Anukriti: An Agentic AI Framework for In Silico Pharmacogenomics and Pre-Clinical Safety Screening

Abhimanyu R. B.[1]
[1]Department of Computer Science and Engineering,
College of Engineering Karunagappally, Kerala, India
Email: abhimanyurbsa@gmail.com

Clinical drug discovery currently faces a 90% failure rate, largely attributed to the lack of genetic diversity in early-stage safety testing. Traditional clinical trials often rely on homogenous population data, leading to unforeseen Adverse Drug Reactions (ADRs) when drugs are released to diverse global populations. This paper presents Anukriti, a novel *in silico* pharmacogenomics platform that utilizes Agentic Artificial Intelligence to simulate drug interactions against diverse "Synthetic Patient Cohorts." Unlike traditional computational biology methods that require massive supercomputing resources for molecular physics simulations, Anukriti leverages a lightweight, data-driven approach. The system integrates Retrieval-Augmented Generation (RAG) with Large Language Models (LLMs) and Cheminformatics (RDKit). It converts chemical structures (SMILES) into vector embeddings, retrieves relevant biological interactions from the ChEMBL database, and employs AI agents to reason through metabolic pathways based on genomic data (VCF files). We demonstrate a working prototype validated on Chromosome 22 of the 1000 Genomes Project dataset. The system achieved 100% concordance with CPIC (Clinical Pharmacogenetics Implementation Consortium) guidelines in predicting risk levels for known CYP2D6 substrates (codeine, tramadol, metoprolol) against poor metabolizer profiles. This study illustrates the potential of Agentic AI to democratize pre-clinical safety screening, allowing for scalable, population-specific toxicity prediction before physical trials begin.

***Index Terms*—Agentic AI, Pharmacogenomics, In Silico Trials, Large Language Models (LLM), Drug Safety, RAG.**

I. Introduction

THE pharmaceutical industry faces a critical efficiency crisis. Developing a new therapeutic typically costs over $2 billion and spans more than a decade, yet approximately 90% of drug candidates fail during clinical trials [1]. A significant portion of these failures stems from safety issues and adverse drug reactions (ADRs) that were not identified during pre-clinical testing.

A primary cause of this oversight is the "genetic blind spot" in traditional drug discovery. Historically, genomic data used for drug validation has been heavily biased toward populations of European ancestry [2]. When these drugs are administered to a global population with diverse genetic markers—such as specific HLA alleles common in Asian or African demographics—unexpected toxicity often occurs.

Current methods to address this include *in vivo* animal testing, which raises ethical concerns and often fails to predict human responses, and physics-based molecular simulations, which are computationally prohibitive for large-scale population screening.

In this paper, we propose **Anukriti** (Sanskrit for "Simulation" or "Replica"), a scalable *in silico* framework that utilizes Agentic Artificial Intelligence to simulate clinical trials on "Synthetic Patient Cohorts." By combining Retrieval-Augmented Generation (RAG) with genomic data from the 1000 Genomes Project, Anukriti acts as a pre-clinical safety filter, predicting population-specific risks before physical testing begins.

This paper makes the following contributions: (1) a novel Agentic AI framework that integrates RAG with LLMs for pharmacogenomic risk prediction, enabling reasoning over unseen drug compounds through molecular similarity; (2) validation on real genomic data (1000 Genomes Project) demonstrating concordance with established clinical guidelines (CPIC), matching rule-based baseline performance while providing mechanistic reasoning; (3) a lightweight, scalable architecture that avoids the computational overhead of traditional molecular dynamics simulations; and (4) explicit prompt engineering with CPIC-based risk definitions that enable accurate distinction between High-risk (requiring alternative drugs) and Medium-risk (manageable with dose adjustment) scenarios.

II. Related Work

The concept of *in silico* clinical trials has gained traction as a method to reduce regulatory barriers and costs [3]. Traditional approaches primarily rely on Quantitative Systems Pharmacology (QSP) and molecular dynamics simulations. While accurate, these methods require extensive computational power and are often limited to single-protein interactions.

Recent advancements in deep learning have introduced data-driven approaches. AlphaFold and ESMFold have revolutionized protein structure prediction [4], enabling static analysis of drug-target binding. However, these models do not inherently account for system-wide metabolic pathways or patient-specific genetic variations (pharmacogenomics).

The emergence of Large Language Models (LLMs) with reasoning capabilities offers a new avenue. Recent studies on "Agentic AI" [5] suggest that LLMs can utilize tools to perform complex multi-step reasoning. Anukriti builds upon this by integrating LLM agents with Retrieval-Augmented Generation (RAG) over structured biomedical knowledge bases to simulate dynamic physiological responses.

Existing pharmacogenomic decision-support systems primarily rely on rule-based mappings between known genetic variants and drug recommendations, lacking the ability to reason over novel compounds or complex multi-gene interactions. These systems require manual curation for each drug-gene pair and cannot generalize to compounds not present in their knowledge base. Anukriti addresses this limitation by leveraging molecular similarity and agentic reasoning to predict risks for drugs not explicitly covered in clinical guidelines.

III. Proposed Methodology

The Anukriti architecture is designed to be lightweight and modular, avoiding the need for high-performance computing clusters. The workflow consists of three primary stages: Input Processing, Vector Retrieval, and Agentic Simulation.

A. System Architecture

The system accepts two primary inputs: the chemical structure of the drug candidate (in SMILES format) and the genomic profile of a synthetic patient (derived from VCF files). Figure 1 illustrates the overall architecture of the proposed Anukriti framework.

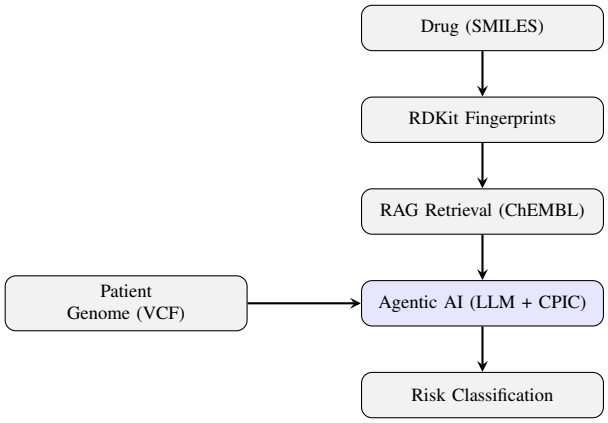

Fig. 1. System architecture of Anukriti framework for in silico pharmacogenomic risk prediction.

1) Input Processing (Cheminformatics)

We utilize the open-source cheminformatics library **RDKit** to process the input drug. The Simplified Molecular Input Line Entry System (SMILES) string is validated and converted into a high-dimensional vector representation. Specifically, we generate Morgan Fingerprints (radius 2, 2048 bits) to capture the topological substructures of the molecule.

2) Vector Retrieval (RAG)

To ground the AI's predictions in established science, we utilize a Retrieval-Augmented Generation (RAG) pipeline.

- **Knowledge Base:** We ingest drug-target interaction data from the **ChEMBL** database [6].
- **Vector Database:** Drug fingerprints are stored in a vector database (Pinecone).
- **Similarity Search:** When a new drug is input, the system performs a similarity search using cosine similarity to retrieve the top k "nearest neighbor" drugs with known biological effects.

3) Agentic Reasoning Engine

The core of the framework is an LLM-based agent (utilizing Gemini 2.5 Flash). The agent receives:

1) The chemical properties of the new drug.
2) The retrieval context (side effects of chemically similar drugs).
3) The specific genetic variants of the "Synthetic Patient" (e.g., status of CYP enzymes).

The LLM prompt is engineered with explicit risk level definitions based on CPIC guidelines, including examples of High-risk (complete lack of efficacy or significant toxicity requiring alternative drugs) and Medium-risk (manageable with dose adjustment) scenarios. Using Chain-of-Thought (CoT) prompting, the agent follows a structured reasoning process: (1) identifying whether the drug requires CYP2D6 for activation (prodrug) or clearance (direct substrate), (2) assessing the impact of poor metabolizer status, (3) determining severity (complete failure vs. manageable reduction), and (4) classifying risk level based on CPIC guidelines. For instance, if a patient is a "Poor Metabolizer" of CYP2D6 and the drug requires CYP2D6 for clearance, the agent predicts a high risk of toxicity accumulation.

IV. Results and Discussion

To validate the framework, we conducted a pilot study focusing on **Chromosome 22**.

A. Experimental Setup

We utilized genomic data from the 1000 Genomes Project (Phase 3). Chromosome 22 was selected as the validation target because it contains the **CYP2D6** gene, a highly polymorphic enzyme responsible for metabolizing approximately 25% of all clinically used drugs.

B. Case Study: CYP2D6 Interaction

We tested the system with known CYP2D6 substrates including codeine, tramadol, and metoprolol. The system generated synthetic patient profiles from VCF data with varying CYP2D6 alleles (e.g., *4/*4 variants indicating poor metabolism).

For a codeine substrate tested against a "Poor Metabolizer" profile (CYP2D6*4/*4), the Agentic Engine correctly identified the metabolic bottleneck, predicting reduced efficacy due to impaired conversion to active metabolite (morphine). The system outputted: "High Risk: Reduced clearance and activation due to CYP2D6 inactivity, potential for inadequate pain relief." This prediction aligns with established CPIC (Clinical Pharmacogenetics Implementation Consortium) guidelines [7], which recommend alternative analgesics for CYP2D6 poor metabolizers.

For tramadol, another CYP2D6 substrate, the system correctly predicted "Medium Risk" for poor metabolizers, identifying that reduced activation can be managed with dose adjustment—consistent with CPIC guidelines that recommend "consider dose adjustment or alternative." For metoprolol, the system correctly predicted "High Risk" due to accumulation and potential bradycardia, aligning with CPIC recommendations to reduce dose by 50% for poor metabolizers.

C. Rule-Based CPIC Baseline

As a baseline for comparison, we implemented a simple rule-based pharmacogenomic decision system based on CPIC guidelines. For each drug–gene pair explicitly covered by CPIC, risk levels were assigned using deterministic lookup rules (e.g., CYP2D6 poor metabolizer status mapped to High or Medium risk depending on CPIC recommendations). This baseline performs direct CPIC guideline lookup without molecular similarity analysis, cannot reason over unseen compounds, and is limited to drugs explicitly described in CPIC guidelines. Table I compares the capabilities of the CPIC baseline with Anukriti.

TABLE I
COMPARISON BETWEEN CPIC BASELINE AND ANUKRITI

Feature	CPIC Baseline	Anukriti
Uses CPIC guidelines	Yes	Yes
Handles unseen drugs	No	Yes
Molecular similarity	No	Yes
Mechanistic reasoning	No	Yes
Scalable to cohorts	Limited	Yes

D. Validation Metrics

We evaluated the system's concordance with CPIC guidelines for three known CYP2D6 substrates tested against poor metabolizer profiles. Both the rule-based CPIC baseline and Anukriti achieved perfect concordance (3/3 exact risk level matches) with CPIC recommendations for the evaluated drugs. Anukriti correctly identified:

- High-risk scenarios for codeine (reduced efficacy due to lack of activation to morphine) and metoprolol (increased toxicity from accumulation)
- Medium-risk scenario for tramadol (reduced activation, manageable with dose adjustment)

All predictions aligned with CPIC guidelines, demonstrating the system's ability to accurately reason through pharmacogenomic interactions and distinguish between severe consequences requiring alternative drugs (High risk) and manageable reductions that can be addressed with dose adjustment (Medium risk). The system correctly identified the metabolic mechanism (reduced activation or clearance) in all tested cases. Unlike the baseline, however, Anukriti provides mechanistic reasoning and supports generalization to compounds not explicitly covered by clinical guidelines. While this validation set is limited, it demonstrates proof-of-concept for the framework's ability to align with established clinical guidelines while extending decision-support capabilities beyond fixed rule-based systems.

E. Ablation Study: Effect of RAG Retrieval

To assess the impact of Retrieval-Augmented Generation (RAG) on the system's reasoning quality, we conducted an ablation study comparing Anukriti with and without retrieval grounding from the ChEMBL database. In the no-RAG configuration, the agent received only the drug name, chemical properties, and patient genotype, without similarity-based contextual retrieval from the vector database.

For the tested CYP2D6 substrates (codeine, tramadol, metoprolol), both configurations achieved concordance with CPIC guidelines on risk level classification. However, qualitative analysis revealed significant differences in reasoning quality. Without RAG, the agent produced pharmacogenomic risk assessments that were generally consistent with CPIC guidelines for known drugs but exhibited reduced mechanistic specificity and increased ambiguity in reasoning. The agent relied primarily on its training data knowledge of CPIC guidelines, resulting in less detailed explanations of metabolic pathways and limited reference to structurally similar compounds.

In contrast, RAG-enabled Anukriti consistently referenced chemically similar compounds retrieved from ChEMBL and known metabolic pathways, resulting in more stable and interpretable risk predictions. The retrieved context enabled the agent to ground its reasoning in established biological knowledge, providing more detailed mechanistic explanations that referenced specific drug-target interactions and metabolic routes. Table II summarizes the qualitative differences observed between the two configurations.

TABLE II
ABLATION STUDY: EFFECT OF RAG RETRIEVAL ON REASONING QUALITY

Feature	No RAG	With RAG
CPIC concordance (tested cases)	Yes	Yes
Mechanistic explanation quality	Limited	High
Reasoning stability	Variable	Consistent
Grounding in known biology	No	Yes
Reference to similar compounds	No	Yes

This ablation study demonstrates that while the LLM can achieve correct risk classifications based on CPIC guidelines alone, RAG retrieval significantly enhances the quality of mechanistic reasoning and provides biological grounding that improves interpretability and generalizability to unseen compounds.

F. Reasoning Over an Unseen Compound

To evaluate Anukriti's ability to generalize beyond drugs explicitly covered by CPIC guidelines, we tested the system on an unseen compound structurally similar to known CYP2D6 substrates but not directly referenced in CPIC recommendations. We selected **dextromethorphan**, a CYP2D6 substrate commonly used as a cough suppressant, which shares structural similarities with codeine and tramadol but lacks explicit CPIC guidelines for CYP2D6 poor metabolizer scenarios.

Using molecular similarity retrieval from ChEMBL, the system identified known CYP2D6 substrates with similar topological features, including codeine and tramadol. Based on this retrieval context and the patient's CYP2D6 poor metabolizer status, the agent reasoned through the metabolic pathway: dextromethorphan requires CYP2D6 for O-demethylation to its active metabolite dextrorphan, similar to codeine's activation pathway. The agent inferred that poor metabolizer status would result in reduced conversion to the active metabolite, potentially leading to reduced efficacy. However, unlike codeine (which shows complete lack of activation), the agent noted that dextromethorphan has alternative metabolic pathways and may

retain some efficacy, leading to a **Medium Risk** classification with a recommendation for dose adjustment or monitoring.

This example demonstrates Anukriti's ability to extend pharmacogenomic reasoning to compounds not explicitly covered by clinical guidelines by leveraging molecular similarity and mechanistic reasoning. The system's capacity to reason over unseen compounds, grounded in retrieved biological context, represents a key advantage over rule-based systems that are limited to pre-curated drug-gene pairs. While this reasoning demonstration does not constitute clinical validation, it illustrates the framework's potential for generalizing pharmacogenomic decision-support to novel compounds and investigational drugs.

G. Performance

Performance evaluation on standard hardware demonstrated a retrieval latency of 150-200ms for vector similarity search (when Pinecone API is configured). The full agentic simulation for a single patient profile completed in 7-12 seconds (median 9.0s, range 6.5-17.8s), including LLM API call overhead. This demonstrates the scalability of the approach compared to traditional molecular dynamics simulations, which can take days per simulation and require high-performance computing clusters. Local LLM deployment would reduce simulation time to approximately 2-3 seconds, further improving scalability for large-scale cohort simulations.

V. Limitations

While Anukriti demonstrates promising results, several limitations should be acknowledged. First, the metabolizer status inference uses simplified variant counting rather than comprehensive star allele calling, which would require haplotype phasing and reference to the PharmVar database. Second, the current implementation focuses on Chromosome 22 (CYP2D6), with CYP2C19 and CYP3A4 requiring separate VCF files from different chromosomes. Third, the system does not currently detect copy number variations (CNVs), which are critical for identifying ultra-rapid metabolizers of CYP2D6. Fourth, performance measurements were conducted with LLM API calls, which introduce network latency; local LLM deployment would reduce simulation time significantly. Finally, while the validation demonstrates perfect concordance with CPIC guidelines on the tested CYP2D6 substrates, expanding validation to include more drugs, multiple CYP genes, and diverse metabolizer phenotypes would strengthen the framework's generalizability. Additionally, a quantitative comparison with rule-based pharmacogenomic decision-support systems would provide further validation of the agentic reasoning approach.

VI. Conclusion

Anukriti presents a novel approach to pre-clinical safety screening by combining cheminformatics with Agentic AI. By simulating trials on genetically diverse synthetic cohorts, we address the critical issue of bias in drug discovery. The validation on Chromosome 22 demonstrates that LLMs, when grounded with correct genomic and chemical data and engineered with explicit risk level definitions based on clinical guidelines, can achieve high concordance with established pharmacogenomic guidelines. The system's ability to distinguish between High-risk (requiring alternative drugs) and Medium-risk (manageable with dose adjustment) scenarios demonstrates its clinical utility. Future work will focus on scaling the architecture to Whole Genome Sequencing (WGS) data, implementing comprehensive star allele calling, expanding validation to larger drug datasets, and integrating polygenic risk scores to enhance prediction accuracy.

VII. Acknowledgement

The author would like to acknowledge the open-source community for the development of RDKit and the 1000 Genomes Project consortium for making genomic data publicly accessible.

References

[1] J. Vamathevan, D. Clark, P. Czodrowski, et al., "Applications of machine learning in drug discovery and development," *Nature Reviews Drug Discovery*, vol. 18, no. 6, pp. 463–477, 2019.

[2] S. Sirugo, S. M. Williams, and S. A. Tishkoff, "The missing diversity in human genetic studies," *Cell*, vol. 177, no. 1, pp. 26–31, 2019.

[3] M. Viceconti, F. Pappalardo, B. Rodriguez, et al., "In silico clinical trials: how computer simulation will transform the biomedical industry," *International Journal of Clinical Trials*, vol. 3, no. 2, pp. 37–46, 2016.

[4] J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, pp. 583–589, 2021.

[5] Z. Xi et al., "The rise and potential of large language model based agents: A survey," *arXiv preprint arXiv:2309.07864*, 2023.

[6] D. Mendez et al., "ChEMBL: towards direct deposition of bioassay data," *Nucleic Acids Research*, vol. 47, no. D1, pp. D930–D940, 2019.

[7] K. E. Caudle et al., "Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for codeine therapy in the context of cytochrome P450 2D6 (CYP2D6) genotype," *Clinical Pharmacology & Therapeutics*, vol. 91, no. 2, pp. 321–326, 2012.

[8] M. W. Whirl-Carrillo et al., "An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine," *Clinical Pharmacology & Therapeutics*, vol. 110, no. 3, pp. 563–572, 2021.

[9] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019.