

# Leveraging Generative AI and Machine Learning for Robust Real-World Cohort Definition and Patient Journey Insights in HIV Research

Chuo CY, PhD<sup>1</sup>; Jin S, PhD<sup>1</sup>; Guillot J, PhD<sup>1</sup>; Rivera JA, MS<sup>2</sup>; Fathollahi M, PhD<sup>3</sup>; Liu L, PhD<sup>1</sup>; Marongiu A, PhD<sup>1</sup>; Lim T, PhD<sup>1</sup>; Arora R, BT<sup>1</sup>; Chong B, BS<sup>1</sup>; Shahraz S, PhD<sup>1</sup>; Christoph MJJ, PhD<sup>1</sup>; Leung G, PhD<sup>1</sup>; Tao L, PhD<sup>1</sup>; Wang W, PhD<sup>1</sup>; Asimwe A, PhD<sup>1</sup>  
<sup>1</sup>Gilead Sciences, Foster City, CA, USA · <sup>2</sup>University of California San Francisco, San Francisco, CA, USA · <sup>3</sup>Amazon Web Services, Palo Alto, CA, USA

ISPOR 2026  
Methodological & Statistical Research

MSR104

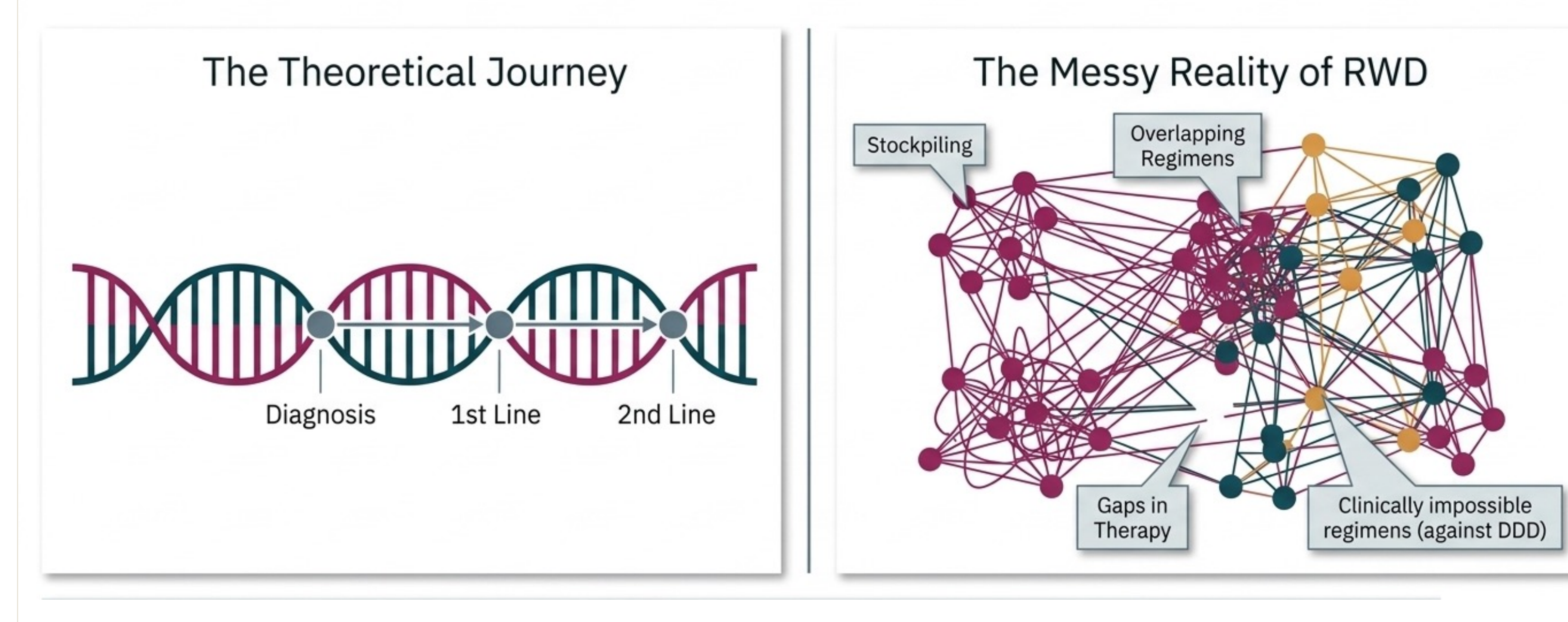
## Plain Language Summary

Using AI, researchers analyzed the treatment records of over 22,000 people with HIV to automatically identify patient groups based on their real-world treatment histories — without requiring manual literature reviews or pre-defined criteria (saving weeks of manual work). This AI-powered tool can help clinicians and researchers better understand how HIV patients move through different treatments in the real world, potentially improving clinical-trial design and identifying patients who may need additional care support.

## Introduction

- Researchers conducting real-world evidence (RWE) studies typically rely on literature and clinical guidelines to define patient cohorts — an approach that is time-consuming and can oversimplify patient journeys.<sup>1</sup>
- Traditional methods constrain definitions toward guideline-driven assumptions, missing real-world treatment patterns and limiting insights in RWE studies.<sup>2</sup>
- Emulating clinical trials with real-world data presents challenges including limited coverage, small samples, and time-intensive criteria curation.<sup>3</sup>

### The Challenge of Real-World Data



**Figure 1.** Guideline-defined journeys are linear; real-world ART journeys involve overlap, stockpiling, adherence gaps, and unplanned switches.

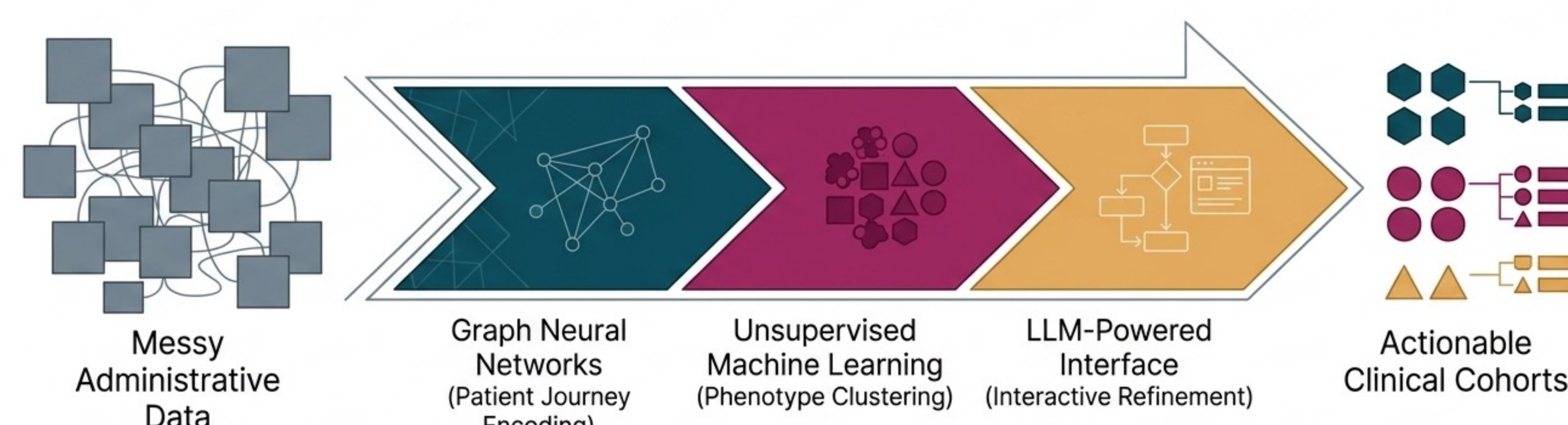
## Objective

Develop an interactive AI-driven framework to streamline cohort definition, uncover antiretroviral treatment (ART) patterns, and enable scalable trial emulation in people with HIV (PWH) — leveraging graph neural networks, clustering algorithms, and a large language model (LLM)-powered interface.

## Methods

- **Data source.** HealthVerity claims and lab data (2022–2025); 22,431 adults with HIV receiving ART with ≥18 months post-ART follow-up.
- **Patient-journey encoding.** Graphs were built representing ART treatments as time-linked nodes; a graph encoder modeled each patient's ART journey using masked treatment prediction as a pretext task.
- **Clustering.** Agglomerative clustering on curated lines-of-therapy features (regimen complexity, gaps, duration).
- **Validation.** Outputs were reviewed by subject-matter experts for interpretability and alignment with epidemiologic standards.
- **LLM interface.** Findings were packaged as a Model Context Protocol (MCP) server and integrated into a Claude-based chat interface for interactive cohort definition.

### The AI Framework

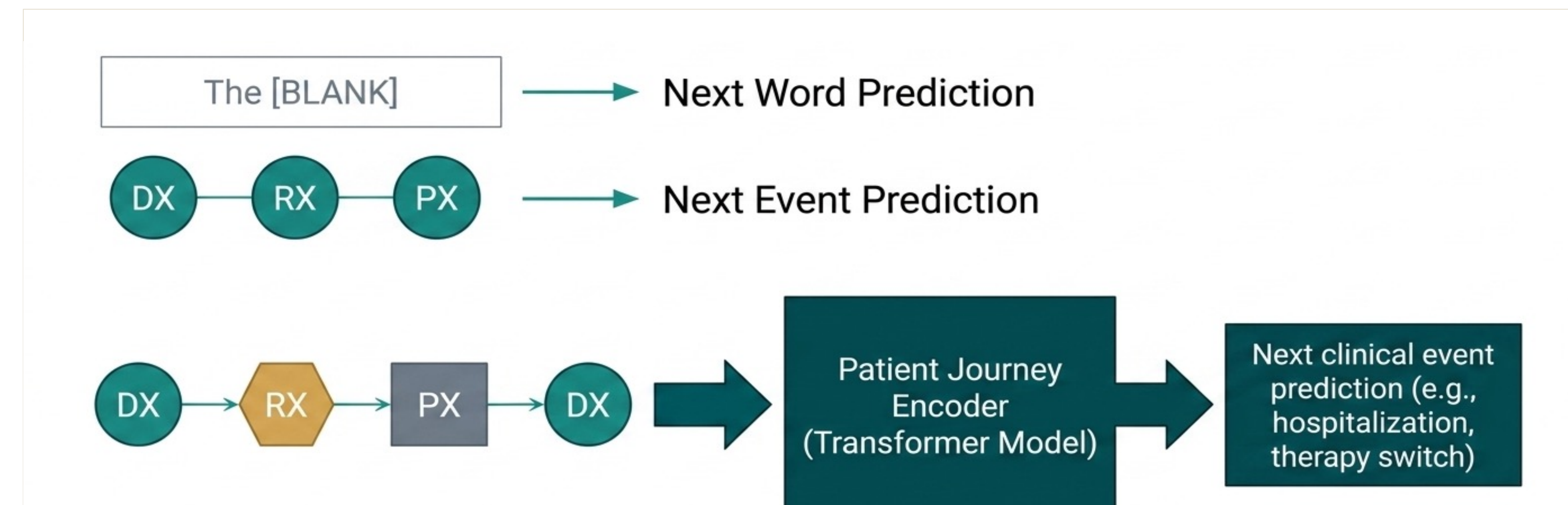


**Objective:** Develop an interactive AI-driven framework to streamline cohort definition, uncover antiretroviral treatment (ART) patterns, and enable scalable trial emulation in people with HIV (PWH) by leveraging graph neural networks, clustering algorithms, and a large language model (LLM)-powered interface.

**Figure 2.** End-to-end framework: messy claims → graph embeddings → unsupervised clustering → LLM-powered refinement → actionable cohorts.

## Patient-Journey Encoder

We treat each patient's journey like a sentence. Just as a language model learns by filling in the blank word from prior context ("The \_\_\_" → "cat"), our encoder learns by masking a diagnosis (DX), prescription (RX), or procedure (PX) event and predicting it from the surrounding clinical history. This pretext task forces the transformer to learn the latent structure of ART — producing a dense embedding of each patient's longitudinal journey.



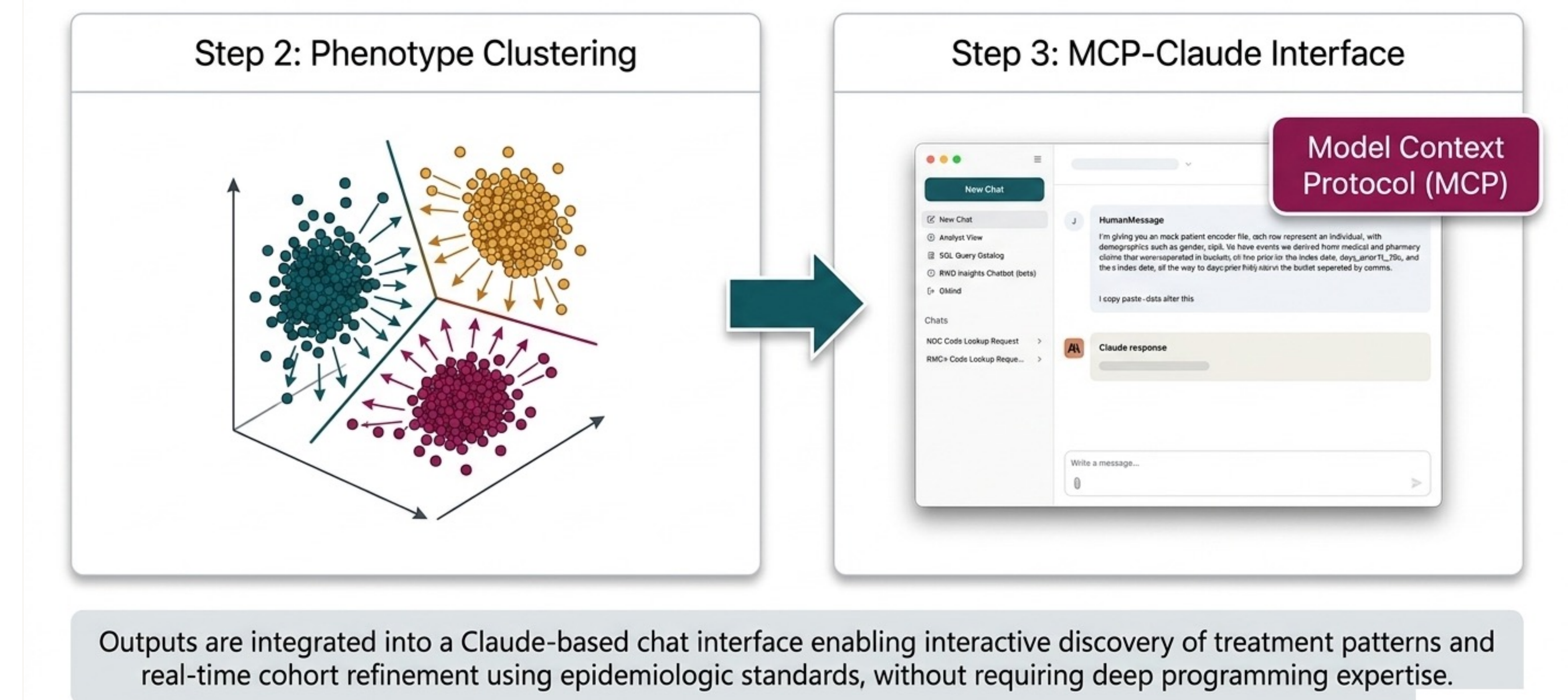
Graphs were built representing ART treatments as time-linked nodes. A graph encoder modeled each patient's ART journey using masked treatment prediction as a pretext task.

**Figure 3. Patient Journey as a Language Model.** Top: the language-model analogy — predicting the next word ("The [BLANK]") is analogous to predicting the next clinical event from a DX → RX → PX sequence. Bottom: time-linked DX / RX / PX nodes are fed into a transformer encoder that predicts held-out events (e.g., hospitalization, therapy switch), yielding embeddings that power downstream clustering and prediction.

<b>&gt;95%</b> Masked-ART prediction accuracy (graph encoder)	<b>~74%</b> Overall HIV-status prediction accuracy (transformer)	<b>22,431</b> Adults with HIV on ART, ≥18 mo follow-up
--	---	---

## Phenotype Clustering & LLM Interface

Patient embeddings and curated line-of-therapy features were fed into agglomerative clustering. The resulting subgroups, with SME-validated definitions, were packaged as an MCP server and exposed through a Claude chat interface.



**Figure 4. Phenotype Clustering & MCP-Claude Interface.** Unsupervised clustering (left) and the MCP-powered Claude chat (right) used by analysts to define, inspect, and refine cohorts interactively.

## Study Population

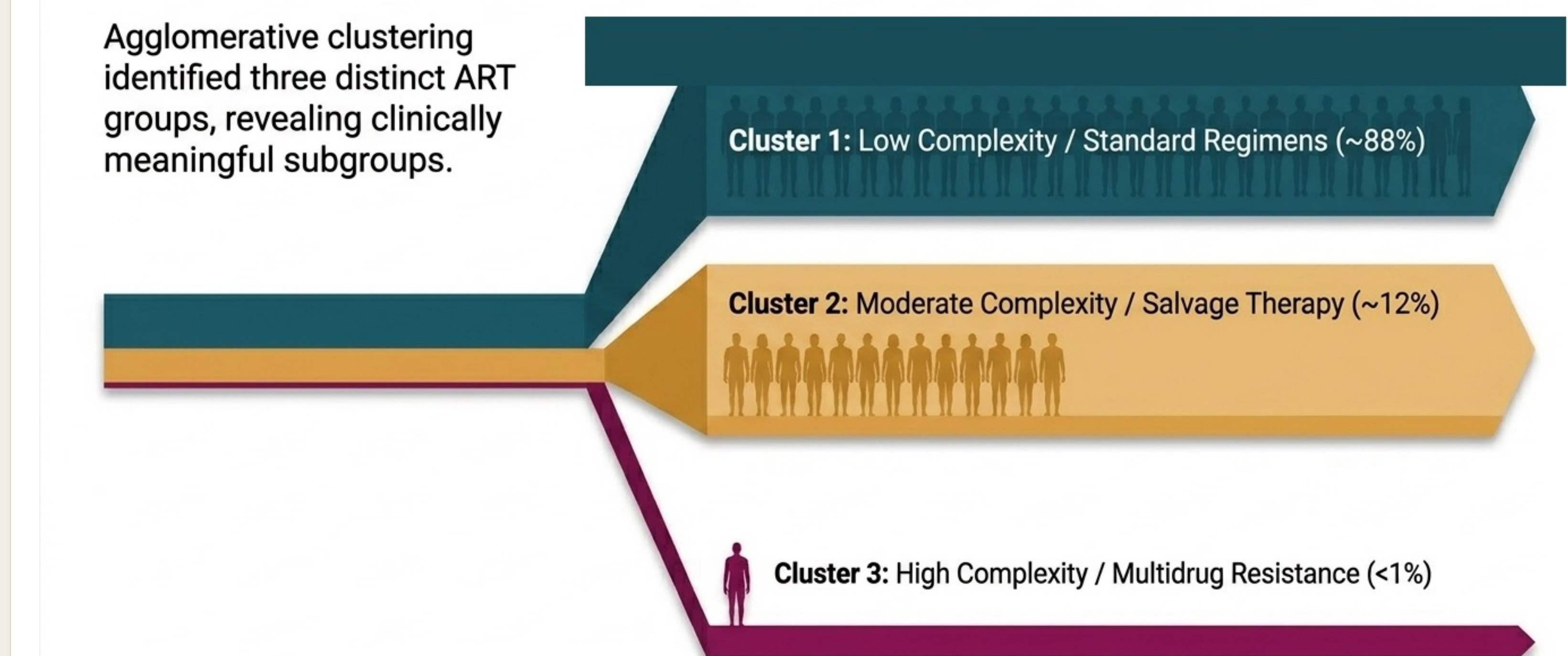
**Table 1. Cohort attrition (HealthVerity, 2022–2025)**

Attrition step	Unique HIVIDs
≥1 ART prescription, 2022-01-01 to 2025-06-30	499,675
+ ≥1 HIV Dx on/prior to index ART date	75,452
+ Age ≥18 at index ART	74,839
+ Non-missing sex	74,796
+ No pregnancy code during study period	71,381
+ ≥18 months follow-up post-index ART	22,431

Mean age 42.7 y (SD 13.2); 79.6% male at birth; South (38.7%) and West (33.7%) dominant; 36.5% Medicaid, 23.8% Commercial.

## Revealing Hidden Subgroups

Agglomerative clustering on encoder-derived features identified three distinct ART groups — revealing clinically meaningful subgroups that guideline-driven definitions would collapse together.



**Figure 5. Revealing the Hidden Patient Subgroups.** Agglomerative clustering identified three clinically meaningful ART subgroups, sized proportionally to cohort share.

## Three ART Patient Clusters

**Table 2. Patient cluster characteristics (N = 22,431)**

	C1 · Standard / 1L maintenance	C2 · Moderate / salvage	C3 · High / HTE
<b>Size (%)</b>	19,672 (~88%)	2,694 (~12%)	65 (<1%)
<b>Mean age, y</b>	42.0	47.0	52.2
<b>Dominant regimens</b>	B/F/TAF (60.6%), DTG/3TC	DRV/c/FTC/TAF (35.3%), CAB/RPV (14.3%)	FTR, MVC (+ B/F/TAF or DTG), LEN
<b>Drug classes</b>	INSTI 92.7%, NRTI 90.6%	PI 57.9%, INSTI 43.7%, NNRTI 33.8%	EI 49%, AI 46%, CI 6%
<b>Median duration</b>	166 d	128 d	84 d
<b>Detectable VL (L1)</b>	9.5%	11.8%	21.5%
<b>CCI, mean (SD)</b>	1.00 (1.93)	1.44 (2.43)	1.83 (2.67)
<b>HTE criteria met</b>	0.5%	8.2%	38.5%

## Real-World Benchmarking

**Table 3. Alignment with established real-world evidence**

ART complexity group	This study (HV)	RWE range (CDC, OPERA)	Alignment
<b>Low (stable INSTI)</b>	88%	80–90%	Strong alignment
<b>Moderate (cycling)</b>	12%	10–18%	Within expected range
<b>High (HTE / multi-class)</b>	<1%	1–5%	Directionally correct

## Conclusions

- Integrating generative AI and deep learning into RWE studies improves efficiency in defining patient cohorts and generating insights, while preserving epidemiologic rigor via validation, sensitivity analyses, and expert-reviewed guideline mapping.
- Three distinct ART clusters emerged — treatment-naïve/stable, PI-heavy treatment-experienced, and heavily treatment-experienced on novel multi-drug regimens — consistent with external CDC / OPERA benchmarks.
- The MCP-Claude interface enables interactive discovery of treatment patterns and real-time cohort refinement without requiring deep programming expertise.
- Future work will evaluate generalizability across other therapeutic areas and clinical-trial emulation settings.

### REFERENCES

1. Yun S, et al. Graph transformer networks. *Adv Neural Inf Process Syst.* 2019;32.
2. Chowdhury G, et al. *AIDS.* 2001;15:2101–8.
3. Jordana B, et al. Clinical phenotypes in HFpEF. *JACC Heart Failure.* 2020.
4. CDC. HIV Surveillance Supplemental Report. 2023;28(4).
5. Hsu RK, et al. Heavily treatment-experienced PLWH in OPERA. *BMC Infect Dis.* 2023;23:303.

### ABBREVIATIONS

AI, artificial intelligence; ART, antiretroviral therapy; CAB, cabotegravir; CCI, Charlson comorbidity index; GNN, graph neural network; HTE, heavily treatment-experienced; INSTI, integrase strand transfer inhibitor; LEN, lenacapavir; LLM, large language model; MCP, Model Context Protocol; NNRTI, non-nucleoside reverse transcriptase inhibitor; NRTI, nucleoside reverse transcriptase inhibitor; PI, protease inhibitor; PWH, people with HIV; RPV, rilpivirine; RWE, real-world evidence; STR, single-tablet regimen.

### ACKNOWLEDGMENTS & DISCLOSURES

Funding. Gilead Sciences Innovation Fund. Computing infrastructure via AWS on Gilead's internal cloud platform.

Disclosures. CYC, SJ, JG, LL, AM, TL, RA, BC, SS, MJJC, GL, LT, WW, AA are employees of Gilead Sciences, Inc. and own stock/shares. JAR reports no conflicts. MF is an employee of Amazon Web Services. Correspondence. Ching-Yi (James) Chuo — james.chuo@gilead.com