Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

PHARMACOGENOMIC ARTIFICIAL INTELLIGENCE BEYOND ANCESTRY LABELS THROUGH POPULATION-AWARE REPRESENTATION, TRANSPORTABILITY, AND UNCERTAINTY

Modibo Traoré[1]*, Awa Coulibaly[2], Souleymane Dembele[3], Boubacar Diallo[4]

1. ***Department of Pharmacogenomic AI and Population Representation, Faculty of Pharmacy, University of Bamako, Bamako, Mali.***
2. ***Department of Beyond-Ancestry Modeling and Transportability, Faculty of Pharmacy, University of Ségou, Ségou, Mali.***
3. ***Department of Uncertainty in Pharmacogenomics, Faculty of Pharmacy, University of Kayes, Kayes, Mali.***
4. ***Department of Population-Aware Drug Response Prediction, Faculty of Pharmacy, University of Sikasso, Sikasso, Mali.***

ARTICLE INFO

Received:
29 January 2025
Received in revised form:
02 April 2025
Accepted:
09 April 2025
Available online:
28 April 2025

Keywords: Pharmacogenomics, Population-aware representation, Genetic ancestry, Treatment-response prediction, Transportability, Calibration

ABSTRACT

Pharmacogenomic artificial intelligence seeks to connect molecular variation with treatment selection, dosing, efficacy, and toxicity, yet its development commonly depends on ancestry categories that compress heterogeneous genetic, environmental, clinical, and social information into a small number of labels. Such labels may describe aspects of study composition or support inequity auditing, but they are inadequate stand-alone representations of pharmacogene haplotypes, structural variation, linkage disequilibrium, admixture, environmental exposure, treatment context, or individual source-to-target similarity. This article develops a proposed population-aware pharmacogenomic representation–transportability–uncertainty framework for organizing these non-equivalent determinants without treating ancestry as a biological shortcut. The framework separates molecular pharmacogenomic representation, population-genetic structure, health and environmental context, treatment and endpoint definition, transportability assessment, calibration, uncertainty characterization, and equity governance. It argues that equitable treatment prediction cannot be established through a single discrimination metric, pooled performance estimate, or subgroup comparison. Instead, claims of usefulness should be conditional on representation adequacy, target-population support, external validation, probability calibration, uncertainty reliability, subgroup harm assessment, and clearly bounded decision use. Ancestry, race, and ethnicity may remain relevant as transparently defined contextual variables, but they should not substitute for measured biological or social determinants. The proposed framework is conceptual rather than empirically validated and does not establish clinical utility, causal treatment effects, regulatory acceptability, or deployment readiness. Its principal implication is that pharmacogenomic artificial intelligence should be evaluated as a context-dependent pharmaceutical evidence system whose reliability depends on what is represented, where predictions are transferred, how uncertainty is communicated, and whether errors and benefits are distributed equitably.

To Cite This Article: Traoré M, Coulibaly A, Dembele S, Diallo B. Pharmacogenomic Artificial Intelligence beyond Ancestry Labels through Population-Aware Representation, Transportability, and Uncertainty. Pharmacophore. 2025;16(2):65-75. https://doi.org/10.51847/3I8xrO74vf

Introduction

Pharmacogenomics examines how inherited variation contributes to differences in drug exposure, efficacy, toxicity, and treatment response, with the broader aim of improving pharmaceutical decisions for individual patients [1]. Artificial intelligence and machine-learning methods may expand the range of genomic, clinical, and treatment variables that can be considered simultaneously, but their potential usefulness depends on the quality and relevance of the evidence from which they learn. Existing pharmacogenomic knowledge is not distributed evenly across populations, sexes, drugs, genes, phenotypes, or health systems, and this imbalance can shape which molecular associations are discovered, which variants are represented in assays, and which patients are well supported by predictive models [2]. The resulting problem is therefore not

Corresponding Author: Modibo Traoré; Department of Pharmacogenomic AI and Population Representation, Faculty of Pharmacy, University of Bamako, Bamako, Mali. E-mail: modibo.traore@ustb.edu.ml.

merely one of algorithm selection. It is a problem of evidence representation, population coverage, biological resolution, contextual relevance, and decision accountability.

The population composition of human genetic research remains concentrated in a limited subset of global populations, leaving substantial variation undercharacterized and constraining the interpretation of findings in groups that are poorly represented [3]. Cross-population studies of genetic prediction further indicate that models developed in highly represented populations may lose performance elsewhere and can reproduce or intensify disparities when portability is assumed rather than demonstrated [4]. These concerns are especially consequential in pharmacogenomics because drug response may depend on uncommon variants, multiallelic haplotypes, copy-number changes, local ancestry, linkage structure, co-medication, organ function, disease state, dose, adherence, and environmental exposure. A broad population label cannot reliably encode this combination of molecular and contextual determinants for an individual patient.

Current approaches frequently handle population variation in one of two unsatisfactory ways. Some models ignore population structure and pool all observations as though the relation between features and treatment outcomes were invariant across settings. Others add a categorical ancestry, race, ethnicity, or geographic variable and implicitly treat it as a sufficient correction for population heterogeneity. The first strategy risks hidden distribution shift, confounding, and underrepresentation. The second risks converting an administrative or socially mediated descriptor into a biological proxy for unmeasured pharmacogenomic variation. Both strategies may produce apparently strong aggregate performance while leaving unanswered whether the model represents the relevant molecular determinants, whether its probabilities remain calibrated in a target group, or whether its errors are concentrated among patients with limited representation.

This article proposes a Population-Aware Pharmacogenomic Representation–Transportability–Uncertainty framework, hereafter termed PA-PGx-RTU, as an original conceptual architecture for equity-aware treatment prediction. PA-PGx-RTU separates five functions that are often collapsed: representing pharmacogenomic variation, characterizing population-genetic and health context, constructing treatment-specific predictions, evaluating source-to-target transportability, and governing uncertainty and equity. The framework does not reject all uses of ancestry, race, or ethnicity variables. Rather, it restricts their role to explicitly defined descriptive, contextual, stratification, or inequity-auditing purposes and requires measured molecular and contextual variables whenever those variables are the actual scientific determinants of interest. The central argument is that equitable pharmacogenomic artificial intelligence cannot be established through a single accuracy measure. It requires an evidence chain linking representation adequacy, transportability, calibration, uncertainty, subgroup consequences, and bounded pharmaceutical decision use.

Why Ancestry Labels are Inadequate Proxies for Pharmacogenomic Variation

Ancestry is not a fixed set of continental biological categories. It describes genealogical relationships that depend on the time depth, reference populations, sampling frame, and genomic scale used to define them [5]. Two individuals assigned the same ancestry label may differ substantially in haplotype composition, local ancestry, allele frequency, admixture history, and genetic distance from a model's development data. Conversely, individuals assigned different categories may share relevant pharmacogene variants or genomic segments. Scientific representations that convert ancestry into a small number of mutually exclusive groups therefore discard continuity, overlap, and locus-specific variation. More appropriate population-genetic descriptions may include continuous genetic similarity, local ancestry, linkage disequilibrium, demographic history, and reference-panel distance, selected according to the pharmacogenomic question rather than inherited from administrative classification [6].

Race, ethnicity, and genetic ancestry also represent non-equivalent domains. Race and ethnicity may reflect social identification, historical classification, migration, discrimination, language, geography, or access to care, whereas genetic ancestry concerns patterns of biological relatedness inferred from genetic data. Treating these constructs as interchangeable can obscure the pathways through which health-system exposure, structural racism, economic conditions, environmental factors, and clinical access influence treatment and outcomes [7]. In pharmacogenomic artificial intelligence, the risk is especially acute when a social category becomes a convenient surrogate for unmeasured molecular variation or when genetic ancestry is used to dismiss social and environmental determinants. PA-PGx-RTU therefore requires parallel rather than substitutive representation: molecular variables should represent pharmacogenomic biology, while social and health-context variables should represent the conditions through which care, exposure, measurement, and treatment response are observed.

Population descriptors additionally vary in how they are collected and interpreted. A category may be self-identified, assigned by an investigator, extracted from a medical record, inferred from geography, or estimated from genetic data. These procedures do not generate equivalent variables, and the resulting labels may change across institutions, countries, time periods, and datasets. Recommendations from large genomic research programs emphasize that race, ethnicity, and ancestry variables should be explicitly defined, scientifically justified, harmonized cautiously, and reported with sufficient methodological detail to permit interpretation [8]. For pharmacogenomic AI, this means that a population variable should carry provenance: who supplied it, how it was measured, why it was included, what construct it is intended to represent, and which interpretation is prohibited. A label lacking such provenance should not be allowed to function as an unexamined model feature.

The inadequacy of ancestry labels is therefore not an argument for population-blind analysis. Removing every population descriptor while leaving sampling imbalance, linkage differences, assay gaps, and contextual heterogeneity unexamined may conceal rather than resolve inequity. The proposed PA-PGx-RTU framework instead distinguishes three roles. First, measured population-genetic structure may be required to understand genomic variation and model transfer. Second, transparently

defined race or ethnicity variables may be required to audit inequities produced by research, health systems, or deployment. Third, neither role justifies using a label as a stand-alone estimate of an individual's genotype, drug metabolism, treatment response, or clinical need. **Table 1** organizes the evidence, constructs, and interpretation boundaries needed to develop why ancestry labels are inadequate proxies for pharmacogenomic variation within the article's central argument.

Table 1. Evidence domains, core questions, scientific requirements, and interpretation boundaries for why ancestry labels are inadequate proxies for pharmacogenomic variation

Construct or evidence domain	Core question	Scientific basis	Role in the article	Failure or overclaiming risk	Boundary statement
Genetic ancestry	What genealogical relationship is being represented, at what temporal and genomic scale, and relative to which reference populations?	Ancestry is multidimensional, continuous, reference-dependent, and scale-dependent.	Replaces essentialized continental categories with explicit population-genetic constructs.	Treating inferred ancestry components as natural, fixed, or clinically homogeneous populations.	Genetic ancestry may describe relatedness but does not determine an individual drug response.
Race and ethnicity	Is the variable intended to represent social identity, structural exposure, discrimination, geography, language, or access to care?	Race and ethnicity are historically and socially produced classifications with context-specific meanings.	Supports inequity auditing and analysis of social or health-system pathways.	Recasting a social category as a biological mechanism or pharmacogene profile.	Race and ethnicity may reveal inequity but cannot substitute for measured molecular variation.
Global population structure	Does broad genetic similarity affect the relevance of the development dataset to the target population?	Allele frequencies and correlation structures may differ across populations because of demographic history.	Supports initial source–target comparison and sampling assessment.	Assuming that global similarity guarantees local pharmacogene similarity or model validity.	Global structure is only one dimension of transportability.
Local ancestry	Does ancestry at a specific genomic region differ from an individual's genome-wide average?	Admixed genomes may contain segments derived from different ancestral source populations.	Supports locus-specific interpretation of pharmacogene regions.	Treating local ancestry as a causal drug-response mechanism without identifying the functional variant or haplotype.	Local ancestry may refine association analysis but does not establish mechanism by itself.
Linkage disequilibrium	Are measured variants correlated with functional variants in the same way across source and target populations?	Demographic history changes correlation patterns between nearby variants.	Explains why an associated marker may not transfer across populations.	Assuming that the same tag variant captures the same functional signal everywhere.	Linkage similarity must be examined at the relevant locus and cannot be inferred from a label alone.
Pharmacogene alleles and haplotypes	Are the relevant star alleles, copy-number changes, structural variants, and functional states directly represented?	Drug-metabolism and transporter genes may contain complex multiallelic variation.	Prioritizes direct molecular representation over group-level approximation.	Predicting an individual genotype or phenotype from population frequency.	Population frequency informs evidence planning but not individual assignment.
Environment and health context	Are diet, co-medication, organ function, disease state, adherence, exposure, and care context measured?	Drug response reflects interacting genetic and non-genetic determinants.	Prevents ancestry variables from absorbing unmeasured clinical or environmental effects.	Attributing modifiable or socially mediated differences to inherited biology.	Contextual variables require direct measurement wherever feasible.
Descriptor provenance	How was each population descriptor collected, inferred, harmonized, and used?	Self-identification, investigator assignment, record extraction, geography, and genetic inference are non-equivalent.	Makes population variables auditable and interpretable.	Combining incompatible categories or leaving their meaning unstated.	A descriptor without provenance should not be treated as a reliable scientific feature.
Equity auditing	Are errors, access, uncertainty, and benefits distributed unequally across socially relevant groups?	Models may reproduce inequities even when social categories are not causal biological variables.	Retains justified group variables for monitoring consequences.	Concluding that metric parity alone establishes equity.	Auditing categories identify patterns of harm but do not explain those patterns without further analysis.

Population Representation, Linkage Structure, Environment, and Health Context

Population representation concerns more than whether multiple demographic categories appear in a dataset. It includes whether the development evidence captures the pharmacogene alleles, haplotypes, structural variants, treatment exposures, clinical states, environments, and health-system conditions relevant to the intended population. A genomically diverse sample may still be unsuitable when phenotypes are inconsistently measured, treatment histories are incomplete, assays omit consequential variation, or certain groups contribute too few outcomes for reliable calibration. Expanding representation therefore requires coordinated changes in recruitment, infrastructure, scientific leadership, analytical capacity, data governance, and benefit sharing rather than the addition of small numbers of underrepresented participants to an otherwise unchanged research system

[9]. In PA-PGx-RTU, representation adequacy is consequently defined as decision-specific coverage of relevant variation, not as demographic diversity reported in isolation.

Linkage structure provides a mechanistic reason why predictive relationships may fail to transfer even when the same variant is measured in two populations. A marker associated with a phenotype in one development cohort may act primarily as a proxy for a nearby functional variant. If the correlation between the marker and the functional variant differs in the target population, the learned association can weaken, reverse, or disappear. Demographic history influences allele frequencies, haplotype structure, linkage disequilibrium, and the extent to which genotyped markers capture causal or functional variation [10]. For pharmacogenomic AI, these effects are particularly important when models rely on sparse genotyping arrays, imputed variants, summary features, or common markers rather than directly resolved pharmacogene haplotypes. Population-aware representation should therefore ask whether the same molecular feature has the same functional meaning and correlation structure across the source and target contexts.

Population heterogeneity also persists within broad ancestry categories. Genetic-prediction accuracy has been shown to vary among individuals placed in the same ancestry group, with differences associated with factors such as age, sex, socioeconomic conditions, cohort composition, and the relationship between the individual and the training data [11]. Analyses of polygenic-score use similarly show that training-set composition, allele-frequency differences, linkage disequilibrium, and analytical choices contribute to unequal performance across populations [12]. These findings should not be transferred uncritically from disease-risk prediction to pharmacogenomic treatment prediction, but they establish a relevant methodological warning: a model cannot be assumed to be transportable merely because source and target participants share a nominal population label. PA-PGx-RTU therefore treats population membership as an insufficient summary of individual support and requires explicit evaluation of molecular, clinical, environmental, and data-generating similarity.

Direct pharmacogenomic evidence further demonstrates that clinically relevant cytochrome P450 alleles are distributed heterogeneously across world populations [13]. Such frequency variation can influence which alleles are observed during model development, which haplotypes are represented adequately, and which functional states remain uncertain. Yet a population frequency cannot identify an individual genotype, and a genotype alone cannot determine treatment response without considering phenotype definition, dose, co-medication, organ function, adherence, disease state, and other determinants of exposure and effect. The proposed framework therefore represents population-genetic structure and health context as complementary but non-interchangeable layers. Linkage structure informs whether genomic associations are likely to transfer; environmental and clinical context informs whether exposure–response relationships are comparable; and direct molecular measurement remains preferable to inference from group membership. Population-aware pharmacogenomic AI begins not with a better ancestry label, but with a more explicit account of what biological, contextual, and treatment information the label fails to contain.

Population-Aware Representations for Treatment Prediction

Population-aware treatment prediction requires more than appending a demographic variable to an otherwise population-blind model. Drug-response machine learning is already challenged by heterogeneous molecular measurements, limited sample sizes, inconsistent endpoints, treatment confounding, incomplete validation, and uncertain generalization across experimental and clinical settings [14]. In pharmacogenomics, these problems are compounded when common single-nucleotide variants are used as substitutes for unresolved haplotypes, copy-number changes, structural variants, or functional phenotypes. The proposed PA-PGx-RTU framework therefore begins with an explicit representation question: which biological and contextual variables are necessary to define the treatment-response problem, and which available variables are merely imperfect proxies? A population-aware representation should preserve the distinction among molecular evidence, population-genetic structure, clinical state, environmental exposure, treatment specification, and data provenance rather than compressing them into one composite ancestry feature.

A clinically oriented representation should also specify the pharmaceutical decision being supported. An example from early rheumatoid arthritis shows that pharmacogenomic machine learning can combine genomic and clinical variables to predict methotrexate response within a defined disease, treatment, endpoint, and study population [15]. Such work demonstrates the potential value of multimodal prediction but also illustrates its boundary conditions: a model developed for one treatment, outcome definition, ancestry composition, and clinical pathway cannot be assumed to support another drug, disease, or population. In PA-PGx-RTU, every prediction task is therefore anchored to a treatment-and-endpoint representation containing the drug, dose or exposure definition, indication, comparator where relevant, timing, adherence assumptions, efficacy or toxicity endpoint, and intended decision. This layer prevents a generic “drug response” label from concealing materially different pharmaceutical questions.

Molecular inputs should be structured according to their biological meaning and evidence status. Functional regulatory annotation may improve the portability of genetic prediction by prioritizing variants more likely to reflect biologically relevant effects rather than population-specific tagging relationships [16]. In a directly pharmacogenomic context, local-ancestry-informed analysis of stable warfarin dose in Latino populations demonstrates that locus-specific ancestry may reveal information not captured adequately by genome-wide categories [17]. These findings do not establish that functional annotation or local ancestry will improve every treatment model. They instead support a representation principle: population-aware pharmacogenomic AI should distinguish directly measured functional variants, inferred haplotypes, local genomic context, annotation-supported candidates, and statistical proxies. The model should also preserve uncertainty about phasing,

functional assignment, assay sensitivity, and reference-panel suitability rather than treating all encoded features as equally reliable.

Population-aware integration should permit information sharing across groups without erasing target-population evidence. Multi-population genetic prediction methods show that population-specific effects can be modeled jointly, allowing information to be borrowed across datasets while retaining differences in effect estimates and uncertainty [18]. PA-PGx-RTU extends this logic conceptually to pharmacogenomic treatment prediction through conditional integration. Shared biological knowledge may inform multiple populations, but population-specific allele frequencies, haplotypes, linkage patterns, treatment practices, and outcome distributions remain visible to the model and evaluator. The proposed architecture therefore rejects both indiscriminate pooling and complete population separation. Its intended function is to determine what information can be shared, what requires target-specific estimation, and where insufficient support should produce wider uncertainty, adaptation, human review, or abstention rather than a forced prediction. **Table 2** organizes the evidence, constructs, and interpretation boundaries needed to develop population-aware representations for treatment prediction within the article's central argument.

Table 2. Components, relationships, uncertainties, and validation needs within Population-aware representations for treatment prediction

Component or layer	Inputs	Core function	Expected output	Uncertainty or limitation	Validation requirement
Molecular pharmacogenomic representation	Variants, phased haplotypes, star alleles, copy-number changes, structural variants, functional annotations, assay quality	Represents the molecular determinants plausibly related to exposure or response	Patient- or sample-level molecular feature set with evidence status	Genotyping error, phasing uncertainty, unresolved function, omitted rare or structural variation	Orthogonal assay confirmation, haplotype concordance, functional-evidence review, platform-specific validation
Population-genetic representation	Continuous genetic similarity, local ancestry, linkage disequilibrium, allele frequencies, admixture, demographic history, reference-panel distance	Characterizes genomic structure relevant to feature meaning and transfer	Locus-aware and individual-level population context	Reference-panel mismatch, unstable ancestry inference, overinterpretation of broad components	Validation across diverse and admixed populations using prespecified population-genetic assessments
Health and environmental context	Age, organ function, disease state, co-medications, diet, exposure, adherence, socioeconomic and health-system variables	Represents non-genetic determinants of drug exposure, observation, and response	Context profile linked to the treatment and outcome	Missingness, temporal variability, measurement error, unmeasured confounding	Longitudinal data-quality assessment and external replication in relevant care settings
Treatment and endpoint representation	Drug, dose, route, formulation, indication, comparator, timing, exposure measure, efficacy or toxicity endpoint	Defines the estimand and pharmaceutical decision	Explicit treatment-response target	Endpoint heterogeneity, exposure misclassification, treatment-confounding	Drug-specific and endpoint-specific validation with clinically meaningful definitions
Evidence provenance layer	Recruitment method, site, platform, phenotype source, descriptor collection, missingness, versioning, data lineage	Makes feature meaning and evidence origin auditable	Provenance record accompanying model inputs and outputs	Incomplete historical metadata and incompatible source definitions	Independent documentation audit and reproducibility assessment
Information-preserving integration core	Molecular, population, context, treatment, and provenance representations	Combines non-equivalent inputs without reducing them to one ancestry feature	Conditional treatment prediction with traceable contributing domains	Shortcut learning, overfitting, omitted interactions, unstable feature importance	External validation, ablation analysis, sensitivity testing, and comparison with transparent baselines
Population-specific adaptation layer	Source and target distributions, target outcomes, local calibration data, prior biological knowledge	Determines what can be transferred, recalibrated, re-estimated, or withheld	Adapted or explicitly non-transferable prediction state	Sparse target data, overadaptation, unstable target estimates	Prospective or temporally separated target-population validation
Uncertainty and abstention layer	Predictive distribution, model disagreement, target distance, data-quality indicators, consequence thresholds	Distinguishes supported predictions from cases requiring caution or non-prediction	Bounded prediction, uncertainty statement, referral, or abstention	Miscalibrated uncertainty and unequal abstention across groups	Coverage testing, selective-risk analysis, subgroup evaluation, and monitoring after use

Transportability, Calibration, and Uncertainty across Groups

Transportability asks whether a relationship learned in one source population remains informative for a specified target population, clinical setting, treatment, and decision. It should not be reduced to whether two datasets share the same ancestry category. Genetic-prediction accuracy can vary continuously with an individual's genetic distance from the development data, including among people assigned to the same broad population label [19]. Transfer principles developed for diverse genetic prediction further show that portability depends on interacting differences in allele frequency, linkage disequilibrium, effect

architecture, admixture, environmental context, phenotype definition, and clinical use [20]. PA-PGx-RTU therefore represents transportability as a multidimensional profile rather than a binary attribute. A model may be transferable for one drug, endpoint, assay, and subgroup but unsuitable for another, even within the same institution.

Calibration is a separate requirement from discrimination. A model may rank patients correctly while systematically overestimating or underestimating the probability of response or toxicity. Clinical prediction methodology identifies calibration as essential because decision makers need to know whether a predicted probability corresponds to observed outcome frequency in the target context [21]. For pharmacogenomic AI, calibration should be evaluated overall, across clinically relevant prediction ranges, and within sufficiently supported population and intersectional groups. Recalibration may be appropriate when the ranking relationship remains stable but the baseline outcome distribution changes. It is not an adequate remedy when molecular feature effects, treatment pathways, or data-generating mechanisms differ fundamentally between source and target settings.

Distribution shift may arise from changes in population composition, sequencing platform, phenotype measurement, prescribing practice, co-medication, disease severity, clinical workflow, or calendar time. Health-AI research shows that development-stage performance can deteriorate when the relationships generating inputs and outcomes change after deployment [22]. PA-PGx-RTU consequently requires shift diagnosis before model transfer and continued monitoring afterward. Covariate overlap alone is insufficient because two settings may have similar observed inputs but different causal relationships among treatment, biology, health-system processes, and outcomes. When the relevant relationship appears stable, external validation or recalibration may be justified. When it is uncertain or demonstrably different, model adaptation, new evidence generation, or rejection is more appropriate than automatic transfer.

Uncertainty should describe the limits of evidence rather than decorate a deterministic output. Medical machine-learning scholarship emphasizes that predictive uncertainty should be quantified, communicated, and linked to referral or abstention when a model lacks sufficient support [23]. PA-PGx-RTU distinguishes aleatoric uncertainty arising from irreducible outcome variability, epistemic uncertainty arising from limited knowledge or sparse representation, and distributional uncertainty arising when a target case lies outside the development evidence. These forms of uncertainty have different implications. Irreducible variability may require probabilistic communication, whereas epistemic or distributional uncertainty may justify additional testing, expert review, adaptation, or non-prediction. Uncertainty estimates must themselves be evaluated because numerical confidence can remain misleadingly high in poorly represented populations. **Figure 1** presents the population-aware pharmacogenomic lens, showing how the article's key components and boundaries are connected within transportability, calibration, and uncertainty across groups.

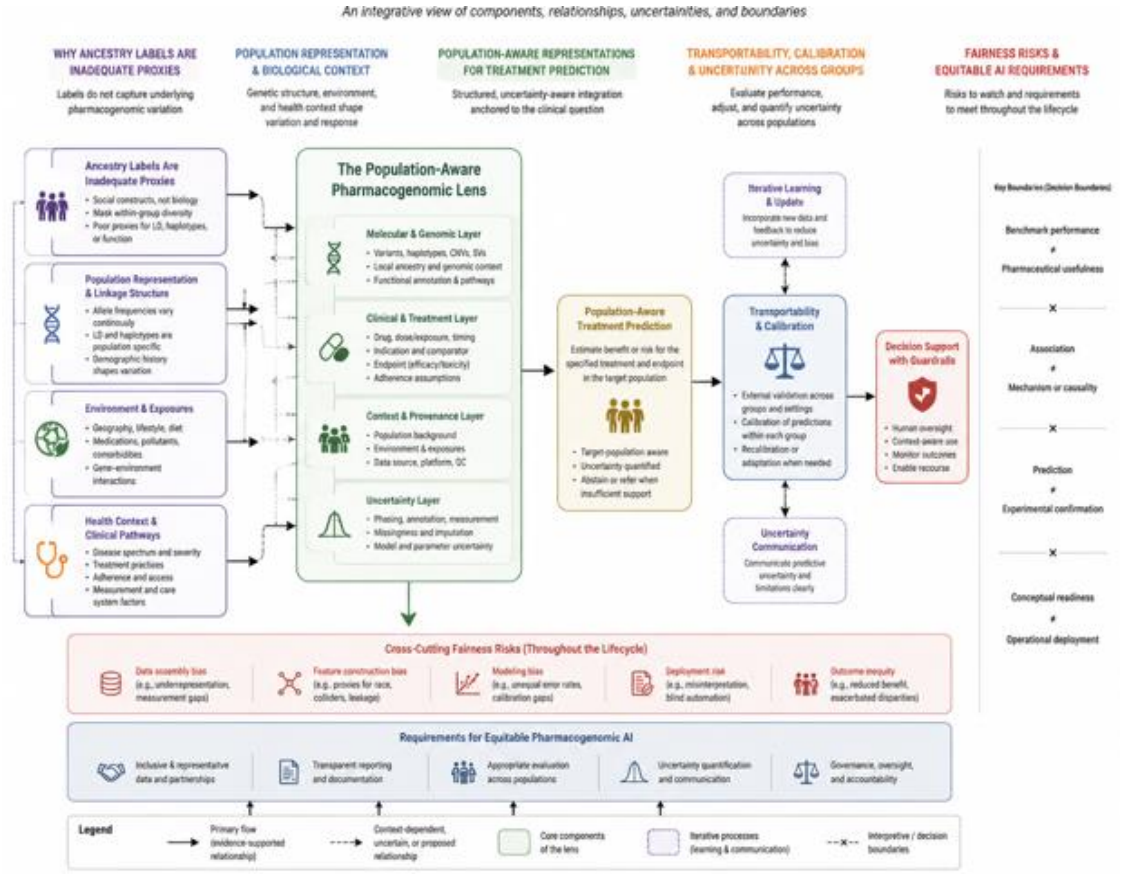

Figure 1. Population-Aware Pharmacogenomic Lens.

The figure is an original conceptual synthesis that organizes the article's central contribution across ancestry labels are inadequate proxies for pharmacogenomic variation, population representation, linkage structure, environment, and health context, population representation, linkage structure, health context, population-aware representations for treatment prediction. Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, regulatory determination, clinical recommendation, or deployment-ready system.

Fairness Risks in Data Assembly, Feature Construction, and Deployment

Fairness risks begin before model training. The variable selected as the prediction target may encode unequal access, treatment, documentation, or expenditure rather than the underlying health state of interest. A widely examined population-health algorithm produced racially unequal decisions because healthcare cost was used as a proxy for health need, even though spending reflected unequal access and patterns of care [24]. The pharmacogenomic analogue is not limited to monetary variables. Recorded treatment response, dose stability, adverse-event detection, medication adherence, and access to genetic testing may all be shaped by clinical systems. A model trained on these observations can reproduce health-system behavior while appearing to learn biological treatment response. PA-PGx-RTU therefore requires explicit justification of the prediction target and examination of the pathway connecting that target to the intended pharmaceutical construct.

Data assembly can introduce selection, measurement, documentation, and missingness biases. Electronic health records are generated through care rather than controlled observation, and their data reflect who accesses care, which tests are ordered, how diagnoses are coded, and which outcomes are recorded [25]. Populations with sparse, fragmented, or poor-quality digital data may consequently be excluded from model development or represented through systematically less reliable measurements. This condition has been described as health data poverty, which can limit access to data-driven benefit while exposing poorly represented groups to models developed elsewhere [26]. In pharmacogenomics, data poverty may include absent sequencing, incomplete structural-variant detection, weak phenotype linkage, limited treatment follow-up, and insufficient local functional evidence. Data availability should therefore never be equated with data suitability.

Feature construction creates additional risks. An ancestry label may become a high-level shortcut that absorbs unmeasured variation in genotype, environment, access, adherence, or clinical practice. Removing the label does not necessarily eliminate the problem because other variables may reconstruct it indirectly, and their predictive role may still reflect structural inequity. Reviews of algorithmic fairness in medicine show that bias can arise across data acquisition, labeling, representation, model development, validation, and deployment rather than at one isolated stage [27]. PA-PGx-RTU accordingly treats fairness as a lifecycle property. Feature review should ask whether each variable represents the intended construct, whether a less ambiguous direct measurement is available, whether the feature's meaning changes across settings, and whether reliance on that feature could generate clinically unequal errors.

Deployment may reveal harms that average performance conceals. Empirical evaluation of medical imaging algorithms has shown that apparently adequate overall performance can coexist with systematic underdiagnosis in underserved populations and intersectional subgroups [28]. The same concern applies to pharmacogenomic AI when false reassurance, missed toxicity, inappropriate dose recommendations, or excessive abstention are concentrated among patients with limited representation. Equity evaluation should therefore include subgroup calibration, error burden, uncertainty reliability, access to confirmatory testing, and the consequences of both prediction and non-prediction. Metric parity alone is insufficient because equal error rates can still produce unequal harm when baseline risks, treatment alternatives, monitoring resources, or consequences differ.

Requirements for Equitable Pharmacogenomic AI

Equitable pharmacogenomic AI requires locally relevant molecular evidence and scientific capacity. Pharmacogenomic research in Africa illustrates the importance of population-specific allele and haplotype knowledge, appropriate testing platforms, laboratory infrastructure, trained personnel, interoperable data resources, and health-system feasibility [29]. These requirements apply more broadly wherever populations are underrepresented or existing assays and knowledge bases incompletely capture local variation. Equity cannot be achieved by transporting a model into a new population and adding a subgroup-performance table. The evidence-generating system itself must support discovery, functional interpretation, assay development, validation, clinical pharmacology expertise, and continued stewardship in the populations expected to use or be affected by the model.

Responsible development also requires lifecycle accountability. Guidance for health machine learning emphasizes the importance of stakeholder participation, clinically meaningful problem formulation, appropriate data, transparent validation, workflow integration, monitoring, and mechanisms for responding to harm [30]. In PA-PGx-RTU, responsibility is distributed across data contributors, genomic laboratories, bioinformaticians, model developers, pharmacologists, clinicians, pharmacists, health systems, patients, and affected communities. Human oversight should not be reduced to a final approval step. It should shape the treatment question, determine acceptable errors, review population descriptors, define abstention pathways, interpret uncertainty, and decide whether a model's evidence is sufficient for a particular use.

Technical validation remains necessary but cannot establish pharmaceutical usefulness by itself. Clinical-impact analyses of artificial intelligence identify external validation, usability, workflow compatibility, prospective evaluation, and comparison with meaningful standards of care as critical requirements beyond internal benchmark performance [31]. For pharmacogenomic AI, this means that a model with strong retrospective discrimination still cannot be described as clinically useful unless it has been evaluated for the intended drug, endpoint, population, platform, care pathway, and decision

consequence. Analytical validity of the genomic input, calibration of the prediction, actionability of the output, availability of treatment alternatives, and clinician and patient understanding are separate evidentiary requirements.

Equity should be evaluated in relation to health consequences rather than reduced to one mathematical parity condition. Fairness in healthcare machine learning depends on the clinical objective, the distribution of benefit and harm, and the historical processes embedded in the data and decision pathway [32]. Broad claims of generalisability are likewise inappropriate when usefulness is conditional on a specified population, setting, time, treatment, and workflow [33]. PA-PGx-RTU therefore proposes that every model claim be paired with a bounded use statement specifying who is represented, which treatment decision is supported, where validation occurred, how calibration and uncertainty were assessed, which groups remain poorly supported, and what action follows when the evidence boundary is crossed. **Table 3** organizes the evidence, constructs, and interpretation boundaries needed to develop requirements for equitable pharmacogenomic ai within the article's central argument.

Table 3. Evaluation, implementation, and research priorities arising from Pharmacogenomic Artificial Intelligence beyond Ancestry Labels through Population-Aware Representation, Transportability, and Uncertainty

Evaluation dimension	What must be tested	Suitable evidence or assessment	Failure signal	Interpretive limitation	Decision relevance
Population-representation adequacy	Whether development data cover the relevant molecular, clinical, environmental, treatment, and health-system variation	Coverage analysis; haplotype and structural-variant representation; missingness assessment; source–target comparison	Important target-population variants, contexts, or outcomes are absent or too sparse for evaluation	Demographic counts alone do not establish scientific representativeness	Determines whether model development, adaptation, or new evidence generation is justified
Molecular analytical validity	Whether pharmacogenomic inputs are measured and interpreted accurately	Orthogonal genotyping; sequencing concordance; phasing assessment; star-allele and copy-number validation	Discordant genotype or phenotype assignment, unresolved structural variation, platform-specific failure	Analytical validity does not establish clinical utility	Determines whether molecular inputs are reliable enough to enter prediction
Treatment and endpoint validity	Whether the model predicts a clearly defined and clinically meaningful pharmaceutical outcome	Drug-, dose-, indication-, exposure-, and endpoint-specific review; phenotype adjudication	Ambiguous "response" labels, treatment-confounding, inconsistent follow-up	Observational outcome prediction does not automatically estimate causal treatment benefit	Determines what decision the output can legitimately support
Transportability	Whether relationships remain informative in the intended target population and setting	External multicenter validation; genetic and contextual distance analysis; temporal and platform testing	Performance or feature relationships deteriorate after source-to-target transfer	Validation in one external cohort does not establish universal transfer	Determines direct use, recalibration, adaptation, or rejection
Calibration	Whether predicted probabilities or response magnitudes correspond to observed target outcomes	Calibration plots; calibration intercept and slope; expected-versus-observed assessment; proper scoring rules	Systematic overprediction or underprediction overall or within groups	Stable estimates require sufficient observations across the prediction range	Determines whether numerical predictions can inform threshold-based decisions
Uncertainty reliability	Whether intervals, confidence states, or abstention rules correspond to actual error	Coverage assessment; selective-risk analysis; out-of-distribution stress testing; subgroup uncertainty evaluation	High confidence persists in unsupported cases or uncertainty is systematically wider for one group without remediation	Uncertainty estimates can themselves be miscalibrated	Determines communication, additional testing, expert review, or abstention
Subgroup and intersectional harm	Whether errors and benefits are distributed equitably	Group-specific calibration; false-positive and false-negative burden; worst-group analysis; consequence-weighted evaluation	Aggregate performance conceals concentrated toxicity, non-response, or inappropriate treatment recommendations	Fairness metrics may conflict and require clinical interpretation	Determines whether use would worsen or reduce inequity
Human oversight and usability	Whether users understand model scope, uncertainty, and appropriate override behavior	Human-factors studies; comprehension testing; override review; escalation-path assessment	Automation bias, inappropriate reliance, unclear accountability, or inaccessible outputs	User satisfaction alone does not establish therapeutic benefit	Determines whether outputs can be integrated safely into clinical workflows
Clinical and pharmaceutical utility	Whether using the model improves a specified treatment decision or patient-relevant outcome	Prospective comparative evaluation; decision analysis; treatment and safety outcomes; workflow assessment	Technical performance fails to change decisions appropriately or causes unintended harm	Retrospective modeling cannot establish prospective benefit	Determines progression from research use toward bounded clinical evaluation

Monitoring and governance	Whether performance, calibration, representation, and harm remain acceptable over time	Drift detection; versioned audit trails; incident review; periodic subgroup reassessment	Unrecognized data drift, changing prescribing patterns, or accumulating subgroup harm	Monitoring cannot compensate for an invalid initial use case	Determines continuation, modification, suspension, or withdrawal

Limitations and Boundary Conditions

The proposed PA-PGx-RTU framework is a conceptual synthesis rather than an empirically validated architecture. Its components are supported by evidence from pharmacogenomics, population genetics, clinical prediction, drug-response machine learning, transportability research, and algorithmic fairness, but the complete framework has not been tested as one integrated system. Several supporting methods were developed for disease-risk prediction or general health AI rather than pharmacogenomic treatment response. Their relevance lies in the mechanisms they identify—such as linkage-dependent portability, calibration failure, dataset shift, proxy bias, and subgroup harm—not in proof that identical effects will occur for every drug, gene, or clinical setting. Future evaluation must therefore be drug-specific, endpoint-specific, population-specific, and decision-specific.

Population-aware representation also cannot eliminate all uncertainty or identify every relevant determinant of treatment response. Genetic ancestry, local ancestry, functional annotation, and linkage structure remain imperfect scientific constructs whose accuracy depends on reference data, measurement methods, demographic assumptions, and genomic resolution. Environmental, social, behavioral, and clinical variables are similarly incomplete and may change over time. A more detailed representation may reduce reliance on ancestry labels while simultaneously increasing data requirements, missingness, privacy concerns, computational complexity, and the risk of overfitting. More variables do not guarantee better science unless their measurement, meaning, temporal relation, and decision relevance are justified.

The framework does not establish causal treatment effects, mechanistic confirmation, dose recommendations, clinical benefit, regulatory acceptability, or operational readiness. A model may identify associations or predict outcomes without demonstrating that changing treatment on the basis of its output improves patient care. Equitable performance in one evaluation cannot guarantee equitable access, interpretation, or outcomes after implementation. Abstention can prevent unsupported prediction, but it may also withhold assistance disproportionately from populations already excluded from evidence generation. The framework must therefore be interpreted as a structure for defining questions, evaluating evidence, exposing uncertainty, and setting boundaries—not as a substitute for prospective pharmaceutical, clinical, ethical, or regulatory evaluation.

Research Agenda and Implementation Priorities

The first research priority is to develop population-complete pharmacogenomic evidence resources that resolve complex haplotypes, structural variants, copy-number variation, rare variants, functional effects, and treatment phenotypes across globally diverse and admixed populations. These resources should be connected to longitudinal information on dose, exposure, adherence, co-medication, organ function, disease state, environment, and care context. Methodological studies should compare categorical descriptors with continuous genetic distance, local ancestry, locus-specific linkage structure, and direct molecular measurements. The goal should not be to identify a single superior ancestry representation, but to determine which representation is scientifically necessary for each pharmacogenomic question and when a population variable adds information beyond the measured molecular and contextual data.

A second priority is to establish transportability and uncertainty evaluation designs tailored to treatment prediction. Multicohort studies should test models across populations, clinical sites, sequencing platforms, time periods, prescribing systems, and treatment pathways. Evaluations should report external discrimination, calibration, uncertainty coverage, selective prediction, subgroup error burden, and consequences of adaptation or abstention. Hybrid pharmacometric and machine-learning approaches warrant investigation because mechanistic knowledge of exposure, metabolism, dose, and response may constrain extrapolation more meaningfully than unconstrained statistical learning alone. Such studies should compare transparent baselines, population-blind models, label-adjusted models, and population-aware representations without assuming that greater architectural complexity produces better pharmaceutical decisions.

Implementation should proceed through staged evidence gates. Initial work should focus on analytical validity, representation adequacy, reproducibility, and retrospective external validation. Subsequent stages should examine human interpretation, workflow fit, target-population recalibration, subgroup harm, and prospective decision consequences. Deployment consideration should occur only when a clearly defined use case, reliable genomic input, clinically interpretable output, validated uncertainty behavior, oversight pathway, and monitoring system are present. Infrastructure investment, local scientific leadership, community participation, data governance, and benefit sharing should accompany technical development so that populations contributing data also shape the questions, interpretation, and distribution of benefits. When evidence remains inadequate, the correct output is not a more confident ancestry-based approximation but an explicit research priority, referral pathway, or decision not to predict.

Conclusion

Pharmacogenomic artificial intelligence cannot move beyond ancestry labels by simply deleting them, renaming them, or adding them as adjustment variables. It must replace their inappropriate proxy function with direct and scientifically justified

representations of pharmacogenomic variation, population-genetic structure, health and environmental context, treatment specification, and evidence provenance. The proposed PA-PGx-RTU framework organizes these inputs around transportability, calibration, uncertainty, equity, and bounded decision use. Its central contribution is to treat reliability as a property of an evidence relationship between a defined model, target population, pharmaceutical decision, and clinical context rather than as a single performance score. The framework remains conceptual and requires rigorous molecular, statistical, clinical, implementation, and equity evaluation. It does not establish causal treatment effects, clinical utility, regulatory acceptance, or deployment readiness. Its practical value lies in making unsupported assumptions visible and requiring pharmacogenomic predictions to state not only what they estimate, but for whom, under which conditions, with what uncertainty, and beyond which boundary they should not be used.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Roden DM, McLeod HL, Relling MV, Williams MS, Mensah GA, Peterson JF, et al. Pharmacogenomics. Lancet. 2019;394(10197):521-32. doi:10.1016/S0140-6736(19)31276-0
2. Corpas M, Siddiqui MK, Soremekun O, Mathur R, Gill D, Fatumo S. Addressing ancestry and sex bias in pharmacogenomics. Annu Rev Pharmacol Toxicol. 2024;64:53-64. doi:10.1146/annurev-pharmtox-030823-111731
3. Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. Cell. 2019;177(1):26-31. doi:10.1016/j.cell.2019.02.048
4. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51(4):584-91. doi:10.1038/s41588-019-0379-x
5. Mathieson I, Scally A. What is ancestry? PLoS Genet. 2020;16(3). doi:10.1371/journal.pgen.1008624
6. Lewis ACF, Molina SJ, Appelbaum PS, Dauda B, Di Rienzo A, Fuentes A, et al. Getting genetic ancestry right for science and society. Science. 2022;376(6590):250-2. doi:10.1126/science.abm7530
7. Borrell LN, Elhawary JR, Fuentes-Afflick E, Witonsky J, Bhakta N, Wu AHB, et al. Race and genetic ancestry in medicine—A time for reckoning with racism. N Engl J Med. 2021;384(5):474-80. doi:10.1056/NEJMms2029562
8. Khan AT, Gogarten SM, McHugh CP, Stilp AM, Sofer T, Bowers ML, et al. Recommendations on the use and reporting of race, ethnicity, and ancestry in genetic research: Experiences from the NHLBI TOPMed program. Cell Genom. 2022;2(8):100155. doi:10.1016/j.xgen.2022.100155
9. Fatumo S, Chikowore T, Choudhury A, Ayub M, Martin AR, Kuchenbaecker K. A roadmap to increase diversity in genomic studies. Nat Med. 2022;28(2):243-50. doi:10.1038/s41591-021-01672-4
10. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. Am J Hum Genet. 2017;100(4):635-49. doi:10.1016/j.ajhg.2017.03.004
11. Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. Elife. 2020;9. doi:10.7554/eLife.48376
12. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler KJ, Feldman MW, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun. 2019;10(1):3328. doi:10.1038/s41467-019-11112-0
13. Zhou Y, Ingelman-Sundberg M, Lauschke VM. Worldwide distribution of cytochrome P450 alleles: A meta-analysis of population-scale sequencing projects. Clin Pharmacol Ther. 2017;102(4):688-700. doi:10.1002/cpt.690
14. Adam G, Rampášek L, Safikhani Z, Smirnov P, Haibe-Kains B, Goldenberg A. Machine learning approaches to drug response prediction: Challenges and recent progress. NPJ Precis Oncol. 2020;4(1):19. doi:10.1038/s41698-020-0122-1
15. Myasoedova E, Athreya AP, Crowson CS, Davis JM 3rd, Warrington KJ, Walchak RC, et al. Toward individualized prediction of response to methotrexate in early rheumatoid arthritis: A pharmacogenomics-driven machine learning approach. Arthritis Care Res (Hoboken). 2022;74(6):879-88. doi:10.1002/acr.24834
16. Amariuta T, Ishigaki K, Sugishita H, Ohta T, Koido M, Dey KK, et al. Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. Nat Genet. 2020;52(12):1346-54. doi:10.1038/s41588-020-00740-8
17. Steiner HE, Carrion KC, Giles JB, Lima AR, Yee K, Sun X, et al. Local ancestry-informed candidate pathway analysis of warfarin stable dose in Latino populations. Clin Pharmacol Ther. 2023;113(3):680-91. doi:10.1002/cpt.2787
18. Ruan Y, Lin YF, Feng YCA, Chen CY, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. Nat Genet. 2022;54(5):573-80. doi:10.1038/s41588-022-01054-7

19. Ding Y, Hou K, Xu Z, Pimplaskar A, Petter E, Boulier K, et al. Polygenic scoring accuracy varies across the genetic ancestry continuum. Nature. 2023;618(7966):774-81. doi:10.1038/s41586-023-06079-4
20. Kachuri L, Chatterjee N, Hirbo J, Schaid DJ, Martin I, Kullo IJ, et al. Principles and methods for transferring polygenic risk scores across global populations. Nat Rev Genet. 2024;25(1):8-25. doi:10.1038/s41576-023-00637-2
21. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group Evaluating Diagnostic Tests and Prediction Models of the STRATOS Initiative. Calibration: The Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. doi:10.1186/s12916-019-1466-7
22. Subbaswamy A, Saria S. From development to deployment: Dataset shift, causality, and shift-stable models in health AI. Biostatistics. 2020;21(2):345-52. doi:10.1093/biostatistics/kxz041
23. Kompa B, Snoek J, Beam AL. Second opinion needed: Communicating uncertainty in medical machine learning. NPJ Digit Med. 2021;4(1):4. doi:10.1038/s41746-020-00367-3
24. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53. doi:10.1126/science.aax2342
25. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. JAMA Intern Med. 2018;178(11):1544-7. doi:10.1001/jamainternmed.2018.3763
26. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: An assailable barrier to equitable digital health care. Lancet Digit Health. 2021;3(4). doi:10.1016/S2589-7500(20)30317-4
27. Chen RJ, Wang JJ, Williamson DFK, Chen TY, Lipkova J, Lu MY, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng. 2023;7(6):719-42. doi:10.1038/s41551-023-01056-8
28. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med. 2021;27(12):2176-82. doi:10.1038/s41591-021-01595-0
29. Twesigomwe D, Mazhindu TA, Nagy M, Agesa G, Scholefield J, Masimirembwa C. Pharmacogenomics in Africa: A potential catalyst for precision medicine in genetically diverse populations. Annu Rev Genomics Hum Genet. 2025;26(1):321-49. doi:10.1146/annurev-genom-121323-104008
30. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: A roadmap for responsible machine learning for health care. Nat Med. 2019;25(9):1337-40. doi:10.1038/s41591-019-0548-6
31. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195. doi:10.1186/s12916-019-1426-2
32. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med. 2018;169(12):866-72. doi:10.7326/M18-1990
33. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. Lancet Digit Health. 2020;2(9). doi:10.1016/S2589-7500(20)30186-2