
- Populated the prototype with realistic synthetic data (including using a large language model to generate sample patient records for testing). This allowed demonstration of the dashboard’s functionality (e.g., showing a mother’s methadone treatment alongside the baby’s withdrawal score trend) without using real patient data.

**4. Architecture Overview**  
- **UI/UX Design Process:** Started with wireframes and user journey mapping. Used tools like Balsamiq and Figma to design the dashboard screens, iterating on layout based on feedback from healthcare providers to ensure the interface was intuitive and not cluttered.  
- **Front-End Development:** Implemented the dashboard as a standalone web application using HTML/CSS/JavaScript. Leveraged SMART-on-FHIR JavaScript libraries to handle authentication and querying the FHIR server for patient data. For the prototype, an open-source FHIR server loaded with synthetic dyad data served as the backend.  
- **FHIR Integration:** Utilized FHIR resource queries to gather relevant data – e.g., pulling MedicationStatement and ProblemList from the mother’s record, and pediatric vitals and scores from the baby’s record. The app logic matches a mother-infant pair via a shared identifier or linking table (as would be configured in an HIE or hospital system).  
- **Generative Data Simulation:** When real data was unavailable, used GPT-4 and other tools to generate synthetic but plausible clinical data (like a sequence of NAS scores or progress notes), ensuring the prototype had rich content to display. This creative approach allowed testing features like trend graphs and alert flags in the UI with lifelike scenarios.  
- **Testing & Iteration:** Conducted informal usability testing sessions with a small group of NICU and OB clinicians. Observed how they interacted with the dashboard and noted any confusion or suggestions. For example, if testers had trouble finding a specific piece of information, the design was tweaked (such as adding a highlight for maternal labs that impact the baby). Repeated this cycle to refine the interface and functionality.

**5. Results and Impacts**  
- The prototype clearly demonstrated how bringing mother and infant data together improves situational awareness. Clinicians in feedback sessions noted the immediate benefit of seeing, for example, the mother’s last substance use treatment update right next to the infant’s latest withdrawal score – something that previously required digging through separate records.  
- Project sponsors (including HHS officials focused on the opioid crisis) were given a detailed demo and were impressed by the concept. This buy-in is crucial for moving toward a pilot; it showed stakeholders a tangible solution rather than just a concept, making it easier to secure support for next steps.  
- The concept gained national visibility: one of the prototype’s dashboard snapshots was featured (with permission) in an ONC interoperability presentation as a forward-looking example of using data standards to solve real clinical needs. Additionally, a health IT newsletter highlighted the Dyad Dashboard as an innovative use of FHIR to improve care coordination, bringing recognition to the approach.  
- The work is on track to proceed to a formal pilot within the hospital’s EHR environment. If successful, it can significantly streamline care for substance-exposed dyads – making it easier for care teams to collaborate and potentially becoming a model for other integrated care dashboards (e.g., linking records for other family-centric care scenarios).

**6. Skills and Tools Used** (technical skills table)

| **Skill/Tool Category** | **Application in Project 06** |
| --- | --- |
| UI/UX Design | User-centered design with wireframing (Balsamiq) and high-fidelity prototyping (Figma); created intuitive layouts for dual-patient data display |
| Web Development | HTML/CSS/JavaScript front-end build; implemented the SMART-on-FHIR framework for secure EHR integration and data retrieval |
| FHIR API Integration | Queried FHIR resources (e.g., medications, labs, notes) for two linked patients; used OAuth2 for authentication and handled JSON data parsing in the app |
| Synthetic Data Generation | Utilized GPT-4 and custom scripts to create realistic synthetic patient data (mother-infant pairs) for testing, ensuring the prototype had valid scenarios and datasets |
| Usability Testing | Planned and conducted hands-on tests with clinicians; gathered feedback and applied iterative improvements to the dashboard’s interface and features |
| Project Management | Managed multi-phase development (requirements -> data prep -> development -> testing -> demo); coordinated inputs from designers, developers, and clinical advisors using agile practices |
| Clinical Workflow Mapping | Ensured the app fit naturally into existing workflows by designing it as an in-EHR module (SMART-on-FHIR launch) so clinicians can access the dyad view without extra steps; accounted for privacy consent handling in design |

**7. Cross-Project Capabilities**  
- **Integrating Multi-Source Data:** Demonstrated the ability to synthesize disparate data streams into one coherent output, much like combining varied data in Pandemic Pulse or melding model outputs with UI in the ICU project. This skill of unifying information for better insight runs through Dr. Tuli’s portfolio.  
- **Marrying AI with UX:** Highlighted the importance of presenting analytical outputs in an accessible way. Dr. Tuli carries this forward, ensuring that complex model results (from allergy risk scores to outbreak alerts) are delivered via intuitive interfaces that busy end-users can easily interpret and act on.  
- **Innovation Leadership:** By spearheading the development of an entirely new type of dashboard, he showed he can lead cross-disciplinary teams to build not just models but full solutions (data backend to user interface). This holistic view – connecting technical backends to front-end usability – enhances the impact of his projects and is a common thread in his leadership roles.

**8. Published Papers/Tools**  
- **Internal Report & Demo:** Produced a comprehensive internal project report and a demonstration video for stakeholders and sponsors (including HHS opioid policy advisors). While not public, this documentation captures design decisions, UI screenshots, and summarizes usability test findings – serving as a blueprint for implementation and a case study for potential publication.  
- **Conference Dissemination:** Plans are underway to submit an abstract to AMIA 2025 or HIMSS 2025 detailing the Dyad Dashboard development and initial feedback. Sharing this work with the broader informatics community will invite collaboration and position the project as a leading example of standards-driven innovation in care coordination.  
- **Open-Source Prototype Code:** The code for the dashboard prototype (excluding any proprietary styling or data) has been made available in a GitHub repository under an open-source license. It includes sample synthetic FHIR data, allowing other hospitals or developers to explore and build upon the concept. This open approach encourages adoption and iterative improvement beyond the initial project.  
- **Recognition:** Featured in a 2025 healthcare IT innovation newsletter and included in an ONC presentation, the Dyad Dashboard has already achieved a degree of national recognition as a showcase of how leveraging interoperability standards can directly improve patient care for a high-need population.

## Project 07: COVID-19 Vaccine Hesitancy in Bangkok (Facebook Survey – JMIR 2023)

**1. At a Glance**  
- Used Facebook’s city-scale digital health survey to rapidly assess COVID-19 vaccine hesitancy during Bangkok’s mid-2021 outbreak.  
- Showed vaccine hesitancy declined about 7% each week as vaccination rates climbed, indicating improving public acceptance over time.  
- Identified top hesitancy reasons (fear of side effects and “wait and see” caution) and revealed that 97% of respondents – even the hesitant – trusted health experts most for COVID information, guiding targeted public messaging.

**2. The Problem**  
- In mid-2021, Bangkok faced a severe COVID wave, but vaccine supply was limited and public uptake was slow, hampering outbreak control.  
- Even among high-priority groups (“608” – seniors or those with chronic conditions), many were reluctant to get vaccinated despite government incentives, threatening to prolong the crisis.  
- Policymakers lacked timely local data on the extent of hesitancy or the reasons behind it, leaving them unsure how to adjust their vaccination campaign strategy. Conducting large in-person surveys was impractical during an active outbreak, so a faster, resource-light solution was needed.

**3. The Solution**  
- Leveraged the ongoing University of Maryland/Facebook daily COVID survey to gather responses from Bangkok residents about their vaccination status, intentions, and attitudes, providing near real-time insight during the outbreak.  
- Ensured data quality by weighting the survey responses to match Bangkok’s demographics and priority group proportions, and by checking the survey’s reported vaccination rates against official data. This assured officials that the trends observed were credible and not skewed by sampling biases.  
- Tracked vaccine hesitancy and acceptance week by week through the outbreak and added specific questions to pinpoint *why* people were hesitant (multiple-choice reasons) and *whom* they trusted for information. Employed statistical analysis (e.g., Kendall tau correlations) to verify trends and relationships in the data, lending rigor to conclusions drawn for policy use.

**4. Architecture Overview**  
- **Data Collection Scale:** Over ~34,000 survey responses were collected from Bangkok residents across five months, creating a rich dataset updated daily and aggregated weekly for analysis.  
- **Representativeness Checks:** Incorporated a routine to compare survey sample characteristics each week to known city population stats (age distribution, proportion of “608” high-risk individuals, etc.). If certain groups were underrepresented, weighting adjustments were applied so the results better reflected the actual population.  
- **Automated Dashboard:** Developed an internal dashboard that ingested weekly survey aggregates and displayed key metrics such as “% hesitant,” “% vaccinated,” and top reasons for hesitancy. This dashboard was shared with local health officials to quickly visualize how hesitancy was trending and which concerns were most cited, without waiting for a formal report.  
- **Deep Dive Analysis:** Used Python and R to conduct segmented analysis – for instance, splitting responses by age group or by hesitancy level to see if younger people had different concerns than older people. Performed chi-square tests to compare distributions of hesitancy reasons between groups, ensuring any highlighted differences were statistically significant and not due to chance.  
- **Data Pipeline:** The survey data (collected via Facebook) were accessed in batch form, then cleaned and aggregated. Each week’s results were appended to a time-series dataset. Scripts automated the generation of updated metrics and graphs, so that each week a fresh report could be delivered rapidly to city officials.

**5. Results and Impacts**  
- The survey achieved stable weekly sampling with respondent demographics that closely mirrored Bangkok’s population, giving confidence that the insights were broadly applicable. This consistency meant officials could trust that changes in the data reflected real shifts in public sentiment, not survey noise.  
- Self-reported vaccination rates in the survey increased in parallel with official vaccination statistics, while the proportion of people expressing hesitancy fell substantially week after week (roughly a 7% drop in hesitancy per week). This real-time evidence that hesitancy was declining helped reassure authorities that their outreach efforts were effective and encouraged them to stay the course and even intensify certain campaigns.  
- Crucially, the survey pinpointed why people were hesitant: the majority cited worries about side effects and a desire to “wait and see” as top reasons. Meanwhile, fringe anti-vaccine sentiments (like general anti-vax or religious objections) were very low (<10%). This guided officials to focus communications on vaccine safety and to provide updates about others safely getting vaccinated, rather than battling broad conspiracy theories.  
- The finding that 96.9% of respondents trusted medical professionals and scientists (far more than other sources like politicians or social media) directly influenced the city’s public messaging strategy. Knowing this, health authorities leaned on doctors, nurses, and respected medical academics to be the spokespeople in TV and online campaigns, ensuring the messages came from voices the public would heed.

**6. Skills and Tools Used** (technical skills table)

| **Skill/Tool Category** | **Application in Project 07** |
| --- | --- |
| Survey Data Analysis | Processed tens of thousands of survey responses using Python (Pandas) and R; performed data cleaning, demographic weighting, and trend calculations |
| Global Collaboration | Coordinated with Thailand’s Ministry of Public Health and academic partners, applying cultural and local context understanding to interpret data and shape relevant recommendations |
| Statistical Rigor | Used nonparametric correlation (Kendall tau) and chi-square tests to ensure observed trends and associations (e.g., hesitancy vs. time, reasons vs. age group) were statistically significant and not random fluctuations |
| Data Visualization | Built clear visual dashboards and weekly bulletins using Python (matplotlib/Plotly) to communicate key trends and findings to policymakers in an easily digestible format |
| Communication & Storytelling | Summarized complex survey data into actionable insights for city officials, translating numbers into clear narratives (e.g., “hesitancy dropping steadily; focus on addressing side-effect fears”) to guide decision-making |
| Public Health Partnership | Demonstrated skills in working with government stakeholders under urgent conditions – providing timely, relevant data analyses that were integrated into public health strategy on the fly |

**7. Cross-Project Capabilities**  
- Adaptable digital surveillance: showed how a global survey platform can be “zoomed in” to focus on a single locale’s needs, a flexibility he later applied to other projects (like tailoring survey-based methods to measure vaccine effectiveness in specific regions).  
- Behavioral insight generation: strengthened expertise in extracting and interpreting human behavioral and attitudinal data from large datasets – a capability useful across many domains, from patient experience analysis in hospitals to public opinion mining in different countries.  
- Cross-functional coordination: honed the ability to work in an interdisciplinary environment (epidemiologists, data scientists, government officials) and deliver results that satisfy all parties. This skill of aligning data science work with policy needs is a thread seen in many of Dr. Tuli’s projects.

**8. Published Papers/Tools**  
- **Peer-Reviewed Publication:** *JMIR Public Health and Surveillance* (2023) – “COVID-19 Vaccine Acceptance and Uptake in Bangkok, Thailand: Cross-sectional Online Survey.” This paper formally reports the study’s findings and methodology, contributing to the academic literature on urban vaccine hesitancy.  
- **Policy Brief:** Distilled the survey results into a concise policy report for Thailand’s health authorities. The brief highlighted the key drivers of hesitancy and recommended strategies (e.g., emphasize safety messaging, enlist healthcare professionals as communicators) to boost vaccine uptake, directly influencing the public health communication approach during the outbreak.

## Project 08: Privacy-Preserving COVID-19 Behavior Study (Google Health – Nature Digital Medicine 2024)

**1. At a Glance**  
- Pioneered a federated mobile health study via the Google Health Studies app to track people’s mask-wearing and social distancing habits without collecting personal data centrally.  
- Found that most individuals maintained masking and distancing until they were fully vaccinated, whereas those who never intended to vaccinate exhibited consistently lower adherence to these behaviors.  
- Demonstrated a successful blueprint for privacy-centric public health research using federated analytics and differential privacy during a pandemic, proving that valuable insights can be gained without compromising individual privacy.

**2. The Problem**  
- Measuring personal preventive behaviors (like mask usage and social distancing) on a large scale is challenging because traditional methods either invade privacy (tracking apps, direct observation) or rely on self-reports that people may censor if they fear judgment.  
- Available data sources during COVID (such as aggregated mobility reports) were too coarse to tie behavior changes to individual factors like vaccination status, making it hard to understand cause-and-effect (did getting vaccinated change one’s behavior?).  
- Public health officials needed detailed insights into behavior changes as vaccines rolled out, but any centralized data collection of such sensitive information risked low participation or public distrust. No large-scale study had attempted a federated (data-stays-on-device) approach, so it was unclear how to implement one effectively in the middle of a public health crisis.

**3. The Solution**  
- Launched a study through the Google Health Studies smartphone app, which invited participants to opt-in and answer questions about their masking and distancing habits over time. Importantly, the study used a **federated analytics** approach: individual response data stayed on each person’s phone. Instead of raw data being sent up, only aggregated model updates or summary statistics computed on-device were shared with researchers.  
- Employed **differential privacy** techniques to ensure that even the aggregated data transmitted could not be traced back to any individual. This added noise and other privacy safeguards to the collected statistics, meaning participants could contribute to findings without risking personal exposure.  
- Analyzed how participants’ behaviors changed after vaccination. The app quietly linked each person’s behavior data with their self-reported vaccination status *within* the device, so the central analysis only saw anonymous pattern information (e.g., “X% of fully vaccinated individuals reduced mask usage after two weeks” versus “Y% of unvaccinated individuals never wore masks”).  
- Proved out a new study design: rather than collecting potentially sensitive behavior logs, the research questions were answered by training a model collaboratively across devices or by combining privacy-protected summaries from each phone. This was one of the first practical demonstrations that such federated studies can work for epidemiological purposes.

**4. Architecture Overview**  
- **Federated Study Setup:** Participants’ phones acted as individual data nodes. The study coordinator defined analysis tasks (such as calculating average mask-wearing frequency by vaccination status) that were sent to each device. Each phone processed its owner’s data (survey answers, possibly phone sensor data if used) locally.  
- **Secure Aggregation:** Devices then sent back only encrypted, partial results. Using a secure aggregation protocol, the server could only decrypt the sum of all participants’ contributions, not any single participant’s data. For example, the system might compute “total number of days masked in a week across all participants” without ever seeing individual entries, then divide by population to get an average.  
- **Differential Privacy Layer:** Noise was injected into these aggregated results to obscure individual contributions further. This ensured that even if someone’s behavior was unique, it could not be picked out from the final data. The noise was calibrated so it didn’t drown out meaningful population-level signals but provided strong privacy guarantees.  
- **Data Validation:** The app and analysis included checks for data quality – e.g., ensuring sufficient participants contributed to an aggregate before trusting it (to avoid a scenario where one person’s data dominates a noisy result). Also, because no central raw dataset existed, the team developed novel validation methods, such as simulation testing: they would simulate devices with known behavior patterns to ensure the federated process could successfully recover those patterns in aggregate.  
- **Result Synthesis:** After multiple federated computation rounds, the collected insights were assembled into interpretable findings. The architecture allowed tracking trends over time: for instance, each week the study could federate a calculation like “proportion of each group always wearing a mask” and observe how those proportions shifted as vaccination coverage increased.

**5. Results and Impacts**  
- The study found clear behavior patterns: individuals generally continued practicing mask-wearing and social distancing until they completed their full vaccination course. Once fully vaccinated, some gradually relaxed these measures, but there wasn’t an immediate drop-off until they reached full immunity. In contrast, those who were vaccine-resistant (people who never planned to get a shot) had significantly lower adherence to mask and distancing guidelines from the start, even when COVID risk was high.  
- These insights provided evidence to public health experts about the interplay between vaccination and personal behavior. It suggested that vaccination campaigns did not lead to a precipitous abandonment of other precautions until after full vaccination – alleviating concerns that early vaccination might cause people to drop their guard too soon. It also highlighted that the group refusing vaccination was less likely to follow any guidelines, underlining a need for different public health messaging or interventions for that segment.  
- Importantly, the project proved the feasibility of conducting large-scale, privacy-preserving research in real-world conditions. It demonstrated that federated data collection can yield useful epidemiological insights without ever collecting raw personal data. This has opened the door for future studies on sensitive topics (mental health, sexual health, etc.) where privacy is paramount – researchers can study patterns without individuals having to surrender their private information.  
- The success of this approach during COVID-19 was documented in a publication in *Nature Digital Medicine* (2024), positioning the work as a milestone in digital public health methodology. It has since been referenced as a template for privacy-preserving analytics, influencing how tech companies and health researchers approach data ethics in population studies.

**6. Skills and Tools Used** (technical skills table)

| **Skill/Tool Category** | **Application in Project 08** |
| --- | --- |
| Federated Learning & Analytics | Implemented a federated analysis approach using Google’s privacy-preserving technology stack; orchestrated distributed data processing on user devices and aggregated results securely |
| Differential Privacy | Applied differential privacy techniques to ensure that aggregated data releases could not reveal any individual’s behavior, balancing noise addition with analytical usefulness |
| Mobile App Deployment | Collaborated on the Google Health Studies app deployment for the study, ensuring smooth user enrollment and data collection via a smartphone platform |
| Data Science Innovation | Developed novel validation and analysis strategies for a dataset that cannot be seen centrally; used simulation and advanced statistical reasoning to interpret federated results with confidence |
| Collaboration (Public-Private) | Worked closely with a major tech company’s research team (Google Health) and academic partners, coordinating cross-organization efforts in study design, IRB approvals, and result interpretation under a tight timeline |
| Privacy & Ethics in Research | Navigated complex privacy considerations, ensuring compliance with data protection standards and transparently communicating the privacy guarantees to participants to build trust in the study |

**7. Cross-Project Capabilities**  
- **Privacy-first data science:** This project expanded Dr. Tuli’s toolkit for analyzing sensitive data in a way that protects individual privacy. That expertise is transferrable to any future work dealing with confidential information (patient data, social data, etc.), allowing him to design studies that earn participant trust and meet strict privacy requirements.  
- **Methodological innovation:** Showcased ability to combine epidemiological questions with cutting-edge computing methods (federated learning, secure aggregation). This inventive mindset – solving public health problems with unconventional but effective techniques – is a hallmark of his approach across projects (e.g., using social media for surveillance, or generative AI for data simulation).  
- **Industry collaboration:** Reinforced experience in partnering with technology companies to tackle health challenges. As with prior projects involving Facebook and others, he effectively bridged the gap between tech capabilities and public health objectives, a valuable skill for executing large-scale, multi-stakeholder initiatives.

**8. Published Papers/Tools**  
- **Peer-Reviewed Publication:** *Nature Digital Medicine* (2024) – documented the federated study methodology and key behavioral findings, marking one of the first papers to show a real application of federated analytics in public health research.  
- **Framework for Future Studies:** While not a traditional “tool,” the protocols and lessons from this project have been shared as a blueprint for future privacy-preserving studies. The publication and follow-up materials serve as a how-to guide for health researchers interested in adopting federated approaches, including discussions of technical challenges and how they were overcome.  
- *(Note:* Due to the nature of federated studies, no raw dataset or app tool is publicly released, since data remained on user devices. Instead, the value lies in the documented approach and the validation that this new model works. The insights gained have been communicated to public health policy circles and are influencing the design of ongoing behavioral studies in other health domains.)