



Molecular biology & basic cellular physiology
Ethics, innovative research, businesses & IPR

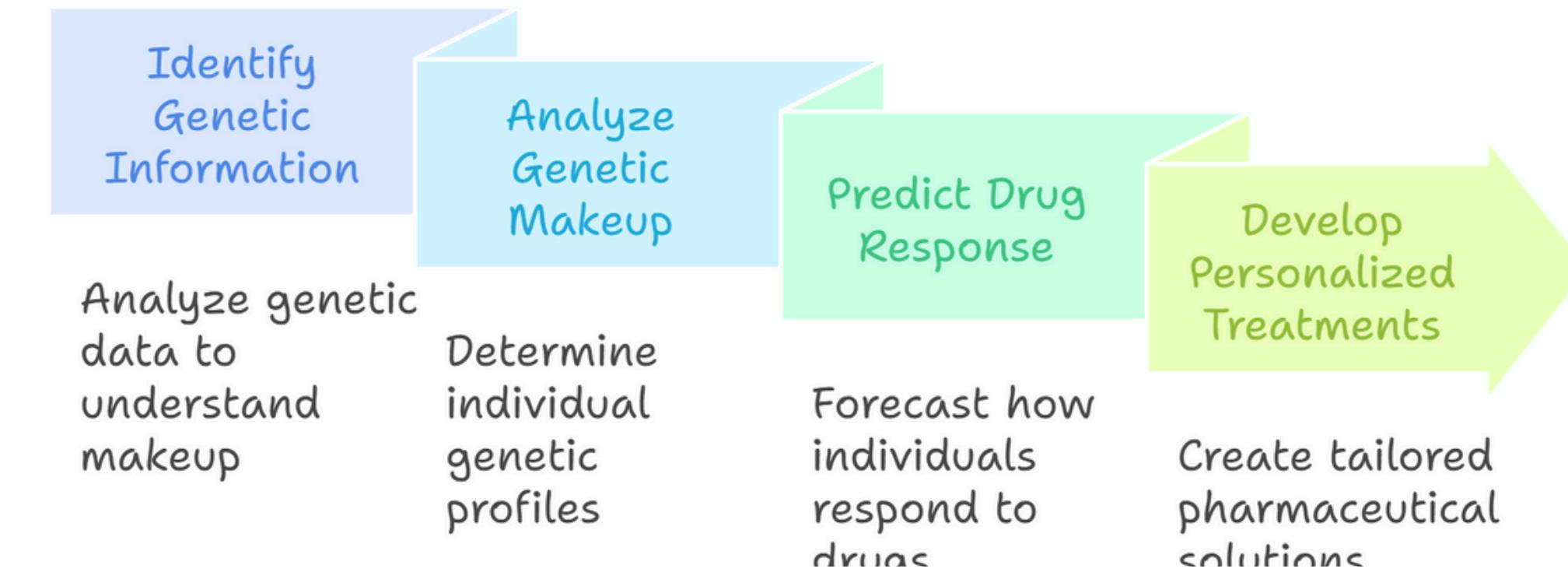
PharmaGen AI: Drug Prediction Through Gene Sequencing

Aayush Mohandas Nair - CB.AI.U4AIM24001
Sahana R - CB.AI.U4AIM24037
Shreya V - CB.AI.U4AIM24041
Siddharth P - CB.AI.U4AIM24043

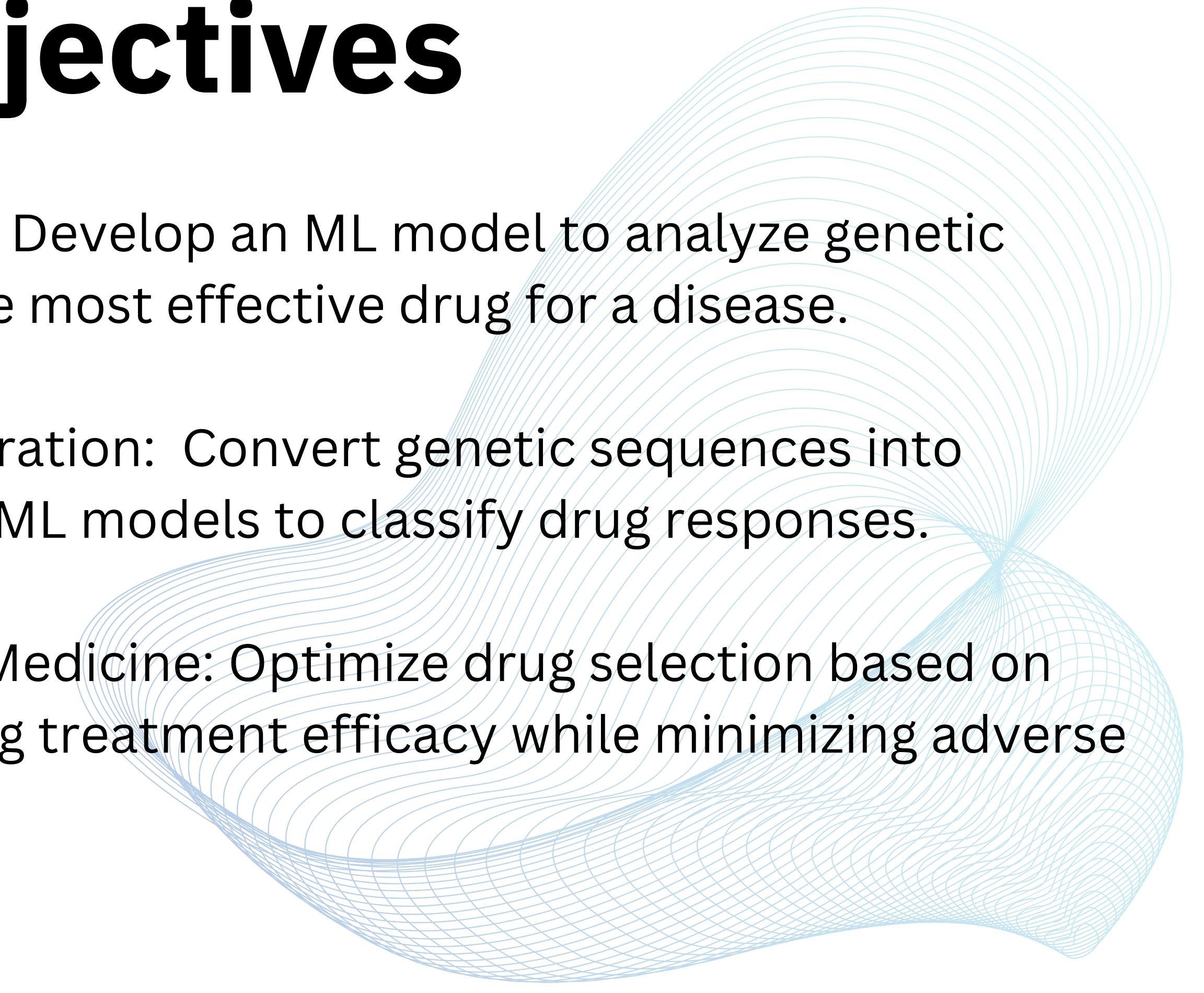
Contents

- Objectives
- Literature Review
- Timeline
- Computational Aspects
- Dataset
- Ethics
- IPR Rules
- Novelty
- References

Pharma Gen AI Drug Prediction Process



Objectives

- 
1. Predict Drug Response: Develop an ML model to analyze genetic variations and identify the most effective drug for a disease.
 2. Machine Learning Integration: Convert genetic sequences into numerical form and train ML models to classify drug responses.
 3. Enhance Personalized Medicine: Optimize drug selection based on genetic profiles, improving treatment efficacy while minimizing adverse effects.

Literature Review

SL.NO.	PAPER NAME	SOURCE	METHOD USED	LIMITATIONS
1.	A Predictive Model for Drug Response Using Machine Learning Techniques	Bioinformatics Journal (Silverchair)	<ul style="list-style-type: none">- Talks about the different machine learning models- Highlights the advantages of SVM for genetic data classification.	<ul style="list-style-type: none">- Limited dataset size, which may impact generalizability.- Does not compare SVM with deep learning techniques.
2.	Personalized Medicine Approach Using Genetic Data for Drug Response Prediction	Springer	<ul style="list-style-type: none">- Genetic variants that affect drug response are identified using bioinformatics tools.- Talks about the role of genetic variations in treatment effectiveness.	<ul style="list-style-type: none">- Does not talk about the machine learning model and how it's applied- Feature selection is not well optimized, introducing potential noise.
3.	Machine Learning for Pharmacogenomics: A Study on Drug-Gene Interactions	IEEE Xplore	<ul style="list-style-type: none">- Uses Regression for drug-gene interactions.- Calculates a score for each drug and then use these scores to rank the drugs for a specific cancer cell line.	<ul style="list-style-type: none">- Lack of enough data points- Lacks a strong validation process, making results less reliable.

Project Timeline for Gene Sequencing Analysis

Identify project scope and finalize problem statement

**January 2025,
Week 1-2**

Preprocess gene sequencing data

**February 2025,
Week 1**

Train models and optimize hyperparameters

**February 2025,
Week 3**

Compare model results and analyze metrics

**March 2025,
Week 1**

Prepare documentation and visualize results

**March 2025,
Week 3**



**January 2025,
Week 3-4**

Gather datasets and assess feasibility

**February 2025,
Week 2**

Select and implement ML models

**February 2025,
Week 4**

Validate models and refine

**March 2025,
Week 2**

Integrate best model into pipeline

**March 2025,
Week 4**

Final submission and presentation

Computational Aspects

- Dataset Preprocessing
 - Handling Missing Values
 - Encoding Categorical Variables
 - Normalizing Numerical Features
- Feature engineering
 - Identify meaningful features from genetic variants, drug properties, and patient response data.
- Data Splitting
- Train SVM Model
- Model Evaluation
- Predict Drug Response

Dataset

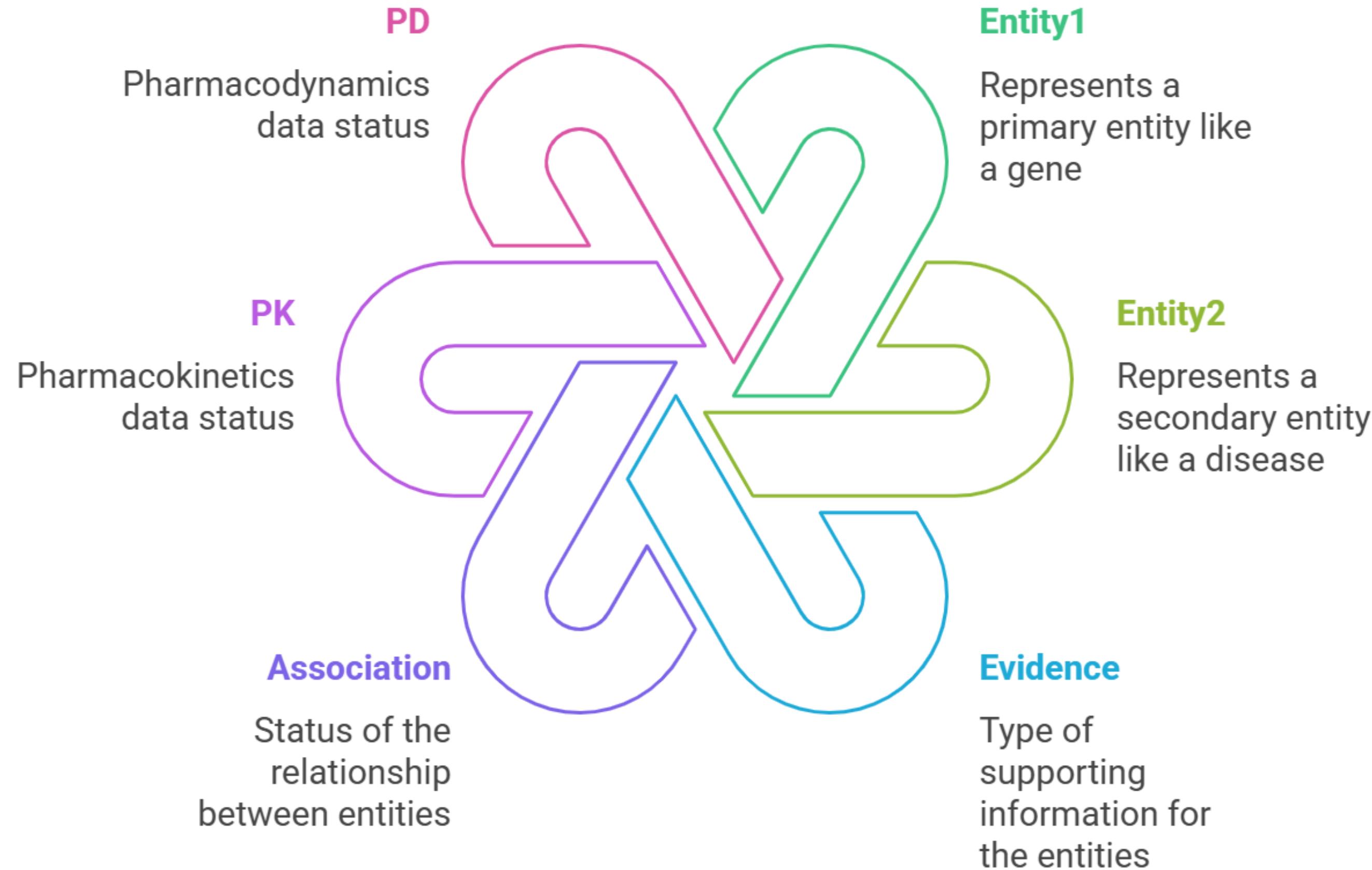
	A	B	C	D	E	F	G	H	I	J	K	L
1	Entity1_id	Entity1_name	Entity1_type	Entity2_id	Entity2_name	Entity2_type	Evidence	Association	PK	PD	PMIDs	
2	PA142672	ANKFN1	Gene	PA447288	Essential	h Disease	ClinicalAnr	associated	PD	25695618		
3	PA142672	ANKFN1	Gene	PA449899	hydrochloride	Chemical	ClinicalAnr	associated	PD	25695618		
4	PA164722	MIR2053	Gene	PA166122	mucositis	Disease	VariantAnr	not associated		28628559		
5	PA164722	MIR2053	Gene	PA443937	Drug	Toxic Disease	VariantAnr	not associated		24614921		
6	PA164722	MIR2053	Gene	PA446155	Precursor	Disease	VariantAnr	not associated		24614921;28628559		
7	PA164722	MIR2053	Gene	PA448492	asparagine	Chemical	VariantAnr	not associated		24614921		
8	PA164722	MIR2053	Gene	PA449165	cyclophosphamide	Chemical	VariantAnr	not associated		24614921		

Source : PharmaGKB

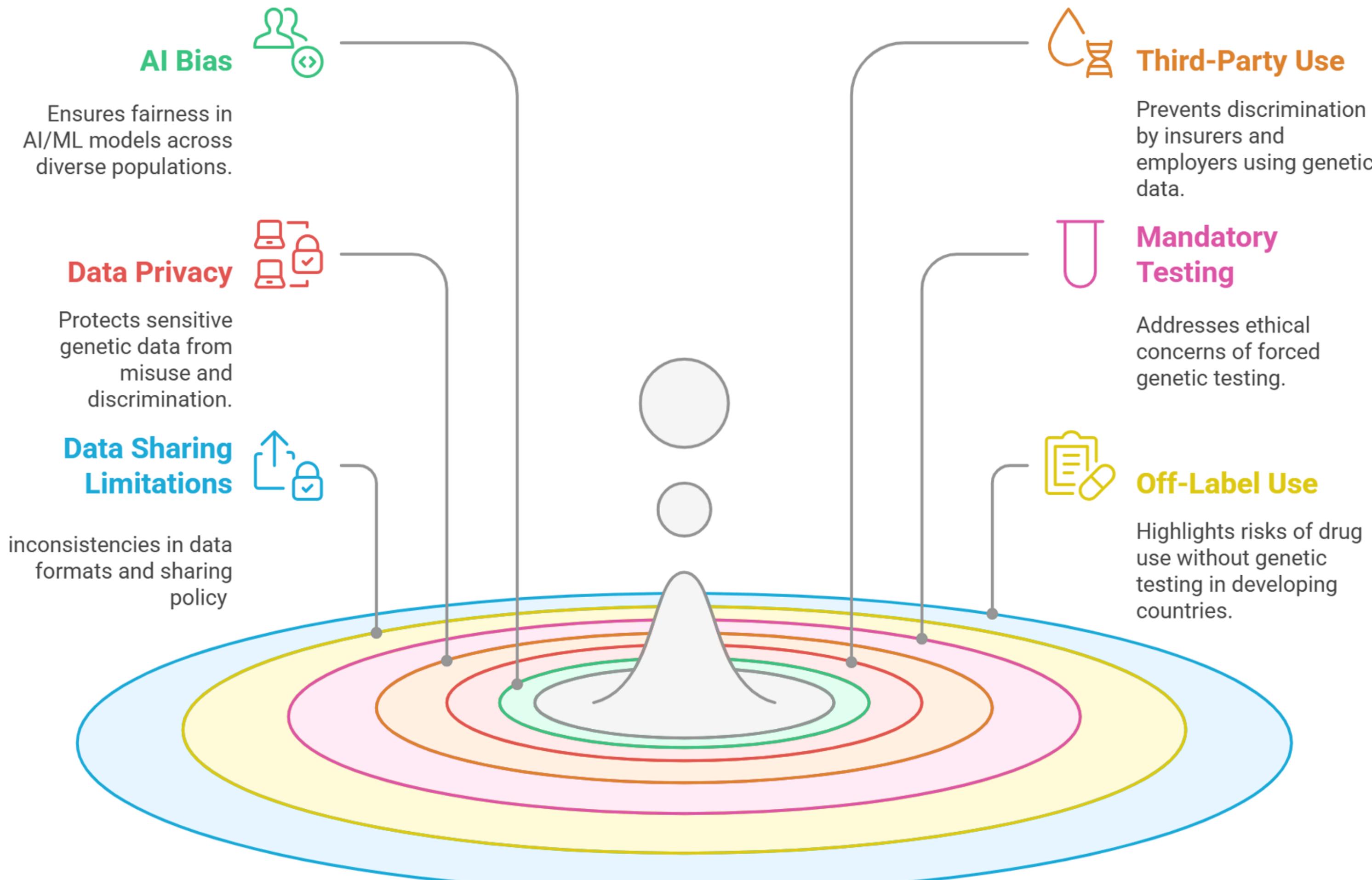
	A	B	C	D	E	F	G	H	I	J	K
1	Entity1_id	Entity1_na	Entity1_ty	Entity2_id	Entity2_na	Entity2_ty	Evidence	Associatio	PD	PMIDs	
2	-1.60285	-1.39459	0.123921	0.933748	-1.02465	-0.52142	0.417589	0.65777	0	0.738207	
3	-1.60285	-1.39459	0.123921	1.010449	0.026924	-1.16676	0.417589	0.65777	0	0.738207	
4	0.587378	-0.57831	0.123921	-1.72703	0.033929	-1.16676	0.183151	-1.52029	0	1.108059	
5	0.587378	-0.57831	0.123921	-1.71182	-0.00655	-1.16676	0.417589	-1.52029	0	-0.69101	
6	0.587378	-0.57831	0.123921	-1.70695	0.127333	-1.16676	-0.14506	-1.52029	0	0.62593	
7	0.587378	-0.57831	0.123921	-1.58885	-0.92813	-0.52142	0.417589	0.65777	0	1.108059	
8	0.587378	-0.57831	0.123921	-1.52311	-0.44632	-1.16676	-2.67699	0.65777	0	0.017655	

After Preprocessing

Breakdown of Data Attributes



Ethical Considerations in Drug Prediction

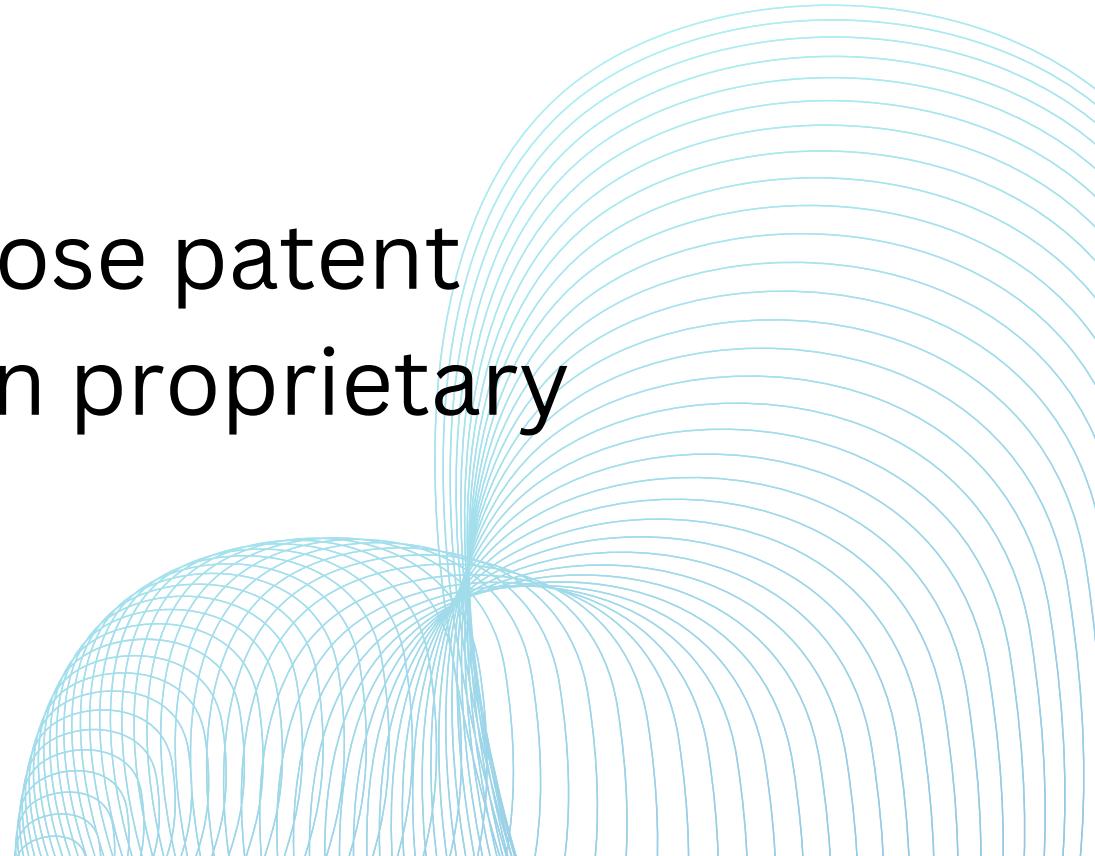


Ethical Insights

- **Pharma monopolies:** Genetic data can drive up drug prices, making essential treatments inaccessible to lower-income or remote populations, limiting access to personalized, life-saving therapies.
- **Over-reliance on technology:** For drug prediction could lead to decisions based solely on algorithms, bypassing the importance of human judgment, patient preferences, and socio-environmental factors that are essential for comprehensive and personalized healthcare.

IPR Rules

- Patentability of AI-Generated Drug Compounds : The AI models can actually design new drug molecules, but patenting an AI-generated molecule is a developing legal area.
- Licensing & Commercialization of AI Models : Companies must license third-party genetic datasets for research and ensure that data-sharing policies are complied with.
- Software & Algorithm Protection : Companies may choose patent protection for novel AI-driven methodologies or rely on proprietary data and algorithms as a competitive advantage.



Patents

US20190164632A1: Predicting drug indications and responses using an artificial intelligence deep learning model that integrates various data types.

US11447830B2: Predict drug responses for cancer, by analysing tissue samples to determine the presence of altered gene expression associated with drug response.

US20190214136A1: Patient-specific drug-response predictions based on gene-expression and drug-response data.

References

Zhengxiang Jiang, Pengyong Li, DeepDR: a deep learning library for drug response prediction, Bioinformatics, Volume 40, Issue 12, December 2024, btae688

Sangeeta. Pharmacogenomics: Personalized medicine and drug response prediction. Pharma Innovation 2019;8(1):845-848.

S. Sotudian and I. C. Paschalidis, "Machine Learning for Pharmacogenomics and Personalized Medicine: A Ranking Model for Drug Sensitivity Prediction," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 4, pp. 2324-2333, 1 July-Aug. 2022, doi: 10.1109/TCBB.2021.3084562. keywords: {Drugs;Sensitivity;Cancer;Genomics;Geneexpression;Bioinformatics;Training;Drug sensitivity prediction;personalized medicine;elastic net regression;cancer;ranking;score function},

THANK YOU

