Because Healthcare Research Deserves AI Support

OHDSI AutoCohort

by Iurii (Yurri) Iurchenko
Key Open Source Contributor

Mission Statement:

Mission aligned with U.S. priorities public health, transparency, data privacy

To empower public health research by simplifying and accelerating cohort analytics on OMOP data. Through intelligent automation, our application enables researchers and agencies, such as the CDC and NIH, as well as independent researchers, to quickly identify critical trends and insights, thereby protecting public health.

Utility Rating:

Why OMOP CDM is critical for the US Healthcare

ATLAS is an open-source web tool from OHDSI for designing and running analyses on standardized healthcare data, including cohort creation, prediction models, and treatment pathways

Nearly 1 billion

unique anonymized patient records, spanning 49 countries including the US, have been standardized using OMOP models, making it one of the largest real-world data networks for health research

More than 300,000 patient records

at major US health systems () have been converted to OMOP format, regularly updated, and made available for research on various diseases, including cancer and COVID-19

Over 340 published studies

the OHDSI community has produced utilizing OMOP standards to answer questions on adverse events, risk assessment, comparative effectiveness, patient-level predictions, and quality of care

Iurii (Yurri) Iurchenko, Open Source Contributor

Vision Statement:

High-level goal of the app

Our application extends OHDSI's Atlas framework by using agentic AI to streamline the design, validation, and execution of observational research on OMOP data. It transforms complex, repetitive workflows into intuitive, AI-assisted tasks, enabling researchers to focus on insights rather than manual operations.

1. Automate Routine Research Workflows

   Eliminate repetitive tasks, such as building concept sets, running diagnostics, or recalculating incidence rates

2. Accelerate Study Design and Execution

   Enable researchers to transition from study ideas to actionable cohort definitions more quickly with AI-generated templates and recommendations

3. Enhance Accuracy and Consistency

   Utilize agentic AI to validate definitions, identify errors, and propose enhancements to cohort logic

4. Democratize OMOP Research

   Lower the barrier for new researchers or organizations by offering natural language interaction and pre-built intelligent workflows

5. Bridge the Gap Between Data and Insights

   Allow researchers to focus on scientific questions while AI handles technical and repetitive processes in ATLAS

Iurii (Yurri) Iurchenko, Open Source Contributor

Problem Overview:

Data-backed explanation
of current research/data challenges in healthcare

- Identifying key markers and influential factors during the research and predictions may consume significant time
- Rapidly create and validate cohorts across diverse data sources, improving reproducibility and scalability
- Leverage AI for intelligent concept set suggestions and quality checks, reducing bias and inconsistencies in observational studies
- Unlock deeper insights from large-scale health datasets (e.g., EHRs, claims data) by enabling near real-time analysis, which supports faster policy decisions, drug safety monitoring, and population health interventions.

Article from NIH

https://pmc.ncbi.nlm.nih.gov/articles/PMC11229257/

Discussion from OHDSI community

https://forums.ohdsi.org/t/seeking-tool-for-simplified-data-extraction-and-transformation-from-omop-cdm/21051/3

Total observational time from OHDSI article

https://www.ohdsi.org/wp-content/uploads/2023/10/104-Blacketer-BriefReport.pdf

Iurii (Yurri) Iurchenko, Open Source Contributor

Iurii (Yurri) Iurchenko, Open Source Contributor

Challenge #1

Clinical Accuracy of AI-Generated Concept Sets

Risk: AI may suggest irrelevant or incomplete codes (e.g., wrong SNOMED or RxNorm concepts), leading to incorrect cohorts

Mitigation:

- Eliminate repetitive tasks, such as building concept sets, running diagnostics, or recalculating incidence rates
- Use reference vocabularies and validation checks (e.g., OMOP standardized vocabulary hierarchy) to verify completeness
- Provide confidence scores and clear explanations for AI suggestions

Challenge #2

Lack of Explainability in AI Decisions

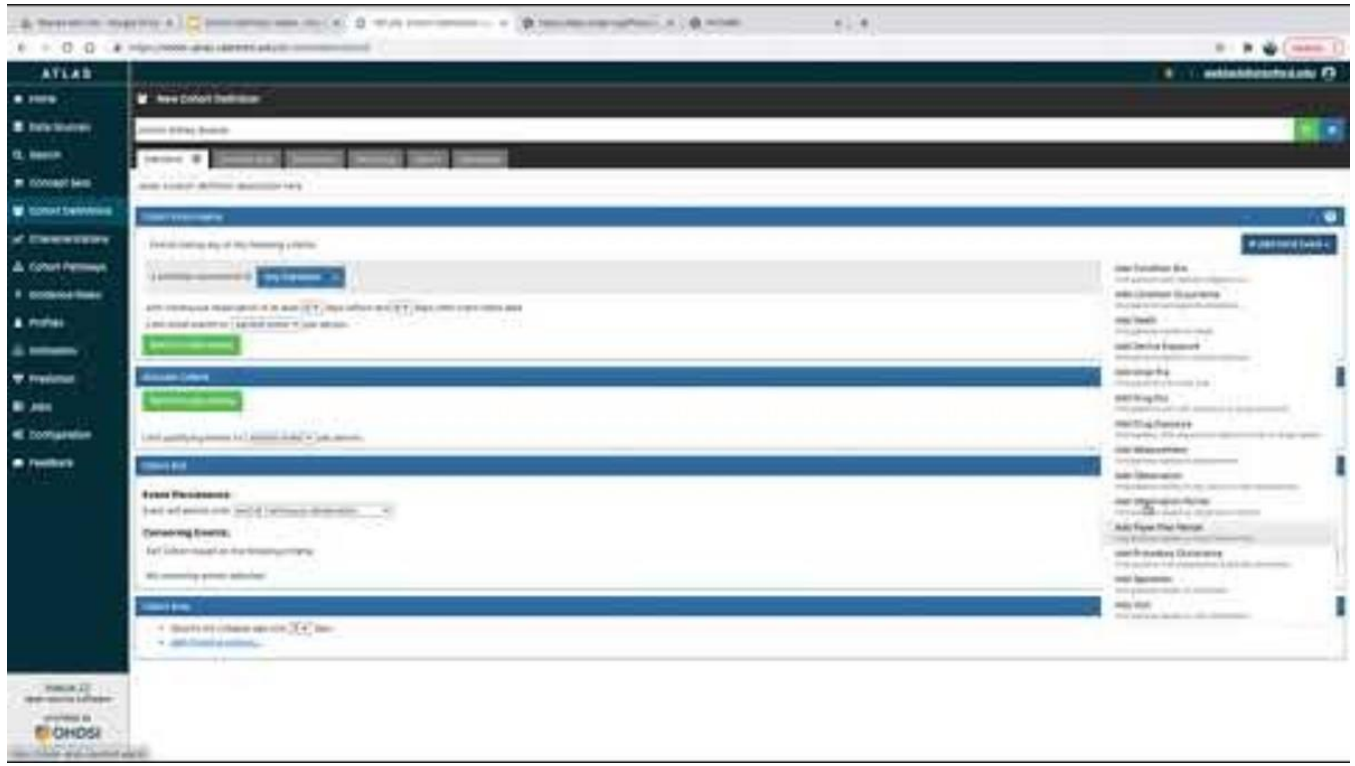
Risk: AI may produce cohort logic or query filters that are not fully transparent, making it hard for researchers to trust or reproduce results

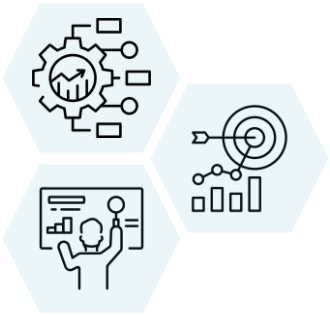
Mitigation:

- Ensure every AI action is logged and can be replayed (audit trail)
- Include plain-language explanations of generated criteria (e.g., “These concepts were added because they are descendants of Condition X in SNOMED”)
- Allow researchers to manually edit or override AI logic

Iurii (Yurri) Iurchenko, Open Source Contributor

Product Overview





Product Overview:

What the App Does

The application extends the OHDSI/OMOP ecosystem with agentic AI to automate repetitive research tasks, such as cohort definition, concept set building, diagnostics, and real-world evidence (RWE) analysis. It transforms natural language instructions into actionable research workflows, reducing the time required to design, validate, and run observational studies.

Target Users

Healthcare researchers (e.g., epidemiologists, data scientists, and statisticians) using the OHDSI framework

Public health organizations (e.g., FDA, CDC, NIH) are monitoring population health and safety signals

Healthcare analytics teams and real-world evidence groups in payers and providers

Value Proposition

Accelerates research: Automates repetitive Atlas/OMOP tasks, reducing study design and analysis time

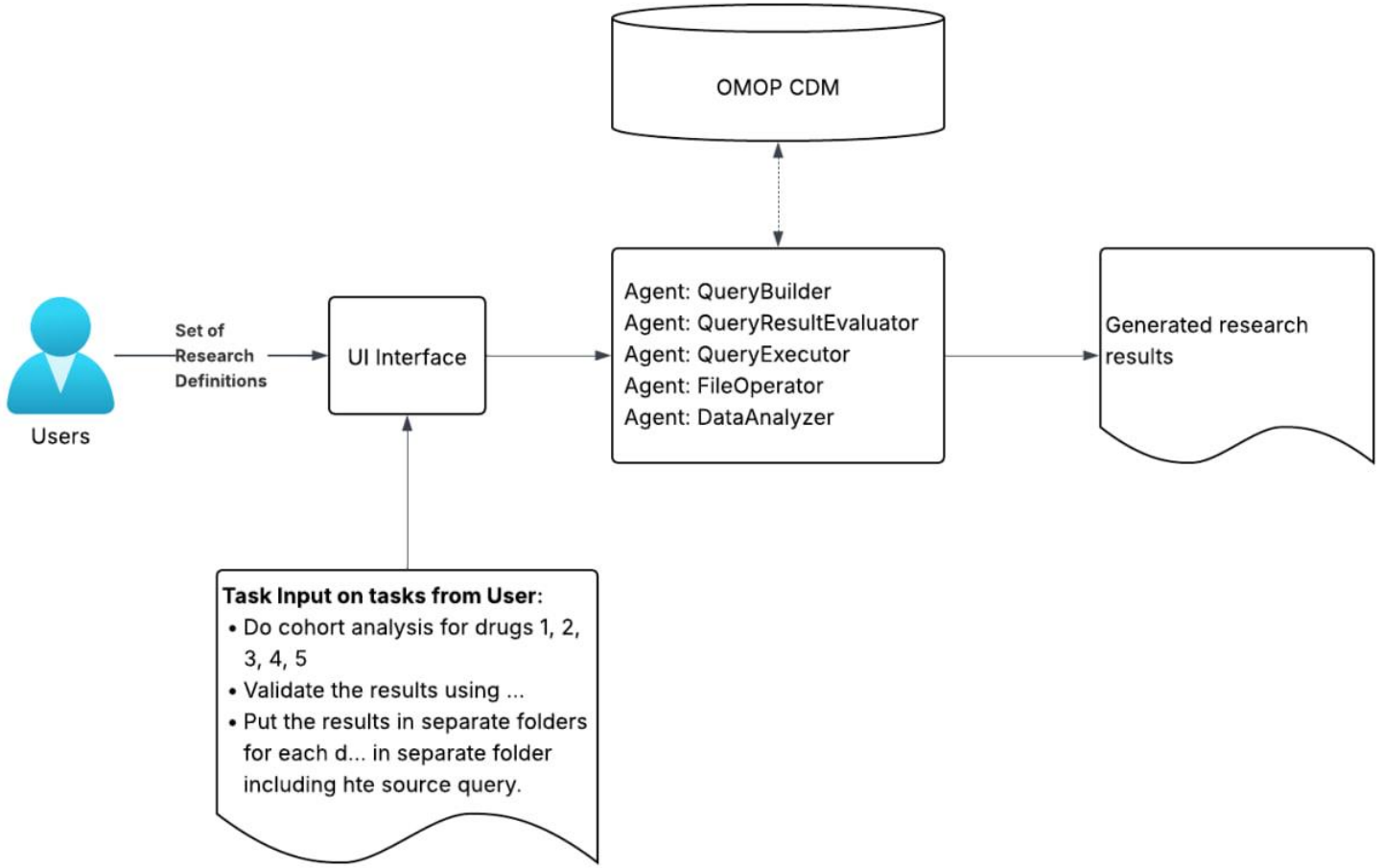
Enhances accuracy & accessibility: AI-driven suggestions with transparent, human-in-the-loop workflows for all skill levels

Open-source innovation: Remains free and community-driven, aligning with OHDSI's collaborative ecosystem

Iurii (Yurri) Iurchenko, Open Source Contributor

Key Features & Functionalities:

Solutions Architecture



Key Feature #1

AI-Powered
Cohort Definition Builder

AI-powered search and suggestion for standardized vocabularies (SNOMED, RxNorm, LOINC)

Auto-inclusion of related or descendant concepts based on ontology graphs

Deduplication and validation of concept sets

Natural language input: “Find all drugs related to Type 2 Diabetes treatment.”

What it does:

- Automatically generates inclusion/exclusion criteria by interpreting research questions (e.g., “Find patients with Type 2 Diabetes diagnosed after 2020”)
- Suggests relevant concept sets and standard vocabularies (SNOMED, ICD) with AI-driven search and validation

Why critical:

- Reduces manual effort in defining cohorts
- Ensures standardized, reproducible definitions

Iurii (Yurri) Iurchenko, Open Source Contributor

Key Feature #2

Integrated Cohort Validation & Diagnostics

AI-driven cohort generation from plain-text study descriptions

Automated inclusion/exclusion criteria recommendations with count preview optimization

Variant generation: Creates multiple cohort definitions for sensitivity analyses

Automatic SQL generation optimized for OMOP CDM

What it does:

- Automatically runs CohortDiagnostics after cohort creation
- Automatically runs CohortDiagnostics after cohort creation
- Generates visual reports and summaries for quick insights.

Why critical:

- Ensures quality, validity, and reproducibility of research findings
- Helps researchers spot data gaps or misaligned criteria quickly

Iurii (Yurri) Iurchenko, Open Source Contributor

Key Feature #3

Outcome Analysis Module

Automated CohortDiagnostics execution to validate definitions

Highlights data gaps, anomalies, or inconsistencies (e.g., missing concepts)

Generates human-readable diagnostic summaries

What it does:

- Calculates incidence, prevalence, and survival metrics for selected cohorts
- Visualizes results (charts, tables) and exports reports

Why critical:

- Turns cohorts into actionable insights
- Automates complex analyses, saving time
- Essential for evidence-based healthcare decisions

Iurii (Yurri) Iurchenko, Open Source Contributor

User Journey Steps

#1. Login & Data Setup

User Action: Logs into the AI-powered app and selects/links OMOP dataset (via Atlas API)

System Action: Verifies data connection and loads available study templates

#2. Define Research Goal

User Action: Describes the research task in natural language (e.g., “Find Type 2 Diabetes patients treated with Metformin.”)

System Action: AI interprets the goal and proposes a workflow plan (Concept Sets → Cohort Definition → Analysis)

#3. Concept Set Generation

User Action: Reviews AI-suggested concept sets (conditions, drugs, procedures)

System Action: Automatically retrieves standardized vocabularies (SNOMED, RxNorm) via Atlas API, highlights missing/related concepts

#4. Cohort Definition & Refinement

User Action: Adjusts inclusion/exclusion criteria with AI’s guidance (e.g., preview counts)

System Action: AI optimizes criteria and generates OMOP-compatible SQL queries

Iurii (Yurri) Iurchenko, Open Source Contributor

Iurii (Yurri) Iurchenko, Open Source Contributor

User Journey

User Journey Steps

#5. Analysis & Diagnostics

User Action: Selects desired analysis (e.g., incidence rates, pathway analysis)

System Action: Executes CohortDiagnostics, characterization, and pathway analysis via Atlas API

#6. Results Visualization

User Action: Reviews interactive dashboards and key findings

System Action: AI summarizes insights (e.g., age distribution, drug patterns)

#7. Export & Reporting

User Action: Exports cohort definitions, reports, and visual outputs

System Action: Generates study documentation (Markdown/Word) with clear descriptions of methods and results

#8. Iterative Refinement

User Action: Asks AI follow-up queries or defines new cohorts (e.g., “Compare to a control group without Metformin”)

System Action: Clones and modifies previous definitions for new analyses

Roadmap & Development Timeline:

12-18 month phased development

Phase #1

Foundation & MVP
0–6 months

Set up core architecture and Atlas API integration

Build MVP with Concept Set Builder and basic cohort creation

Launch pilot version with early feedback

Phase #2

Feature Expansion 6–12 months

Add cohort diagnostics, characterization, and pathway analysis

Introduce AI-driven SQL optimization and natural language query support

Strengthen UI/UX and open-source community presence

Phase #3

AI & Automation
12–18 months

Enable agentic workflows (multi-step automation)

Launch chat-based research assistant for real-time interactions

Add auto-reporting and export tools

Market Gap / Competitive Landscape:

How my tool fills an existing gap

Existing OHDSI tools like Atlas are powerful but require manual, repetitive, and time-intensive workflows (e.g., building concept sets, defining cohorts, running diagnostics). Researchers often spend hours to days performing tasks that could be automated. Moreover, Atlas lacks AI-driven assistance, natural language interfaces, and workflow automation to lower the entry barrier for non-technical researchers.

Feature / Capability	OHDSI Atlas	Commercial Platforms (e.g., TriNetX, IBM Watson Health)	AgentOMOP
Open-Source	Yes	No	Yes
AI Assistance	No	Free	Advanced Agentic AI
Natural Language Interface	No	Limited (mostly rule-based)	Yes
Concept Set Automation	Manual	Rare / Limited	Fully Automated + AI
Cohort Design Optimization	Manual	Partially Automated	AI-driven Optimization
Cohort Diagnostics	Yes (manual)	Yes	Auto-run + Summaries
Workflow Orchestration	No	Limited	End-to-End Agentic Flow
Customization & Extensibility	Moderate	Low (proprietary)	High (Community-Driven)
Cost	Free	High (License Fees)	Free & Open-Source

Iurii (Yurri) Iurchenko, Open Source Contributor

Iurii (Yurri) Iurchenko, Open Source Contributor

Impact Metrics / Success Indicators:

Projected impact

on healthcare research, policy, trust restoration

Research Efficiency Metrics

Cohort Design Time Reduction:
Measure: Average time (in hours/days) to create and validate a cohort before vs. after using the app

Concept Set Automation Rate:
Measure: % of concept sets created or improved by AI with minimal manual input

Workflow Automation:
Measure: Number of tasks completed end-to-end by AI agents (e.g., cohort creation, diagnostics) without manual intervention

User Task Reduction:
Measure: Average reduction in manual clicks or steps compared to standard Atlas workflows.

Iurii (Yurri) Iurchenko, Open Source Contributor

Impact Metrics / Success Indicators:

Projected impact

on healthcare research, policy, trust restoration

Adoption & Usage Metrics

Active Users:
Measure: Number of researchers/teams regularly using the app per month

Open-Source Contributions:
Measure: Number of GitHub contributors, forks, and stars

Institutional Adoption:
Measure: Number of research institutions, pharma companies, or universities adopting the app

Repeat Usage:
Measure: Average number of sessions per user per week or month.

Iurii (Yurri) Iurchenko, Open Source Contributor

Impact Metrics / Success Indicators:

Projected impact

on healthcare research, policy, trust restoration

Analysis & Output Metrics

Studies Completed:
Measure: Number of completed cohorts, diagnostics, or pathways per user/team per month

Query Optimization:
Measure: % improvement in query execution time compared to manually created SQL in Atlas

Report Generation:
Measure: Number of automated study reports generated by AI (vs. manually written)

About me

Open Source Contributor

Iurii (Yurri) Iurchenko

12 years of experience in Data and Big Data Engineering

Master of Science in Chemistry from the top-ranked university in Ukraine; synthesized new sugar-lowering compounds

3 years of experience in Healthcare Data Engineering

Implementing AI solutions since 2019

What's next?

Conclusion / Next Steps

App Value Summary:

- AI-driven automation of cohort building and analysis in OHDSI ATLAS
- Accelerates research: Cuts manual steps and enables near real-time insights
- Improves quality: Standardized, reproducible cohort definitions and diagnostics

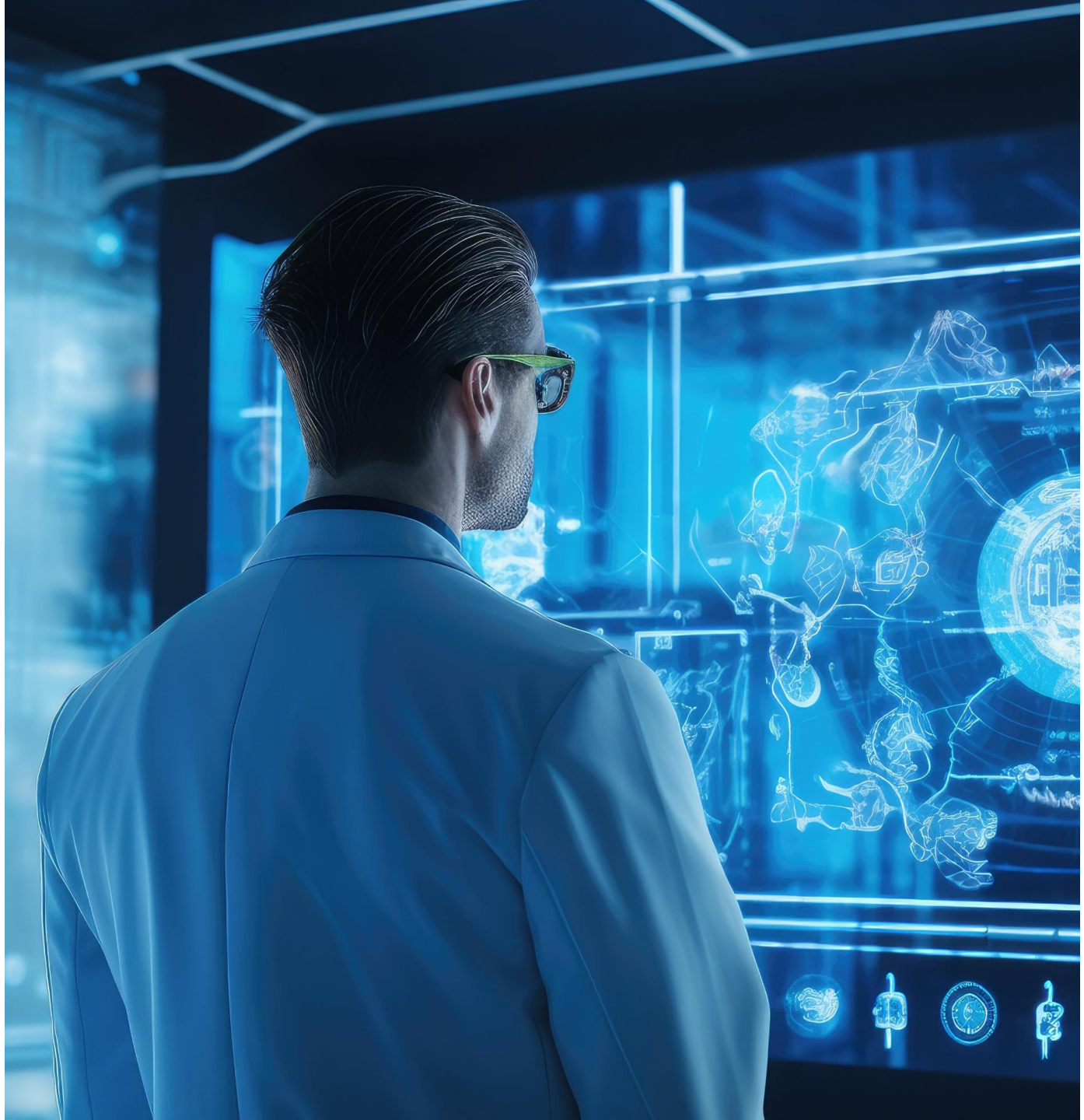
Execution Plan:

- MVP Build (3–6 months): Cohort builder, outcome analysis, diagnostics automation
- Pilot with Healthcare Partners (6–9 months): Validate with real-world data
- Scale & Integrate (9–12 months): Multi-database support, AI optimizations

What's next?

National Benefit

- Supports faster, data-driven healthcare decisions in the U.S.
- Improves drug safety monitoring and population health research
- Boosts research productivity with minimal overhead



Iurii (Yurri) Iurchenko, Open Source Contributor