ADVOCATES FOR ADVANCING HEALTH

METHODS APPENDIX: AI Framework for National, Near Real-Time Detection of Clinical Practice Gaps

INPUT: DATA

1 NATIONAL-SCALE AND DEPTH

AI autonomously captures and integrates different data types, structures, formats, and sources into knowledge domains

National scale addresses traditional small sample limitations

DATA SOURCES capture **85%**[1] of daily digital activity

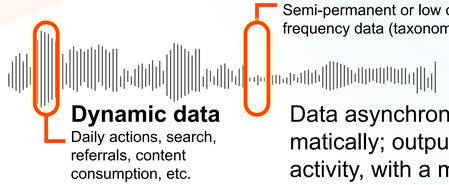
- **VOLUME**
  - Web traffic
  - Links and redirects
  - Geospatial activity *Trillions of location points*
- **VOCABULARY**
  - Entire search domain *Billions of searches*
- **VENUES**
  - Websites
  - 1st party apps *~4 billion sites processed monthly*
- **ENGAGEMENT**
  - Content
    - *Drug information, interaction databases*
    - *PV databases, e.g., SIDER-2*
    - *Medical taxonomies, e.g., NIH-MESH*
    - *Publications: >1 million studies; >5000 publishers*

Social media, paid campaigns, offline activities

2 NEAR REAL-TIME

Digital patterns are ingested daily and re-indexed to reconstruct 90-day patterns on a rolling basis

Rapid access to emerging signals facilitates responsive decision-making



Stable data
Semi-permanent or low change frequency data (taxonomies, etc.)

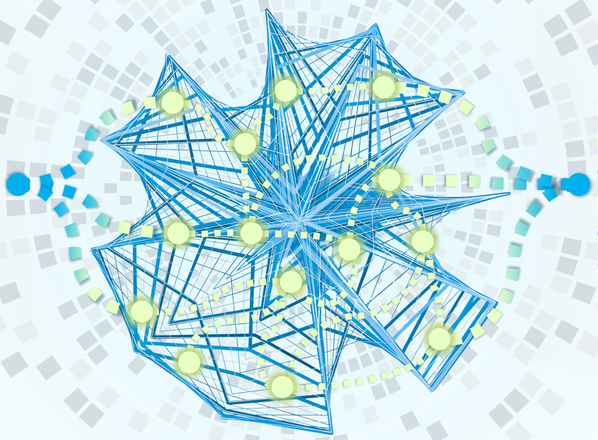
Dynamic data
Daily actions, search, referrals, content consumption, etc.

Data asynchronicities are managed automatically; output represents ~90 days of activity, with a minimum of 30 days history

3 DATA PRIVACY AND SECURITY

GDPR · HIPAA COMPLIANT · COPPA COMPLIANT · ISO 27001

AI ALGORITHM META-SIGNAL



AI systematically extracts deterministic digital activity to assemble pattern chains or **SIGNALS** representing a microcosm of a given construct

Linked signals form a **META-SIGNAL** → yielding ontologies, behavioral cohorts and market segments → making inter-dependencies within complex ecosystems visible

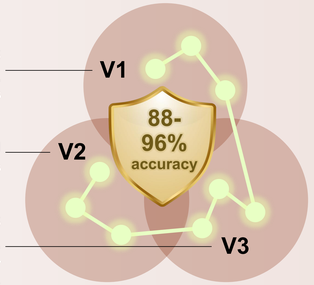
OUTPUT: KNOWLEDGE

1 HCP IDENTIFICATION

HCPs authenticated probabilistically given shift away from deterministic IDs (cookies). Only HCPs active across all 3 independent vectors were identified as target cohort; independent 3rd party audits confirmed 88-96% model accuracy[2,3] despite residual risks associated with probabilistic approaches

Validated, accurate cohort identification supports robust specialty and stage-level analyses

Probabilistic cohort targeting 3-vector (V) model

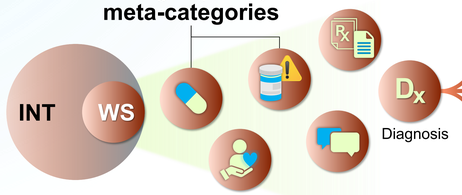


2 WHITESPACE DETECTION

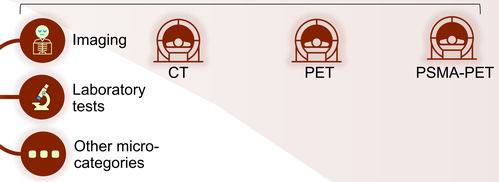
Digital patterns were categorized as INT (interaction events); unmet information needs within INT were flagged as Whitespaces (WS)

Stage, specialty and MDT-specific WS identified, characterized and prioritized at scale, providing clarity on where, what, how to intervene

WS grouped by **meta-categories**



WS meta-categories can be drilled down for deeper analysis through **4 micro-categories**



METHOD LIMITATIONS

WS ≠ CAUSALITY
WS represents unmet informational demand, **not causal Tx effects or clinical outcomes**

TEMPORALITY
Short data processing window (~90 days) may **bias toward seasonal or campaign activity** over sustained demand

RESIDUAL RISKS
Residual blind spots from untracked digital activity (~15%) and probabilistic 3-vector model may affect precision

OPEN WEB BIAS
Captures **organic demand only**; excludes closed networks, social and paid media, and offline behaviors

1. "85% of digital activity" refers to the share of public, anonymized open web activity from sources listed above (search queries, redirect chains, traffic, page loads and related telemetry from ~4 billion websites/1st party apps); it excludes closed-platform content and paid/social media.
2. Model validation showed 88-96% accuracy across healthcare and consumer applications; in healthcare, it achieved 90% accuracy when deployed for contact tracing in South Africa https://www.straitstimes.com/tech/local-health-bodies-and-tech-firms-using-ai-to-fight-covid-19
3. We validated HCPs authenticated by the 3-Vector Model against national medical society registers to confirm accuracy of HCP-cohort universe size.