PROTOCOL Community Learning in Governed AI Health Navigation

2026-03-18

Dexter Hadley, MD/PhD

Hadley Lab CANONIC

hadleylab.org Governed Document. Every claim cited.

0.1. Study Information

Field	Value
Title	Community Learning Patterns in Governed AI Health Navigation
Short Title	CANONIC Community Learning Study
Sponsor	CANONIC Foundation
Principal Investigator	Dexter Hadley, MD/PhD
Co-Investigator	Marisa Nimrod, MD
Version	1.0
Date	2026-03-18

0.2. 1. Background and Rationale

Artificial intelligence systems deployed in health navigation generate conversational data that, when governed properly, constitutes a community learning resource. The CANONIC governance framework provides a structural approach to this problem: every conversation turn is recorded on an append-only ledger, every clinical claim traces to a governed evidence source, and every session is anonymized at the point of capture with identifiers that carry no personally identifiable information.

This protocol describes a retrospective observational study of community learning patterns generated by CANONIC-governed health navigation services. The study analyzes questions asked by patients, caregivers, and clinicians as they interact with governed AI navigation platforms, characterizing the community intelligence that accumulates when conversational data is structurally anonymized and ledgered from inception.

The study is designed to be modular. Each governed TALK scope (a deployed health navigation instance) constitutes a study arm with its own population, evidence layers, and geographic context. The governance architecture is identical across all arms because every TALK scope inherits the same structural constraints from the CANONIC framework.

0.2.1 1.1 Regulatory Context

A 2022 Lancet Global Health systemic assessment of research ethics systems in Latin America and the Caribbean found that among CARPHA member states, only Guyana has adopted comprehensive legislation governing human subjects research. Jamaica and Trinidad and Tobago have institutional ethics committees but no national oversight body, and their research ethics policies are not legally binding. The CANONIC governance framework, which enforces anonymization, append-only immutability, and evidence sourcing as structural constraints rather than policy guidelines, provides protections that exceed the legislative baseline in most jurisdictions where the system operates.

0.3. 2. Study Design

Type: Retrospective observational study of anonymized community learning ledger data.

Design: Multi-arm, with each governed TALK scope constituting an independent arm. Arms share identical governance architecture but differ in population, geography, evidence layers, and clinical domain.

0.4. 3. Study Arms

0.4.1 Arm A: CaribChat (Caribbean Cancer Navigation)

Field	Value
Scope	TALKS/CARIBCHAT
Domain	caribchat.ai / carib.chat
Population	Caribbean cancer patients, caregivers, clinicians

Field	Value
Geography	Trinidad & Tobago, Jamaica, Barbados, Bahamas, Guyana, Saint Lucia, Dominica, Antigua & Barbuda, Suriname
Evidence Layers	CAOH, CARPHA, TT Cancer Society, NCCN Resource Stratification, PAHO/WHO, ClinicalTrials.gov, IAEA, Caribbean healing traditions, screening infrastructure registry, WHO CHW framework
Sessions Ledgered	55+ (as of 2026-03-18)
Active Since	2026-03-02
Launch Event	CAOH 2026 Annual Scientific Conference, July 17-19, Hilton Trinidad

0.4.2 Arm B: MammoChat (US Breast Health Navigation)

Field	Value
Scope	TALKS/MAMMOCHAT
Domain	mammochat.ai
Population	US breast cancer patients, caregivers, clinicians
Geography	United States (Florida primary)
Evidence Layers	NCCN guidelines, ClinicalTrials.gov, SEER, mCODE, institutional screening databases
Sessions Ledgered	20+ (as of 2026-03-18)
Active Since	2026-02-27
Clinical Trial	NCT06604078 (Casey DeSantis Florida Cancer Innovation Award, $2M)

0.4.3 Future Arms

Additional TALK scopes may be added as amendments when their community learning ledgers reach sufficient volume. Each arm inherits the same governance constraints and data architecture described in this protocol.

0.5. 4. Data Description

0.5.1 4.1 Data Source

The community learning ledger for each TALK scope. Ledger entries are generated automatically when a user submits a question to a governed TALK instance. The ledger is append-only; entries cannot be modified or deleted after creation.

0.5.2 4.2 Data Elements

Each ledger entry contains exactly three fields:

Field	Description	Example
Date	Date of the session	2026-03-15
Question	The text of the users question	Where can I get screened in Port of Spain?
Trace ID	Session identifier	Random, non-linkable

0.5.3 4.3 Data NOT Collected

The ledger does not contain and has never contained:

- Names or usernames
- Email addresses or phone numbers
- IP addresses or device identifiers
- Demographic information (age, sex, race, ethnicity)
- Geographic location of the user (only geographic content of questions)
- Clinical data, diagnoses, or treatment information
- Session responses (captured separately, not analyzed)

0.5.4 4.4 Governance Architecture

All data is governed by the CANONIC framework at the structural level:

Constraint	Mechanism
Anonymization	Identifiers assigned at session creation; no PII fields exist in the schema
Immutability	Append-only ledger; entries cannot be modified or deleted
Integrity	Cryptographic hashing of all evidence
Auditability	Full governance audit at continuous compliance score
Evidence sourcing	Every clinical claim in system responses traces to governed sources
Structural enforcement	Constraints are architectural, not policy-based

0.6. 5. Study Objectives

0.6.1 5.1 Primary

Characterize community learning patterns in governed AI health navigation by analyzing the questions patients, caregivers, and clinicians ask when interacting with structurally anonymized, ledger-based navigation systems.

0.6.2 5.2 Secondary

1. Compare community learning patterns across geographic and clinical contexts (Caribbean cancer navigation vs. US breast health navigation).
2. Evaluate whether governance-native data architecture provides adequate human subjects protections in jurisdictions without research ethics legislation.
3. Characterize the compounding community intelligence effect: how accumulated questions improve navigation quality over time.

0.7. 6. Analysis Plan

0.7.1 6.1 Descriptive Analysis

- Question volume by arm, by week, by geographic reference
- Question taxonomy: screening, treatment, navigation, tradition, clinical trial, community care
- Temporal patterns: question frequency trends, topic evolution over time

0.7.2 6.2 Comparative Analysis

- Cross-arm comparison: CaribChat vs. MammoChat question distributions
- Geographic specificity: proportion of questions referencing specific facilities, countries, or regions
- Cultural context: healing tradition queries (CaribChat) vs. clinical pathway queries (MammoChat)

0.7.3 6.3 Spanish Language Subgroup Analysis

A pre-specified subgroup analysis across both arms will characterize community learning patterns among Spanish-language users. Trinidad and Tobago has experienced significant Venezuelan migration, creating a growing Spanish-speaking population that interacts with CaribChat for cancer navigation. La Florida is historically Spanish Caribbean: settled by Ponce de León in 1513 and governed by Spain for three centuries before cession to the United States, its contemporary Hispanic population maintains deep continuity with the Spanish-speaking Caribbean basin. The cross-arm comparison (Trinidad vs. Florida) tests whether Spanish-language community learning patterns differ by geography and clinical context when the underlying governance architecture is identical, and when both populations share a common linguistic heritage rooted in the Spanish Caribbean.

- **Language detection:** Automated classification of ledger questions by language (Spanish vs. English vs. other) using the question text field
- **Cross-arm comparison:** Spanish-language question volume, taxonomy distribution, and temporal trends compared between Trinidad (Arm A, CaribChat) and Florida (Arm B, MammoChat)
- **Topic characterization:** Whether Spanish-language questions in Trinidad differ in clinical domain emphasis (screening access, treatment navigation, traditional medicine, clinical trial inquiry) compared to Spanish-language questions in Florida
- **Navigation adequacy:** Assessment of whether governed evidence layers provide equivalent

coverage for Spanish-language queries across both arms, measured by evidence source match rates

- **Cultural navigation:** Comparison of how Spanish-speaking users in Trinidad (Venezuelan migrants navigating an unfamiliar health system) differ from Spanish-speaking users in Florida (established Hispanic community with existing health system access) in the types of navigation support they seek

0.7.4 6.4 Underserved Language Discovery

Beyond Spanish, the Caribbean basin encompasses a rich linguistic ecology that conventional health systems largely ignore. The community learning ledger captures questions in whatever language the user chooses to write, providing a natural discovery mechanism for underserved language communities that would be invisible in systems that default to English-only interfaces.

- **Creole and Patois detection:** Identification of questions written in Trinidadian Creole, Jamaican Patois, Haitian Kreyòl, Guyanese Creole, and other Caribbean contact languages that reflect the lived linguistic reality of the populations served
- **Volume and distribution:** Characterization of non-English, non-Spanish question volume by arm and by geographic reference, to quantify the scale of underserved language navigation need
- **Code-switching patterns:** Analysis of questions that mix English with Creole, Patois, or Spanish within a single query, reflecting how multilingual users naturally navigate health information
- **Evidence gap identification:** Assessment of whether governed evidence layers provide adequate coverage for queries in underserved languages, identifying where translation or language-specific evidence sourcing is needed
- **Community learning effect:** Whether questions in underserved languages cluster around specific clinical domains (traditional medicine, facility navigation, screening access) that reveal unmet needs invisible in English-language data

This analysis is discovery-oriented: the ledger reveals which languages communities actually use when seeking health navigation, rather than which languages the system was designed to support. Findings will inform future evidence layer development and multilingual governance extensions.

This subgroup and discovery analysis requires no additional data collection because language is an intrinsic property of the question text already captured on the ledger.

0.7.5 6.5 Governance Analysis

- Data protection adequacy: comparison of CANONIC structural constraints vs. legislative requirements across jurisdictions
- Community learning compounding: measurement of question overlap and knowledge reuse over time

---

0.8. 7. Risk Assessment

Risk	Assessment	Mitigation
Re-identification	Minimal: no PII in schema, questions are free-text without identifiers	No linkage table exists
Sensitive content	Low: questions may reference personal health concerns in free text	Analysis at aggregate level; no individual question attribution in publication
Cultural harm	Low: Caribbean healing traditions discussed in evidence context	Evidence-tagged with clinical status; traditions respected, never dismissed
Geographic re-identification	Minimal: facility names are public information already in the system	Questions reference public facilities; no private location data

Overall risk classification: MINIMAL

0.9. 8. Consent and Waiver Justification

0.9.1 8.1 Waiver of Informed Consent

This study requests a waiver of informed consent under 45 CFR 46.116(f) based on the following criteria:

1. **Minimal risk:** The study involves no intervention and analyzes only anonymized, aggregate question patterns. No PII exists in the data.
2. **Impracticability:** Users interact with public-facing navigation services; requiring prospective consent for retrospective analysis of anonymized questions would be impracticable and would fundamentally alter the community learning model.
3. **Rights and welfare:** The waiver does not adversely affect rights or welfare. Data was anonymized at the point of capture by architectural design, not by post-hoc de-identification.
4. **Additional information:** Subjects will not be contacted. No additional information will be provided because subjects cannot be identified.

0.9.2 8.2 Terms of Service

All TALK instances display terms of service noting that anonymized questions contribute to community learning. The community learning model is a core, disclosed feature of the service, not a secondary use of clinical data.

0.10. 9. Data Management

Aspect	Detail
Storage	CANONIC governed repository (append-only, version-controlled)
Retention	Permanent (community learning ledger is the institutional memory)
Access	PI and Co-I; governed by CANONIC access controls
Sharing	Aggregate results published; raw ledger entries not shared

0.11. 10. Dissemination

Primary manuscript targeting peer-reviewed publication. Presentation at CAOH 2026 Annual Scientific Conference (July 17-19, Port of Spain, Trinidad and Tobago). Additional presentations at relevant informatics and digital health venues.

IRBS | CARIBCHAT | PROTOCOL | 2026-03-18