AI Model Card: PhenoGenius v3.1

Compliance Framework: ISO/IEC 42001:2023

1. General Information

1.1. Basic Information

AI System Name	PhenoGenius
Model Identifier	AIMC-02
Model Version	3.1
Release Date	29 April 2026 (SeqOne Platform v2.1)
Model Type	Hybrid AI System — Gene-Phenotype Association Matrix (Knowledge Matrix + score calculation)
Development Team	SSC — SeqOne Nicolas Duforet-Frebourg | Jiri Ruzicka | Alexandre Boulat
Maintenance Status	Active
Related Products	DiagAI Score, DiagAI Short List, DiagAI Smart Picks (SeqOne Platform)
EU AI Act Classification	High-risk AI system (Article 6 & Annex 3.5b) – Ai intended to be used as a safety component in the management and operation of critical digital infrastructure, or as medical device

1.2. Licensing and Access

License Type	Proprietary
Access Level	Restricted to authorized clinical and research users
Terms of Use	Subject to SeqOne Platform Terms of Service and Clinical Use Agreements

2. Intended Use & Constraints

2.1. Intended Purpose

PhenoGenius is an AI-powered clinical decision support system designed to assist healthcare professionals and researchers in identifying candidate genes associated with observed patient phenotypes. The system integrates phenotypic data with gene information to prioritize potentially pathogenic variants by computing phenotype-to-gene correlation scores based on HPO (Human Phenotype Ontology) terms.

Primary Functions:

- Phenotype-to-gene correlation quantification: computing match scores between patient HPO terms and gene-disease associations aggregated from 9 international databases
- Clinical variant prioritization based on patient HPO symptom profile, ranking candidate genes by phenotypic compatibility
- Feeding the DiagAI Score composite prioritization pipeline as the PhenoGenius score feature

2.2. Target Users

Primary Users:

Healthcare professionals responsible for routine clinical genomic analysis: geneticists, bioinformaticians, biologists, engineers, and NGS platform technicians.

Indirect Users:

- Patients receiving diagnostic information
- Healthcare systems managing patient care
- Regulatory authorities overseeing device performance

2.3. In-Scope Uses

- Rare disease diagnosis support using phenotype-driven variant analysis
- Variant prioritization in whole exome sequencing (WES) clinical cases
- Variant prioritization in whole genome sequencing (WGS) clinical cases
- Gene panel selection based on patient phenotypes

2.4. Out-of-Scope Uses

- Standalone diagnostic tool without clinical expert interpretation
- Direct-to-consumer genetic testing interpretation
- Prenatal screening without appropriate clinical genetics oversight
- Somatic variant analysis in oncology (unless specifically validated)
- Population screening programs
- Forensic applications
- Non-medical genetic ancestry analysis
- Automated clinical reporting without human validation

3. Model Architecture & Technical Specifications

3.1. Algorithm Overview

PhenoGenius is a genotype-phenotype matching system that prioritizes genes based on phenotypic information when HPO terms are available. The PhenoGenius score is calculated using a data matrix with genes as row indices and HPO phenotype terms as columns.

Model Family	Hybrid AI System (Knowledge Matrix + score calculation)
Core Structure	Aggregated gene-phenotype association matrix
Total Parameters	123,454,288 (each representing a quantified gene-phenotype association)
Input Format	List of HPO terms (structured phenotype descriptions)
Output Format	Ranked gene lists with phenotype match confidence scores
Modifications	Standard implementation — no architectural modifications applied
Implementation	Python 3.12 | pandas 2.2 | elasticsearch 6.8.23| obonet 1.0

Frameworks and Libraries:

Library	Version
Python	3.12
pandas	2.2
elastisearch	6.8.23
obonet	1.0

3.2. Architecture Components

The model relies on a data structure that is the sum of 9 component data structures. Each component is a dataframe quantifying the level of interaction between genes and phenotypes from a specific information source. Aggregating these 9 dataframes creates an ensemble model combining multiple evidence streams.

Component	Source Type	Description
1	Structured	HPO — Human Phenotype Ontology (hpo.jax.org)
2	Unstructured	OMIM — Online Catalog of Human Genes and Genetic Disorders (omim.org)
3	Unstructured	NCBI LitVar2 — extracted articles from scientific literature (ncbi.nlm.nih.gov/research/litvar2)
4	Unstructured	MedGen — NCBI medical genetics resource (ncbi.nlm.nih.gov/medgen)
5	Structured	PanelApp Australia (panelapp-aus.org)
6	Structured	Monarch Initiative database (monarchinitiative.org)
7	Structured	Orphanet — rare disease and orphan drug database (orphadata.org)
8	Structured	EBI Gene2Phenotype (ebi.ac.uk/gene2phenotype)
9	Structured	Omim clinical synopsis

ℹ️ Sources 2 (Omim) is unstructured — gene-phenotype associations are given in raw text associated with each gene. For this unstructured source of data, associations are inferred via Elasticsearch, returning the top 100 matching genes per HPO term.

ℹ️ Sources 3 (LitVar2) is unstructured — gene-phenotype associations are inferred via Elasticsearch, returning the top 100 matching genes per HPO term.

3.3. Data Processing Method

Association Collection:

For each data source, a list of gene-phenotype associations is built. For unstructured sources (e.g. scientific articles), associations are inferred using Elasticsearch. Results are limited to the best 100 matching genes per HPO term.

Ontology-Based Association Enrichment:

To account for imprecision in phenotyping and the potential matching of related rather than exact terms, associations are enriched. If an HPO term is associated with a gene, all HPO terms at an ontological distance <= 2 are also associated with that gene. The coefficient value is inversely proportional to the ontological distance.

Phenotype-wise Normalization:

To down-weight non-informative phenotypes correlated with many genes, column-wise normalization is applied. For each phenotype column, the average coefficient is calculated. Each value is multiplied by (1 - column average), reducing the influence of globally frequent phenotypes.

Multi-source Aggregation:

Once a dataframe is produced per data source, the final matrix is the average of all component dataframes across all possible gene-phenotype associations. Unknown associations in any single source are set to 0 before averaging.

3.4. Data Preprocessing & Feature Engineering

Feature Engineering:

- For each data structure built on an unstructured data source, the associations retained are the top 100 associated genes for one phenotype
- Phenotype semantic similarity metrics computed to quantify closeness between HPO terms
- Gene-phenotype association strength indicators calculated per source and normalised
- Synonym expansion for phenotypes — related terms included within an ontological distance of 2

3.5. Key Hyperparameters

Ontology distance for neighbor enrichment	2 (HPO graph hops)
Max genes per HPO term (unstructured sources)	100 (top Elasticsearch results)
Similarity step decay threshold	0.5
Weighting factors per source	1

4. Training Data

4.1. Dataset Overview

Total Gene-Phenotype Associations	4,272 million (combined across all 9 sources) DS04, DS05, DS06, DS07
Annotation/Labeling	No manual annotation — raw data aggregated directly from public sources
Geographic Coverage	International multi-centre data (all 9 international databases)
Data Types	Structured HPO terms, gene-disease associations, scientific literature, gene panel data

4.2. Data Sources

All data sourced from publicly available international biomedical databases (see Section 3.2 for full list). No proprietary clinical case data was used for building the gene-phenotype association matrix — only public knowledge bases. Clinical cases were used exclusively for performance evaluation (validation set).

4.3. Data Governance

Privacy Framework	All training data sourced from public databases — no patient PHI involved in model training
Data Access	Public databases accessed under their respective terms; restricted to authorized SeqOne staff for operational use
Pseudonymization	Clinical validation cases fully pseudonymized; no PHI stored in model
Bias Monitoring	HPO resource reviewed to monitor the appearance of new hpo terms.

4.4. Known Training Data Limitations

- Acknowledged overrepresentation of European ancestry in genetic databases — may reduce performance for non-European populations
- Potential bias toward well-described 'classical' phenotypes; atypical presentations may score lower
- Publication bias: training data weighted toward published, positive findings — genes with fewer publications may have weaker scores
- Demographic data (ancestry distribution, age ranges, sex) not systematically recorded across all sources
- Novel HPO terms not yet integrated in the model may reduce accuracy for recently described phenotypes

5. Performance Metrics

5.1. Evaluation Methodology

Validation Set	1701 independent held-out clinical cases with confirmed genetic diagnoses from 4 independent datasets. DS08
Primary Metric	Gene recall performance — rank of causal gene among all prioritized candidates based on patient HPO profile
Methodology	Recall performance on gene ranking based on phenotypic profile of each patient
Overfitting	No statistical evidence of overfitting; performance stable across validation folds
Disease Coverage	485 different Orphanet disease categories represented in the validation dataset

5.2. Primary Performance Metrics

Metric	Size	Median rank	top1	top10	top100	top200	AUC
Validation set 1	316	96	6%	21	50	61	306
Validation set 2	304	67	4%	20	54	67	322
Validation set 3	140	27.5	6%	30	71	79	390
Validation set 4	941	19	16%	43	72	79	386

5.3. Known Performance Limitations

- Variant presence missing information. The ranking based solely on symptoms are usually not good at finding diagnostic genes among the first candidates. This is because a large number of genes can match a set of a few symptoms describing patient pathology. However, combined with the genetic information of presence or not of a variant in the genes, the ranking becomes better.
- Phenotype Specificity: Performance depends heavily on quality and specificity of HPO term selection
- Novel Genes: Reduced accuracy for genes without established disease associations in any of the 9 source databases
- Intergenic Variants: Limited performance for variants outside well-characterized gene regions
- Ultra-rare Diseases: Lower accuracy for diseases with fewer than 5 reported cases in the literature
- Phenocopy Scenarios: Cannot distinguish genetic from environmental causes presenting with identical phenotypes
- Novel HPO Terms: Reduced accuracy for patients with HPO terms not yet integrated in the current model version
- Oligogenic/Polygenic Cases: Limited performance for cases requiring multiple genetic hits — system optimized for monogenic disease

6. AI Risk Assessment & Impact

6.1. Identified Risks

ID	Hazard Description	Consequence / Harm	P	S	Risk Level	Mitigation Strategy	Residual Risk
RSK-76	Explainability Transparency - User interface is not understood / Misunderstanding of scores, rankings, or evidence. - AI-powered "black box" logic provides clinical scores without clear biological reasoning.	Incorrect results - Misinterpretation of data	2	2	Low	Mitigated by comprehensive training and clear UI design (in-product tooltips and contextual help on each score component; user manual section dedicated to the interpretation of evidence and rankings.) The system shall display a DiagAI score summary, with the different score sub-components : UP2, Phenogenius, Family Pattern, Inheritance compatibility, Quality Checks, PanelApp. (SR-616)	Low
RSK-81	Bias in predictions / ancestry disparities Reduced performance for patients from ancestry groups underrepresented in training data Bias in predictions: underperformance on under-represented genomic populations or variant types	Disparate performance across populations; health equity concerns; potential missed diagnoses in certain groups	3	2	Medium	Training data limitation documented; performance stratified by ancestry where possible; user guidance provided Fairness subgroup evaluation ; ethics review prior to deployment ; Direct ancestry-based fairness evaluation is not feasible due to absence of ancestry metadata in raw data source (ClinVar) Addition of new datasets to improve the AI model	Low
RSK-84	Over-reliance on AI / clinical misinterpretation Predictions used without appropriate clinical context or expert review	Human fails to catch AI errors — incorrect decisions acted upon uncritically; missed pathogenic variant; erroneous clinical decision; potential patient harm	3	2	Medium	Mandatory human-in-the-loop validation; clear UI warnings; system labeled as decision support; predictions accompanied by confidence scores	Low
RSK-86	"False negative resutls such as : - pathogenic variant scored below threshold, as Benign - miss an article about the variant of interest"	Incorrect results — misinterpretation of data; missed pathogenic variant; delayed or incorrect diagnosis; potential patient harm	2	2	Low	Mandatory human review before clinical reporting; Score is explainable to users (SR-252 to SR-257). Module retraining according to users false negative report. Thresholds are calibrated and displayed via tooltip.	Low
RSK-87	False positive results such as: - benign variant scored incorrectly ranked highly, above threshold (ay lead to incorrect pathogenic classification) - suggestion of an article that is not relevant for the user	Incorrect pathogenic classification; unnecessary downstream investigation; wasted clinical resources; erosion of user trust	3	2	Medium	Mandatory human review before clinical reporting; Score is explainabled to users (SR-252 to SR-257). Module retraining according to users false positive report. Thresholds are calibrated and displayed via tooltip.	Low

ID	Hazard Description	Consequence / Harm	P	S	Risk Level	Mitigation Strategy	Residual Risk
RSK-88	Data Management & Data Quality Reference database error propagated into DiagAI ClinVar, gnomAD, COSMIC or other reference databases used by DiagAI may contain erroneous, outdated, or conflicting variant annotations. These errors propagate into DiagAI pathogenicity classifications without the model detecting the source error, potentially causing systematic misclassification of variants across all patients.	Incorrect results - Misinterpretation of data	3	2	Medium	Independent validation step before database release to production.	Low
RSK-89	DiagAI model update regression in production A new version of DiagAI passes pre-prod validation but introduces performance regression in production (e.g., due to dataset shift between the validation cohort and live patient population, edge cases not covered in pre-prod, or interaction with updated reference databases). Regression may not be immediately detectable, delaying rollback and exposing patients to degraded diagnostic support.	Incorrect results - Misinterpretation of data	3	2	Medium	Implement automated performance regression testing comparing new model vs. current production on a fixed holdout set before any production release.	Low
RSK-91	Model Drift / Knowledge Base Obsolescence Performance degradation as new genes, diseases, and variants are discovered after model training, and as genomic databases and evidence evolve post-training	Reduced accuracy; outdated pathogenicity predictions; missed clinically relevant variants	2	2	Low	Major knowledge base updates in our bioresources. Some DB (ClinVar, OMIM) are also plugged to directly use the latest information from the pipeline, minimizing the risk of missing an important annotation. Models are retrained at least annually	Low
RSK-93	Phenotype Description Bias / Performance varies with clinical documentation quality or language	Inconsistent results across institutions	2	2	Low	User training on HPO term selection ; guided phenotype entry interface Additionnaly, DiagAI HPO extracts hpo terms by including "lay person terms" from hpo.jax Interface displayed best matching HPO terms.	Low
RSK-94	Out-of-distribution (OOD) input / scope creep	Unreliable outputs on unseen data patterns without detection	2	2	Low	Models are gatekept in variant specific pipelines.	Low

ID	Hazard Description	Consequence / Harm	P	S	Risk Level	Mitigation Strategy	Residual Risk
	AI model used outside validated use cases						
RSK-102	Overfitting / underfitting of the ML model A classification/scoring model (DiagAI Score, UP2, Phenogenius) fits its development / validation dataset too closely, or too loosely, so that its performance on new, real-world samples not represented in that dataset differs materially from its performance measured during validation.	Incorrect results - Misinterpretation of data	3	2	Medium	Training/Validation are performed on multiple datasets if possible (for some models) Train/validation/test split enforced with an independent holdout set for each model version; performance is monitored on production data via Post-market Performance Follow-up and compared against validation metrics; retraining/rollback is triggered if a drop is observed (see RSK-089, RSK-091).	Low
RSK-104	Data management & data quality Training data used to build or retrain a model is maliciously altered (data poisoning), a model artifact/weights file is substituted or tampered with without authorization, or crafted ("adversarial") input files are used to deliberately manipulate a model's output.	Incorrect results - Misinterpretation of data	3	2	Medium	Model artifacts are version-controlled and integrity-checked before deployment; access to training pipelines and datasets is restricted and logged in accordance with the ISMS (ISO/IEC 27001:2022/A1:2024); the release process requires validation of model provenance before a new version reaches production (see RSK-089).	Low
RSK-108	Post-production monitoring & performance degradation over time No structured post-market monitoring of ML-specific performance indicators (data drift, concept drift) is in place beyond general PMS activities. Gradual degradation of classification/scoring accuracy in the field is not detected in a timely manner	Incorrect results - Misinterpretation of data	2	2	Low	Post-market surveillance activities are documented in the PMS Procedure, and will be followed on a quarterly basis.	Low
RSK-110	Insufficient evaluation & testing of the trained model Model performance is overstated at release and not caught prior to production use	User may have incorrect results for the patient Misrepresentation of clinical information	2	2	Low	Analytical performances are executed before each release, as per PRC-027	Low

P = Probability (1=Improbable — 1 in 10 000 000; 2=Occasional — 1 in 1 000; 3=Probable — 1 in 100; 4=Frequent — 1 in 1) | S = Severity (1=Negligible — no/negligible patient risk; 2=Minor — temporary impairment, incorrect/missing variant; 3=Major — injury requiring medical intervention, false diagnostic; 4=Critical — death, permanent impairment) | Risk Level derived from SEQ-011 Risk Acceptability Matrix.

6.2. Impact Assessment

Question	Answer
High-impact individual decisions?	Yes — AI-assisted clinical variant reporting and diagnostic support
Sensitive data processed?	Yes — genomic data and clinical health information (PHI)
Overall risk level	High (prior to controls) — Low/Medium (post mitigation)
Referenced AIIA document	AIIA-001_DiagAI

7. Ethical Considerations

7.1. Fairness & Bias

Identified Biases:

- Training Data Bias: Acknowledged overrepresentation of European ancestry in the genetic databases used — performance may be systematically lower for non-European populations
- Phenotype Bias: Potential bias toward well-described 'classical' phenotypes; atypical or poorly documented phenotypes may yield lower-quality prioritization
- Publication Bias: Training data weighted toward published, positive findings — rare or under-studied gene-disease associations may be underweighted

Bias Assessment:

Analysis of the validation dataset (1701 cases) was performed to identify potential sources of bias. The dataset covers 849 different genes and 2526 different HPO terms, providing broad disease-group representation as a mitigation for disease-specific over-fitting.

Mitigation Strategies:

- Active recruitment of diverse ancestry training data
- Collaboration with international consortia for data diversity
- Regular bias audits across ancestry groups (quarterly performance analysis)
- Transparent reporting of performance by population
- Adjustment of priors based on patient ancestry when appropriate

7.2. Transparency & Explainability

Evidence Summaries:

Each ranked gene includes a breakdown of contributing evidence:

- Variant pathogenicity evidence (ACMG criteria)
- Phenotype match scores with specific matching HPO terms highlighted
- Inheritance pattern compatibility
- Population frequency considerations

Additional Explainability Mechanisms:

- Attention Visualization: Key phenotypes contributing to gene ranking are displayed
- Feature Importance: Breakdown of scoring components per gene
- Color-coded confidence levels (high/medium/low) with plain language explanations
- Links to primary literature and database entries

7.3. Privacy & Data Protection

Legal Framework	GDPR (EU); ISO 27001; ISO 27701; HIPAA (US);
Encryption	AES-256 at rest; TLS 1.3 in transit
Security Certification	ISO 27001 certified; Annual penetration testing

Access Control	Role-based access control (RBAC); multi-factor authentication (MFA) required ; principle of least privilege
Data Pseudonymization	All training data de-identified per Safe Harbor method; no PHI stored in model
Audit Logging	All data access events logged (HIPAA compliant)
Data Retention	GDPR Article 17 right to erasure implemented; automated deletion after retention period; secure disposal procedures

7.4. Clinical Ethics

- System designed to augment — not replace — clinical judgment; no automated clinical decisions without human oversight
- Clear delineation between AI suggestions and clinical determinations
- Informed consent required: clear communication of AI involvement, explanation of limitations
- Genetic counseling must be involved in result disclosure to patients

8. Monitoring & Maintenance

8.1. Known Limitations

⚠ *The following limitations have been formally identified. They must be communicated to all operators prior to deployment and upon each model update.*

Limitation	Impact	Communication Required?
Performance depends on completeness and accuracy of HPO terms entered for the patient	Sparse, ambiguous, or incorrect HPO input lowers ranking quality and may demote the true causative gene	Yes — users must be instructed to provide a complete and accurate HPO profile
Knowledge limited to gene–disease associations from 9 aggregated databases	Genes/diseases not yet curated in any source database cannot be prioritized, novel gene discovery is out of scope	Yes — users must be aware that absence of a high score is not evidence of absence of disease association
Output is a phenotype–gene correlation score, not a causality or pathogenicity verdict	A high PhenoGenius score does not establish that a variant is pathogenic, a low score does not rule it out	Yes — score must be interpreted alongside variant-level evidence (ACMG, segregation, etc.)
Source databases reflect a knowledge snapshot and a literature/population bias	Recently described gene–disease links may be missing, under-represented populations or ultra-rare conditions may be deprioritized	Yes — users must know the database refresh cadence and inherent curation bias
Validated only as a decision-support tool, not a diagnostic device	Cannot replace clinical/biological expert review and final medical judgment	Yes — labelling and IFU must state decision-support-only status
Designed for rare disease / Mendelian context	Reduced relevance for common, polygenic, or somatic disease contexts	Yes — out-of-scope contexts must be explicitly listed
Clinical scope limited to WES, WGS, and phenotype-driven gene panel selection	Use on other assay types (e.g. targeted somatic panels, RNA-only, methylation) is unvalidated	Yes — IFU must restrict use to validated assay types
When consumed by DiagAI, PhenoGenius contributes only one feature among several	Prioritization behaviour observed in DiagAI cannot be attributed solely to PhenoGenius, isolated interpretation of the score is misleading	Yes — downstream DiagAI documentation must clarify the feature's weight and role

Limitation	Impact	Communication Required?
Reproducibility tied to the version of HPO ontology and database snapshot used	Re-analysis at a later date may yield different rankings for the same case	Yes — version of HPO and database snapshot must be recorded with each analysis
Performance not characterised for paediatric vs. adult, prenatal, or specific ethnic subgroups	Diagnostic yield may vary across patient populations	Yes — known performance gaps must be communicated to users and regulators

8.2. Knowledge Base Update Schedule

Update Type	Frequency	Sources / Triggers	Approval
Knowledge Base Updates	Monthly	ClinVar, OMIM, HPO, literature monitoring	Clinical genetics board review; regression testing on benchmark dataset
Model Retraining	Quarterly or as needed	1,000 new high-quality training cases; major knowledge base updates	Clinical Safety Board + CTO; must meet or exceed previous version performance
Performance Reports	Quaterly	Automated generation with performance evaluation	Internal review
Risk Assessment	Quarterly	Scheduled review + any model changes	Compliance team

8.3. Documentation Update Triggers

Document	Update Trigger
Model Card (this document)	Within 30 days of any model changes
User Manual	Within 14 days of UI or functionality changes
Risk Assessment	Quarterly review; updated as needed
Performance Reports	Monthly generation

Retraining trigger criteria:

- Performance drop > 5% vs. validated baseline on reference cohort
- Major ClinVar update with significant change in training label distribution
- New gnomAD release with substantially revised population frequencies
- Serious incident linked to UP² output contributing to a significant adverse outcome
- Regulatory or policy framework change affecting this AI system
- Scheduled annual review — regardless of drift or incident

8.4. Model Drift Monitoring

Monitored Signal	Detection Method	Monitoring Frequency	Alert Threshold	Response Action
Concept drift (output performance)	Performance metrics tracked vs. validated baseline	Quarterly	ΔAccuracy > 5% or ΔSensitivity > 5%	Formal re-evaluation triggered — §8.2 retraining criteria
Output anomalies (score distribution)	Statistical monitoring of Phenogenius coefficiants distribution (mean, variance, outliers)	Continuous	Outside μ ± 3σ sustained	Manual inspection — escalate if patient impact confirmed

8.5. Decommissioning Plan

- Formal decommissioning decision approved by AI System Owner and top management
- All model artifacts (weights, training data, configurations) archived securely per data retention policy
- All integrations and API endpoints disabled; operators notified with 5 days' notice
- Final version of Model Card and all AIMS records archived in the document management system
- Post-decommissioning review: lessons learned captured for future model development

9. Governance & Compliance

9.1. Regulatory Compliance

AI Management	ISO/IEC 42001:2023 — AI Management System
Medical Device	EU IVDR (2017/746) — In Vitro Diagnostic Regulation
Data Protection	GDPR (EU); HIPAA (US); PIPEDA (Canada); national data protection laws as applicable
Clinical Standards	ACMG/AMP Guidelines for Sequence Variant Interpretation; ACMG Recommendations for Reporting Secondary Findings
Clinical AI	EU AI Act — High-risk AI system classification assessment ongoing
Digital Health	NICE Evidence Standards Framework for Digital Health Technologies
US FDA	FDA guidance on Clinical Decision Support Software

9.2. Approval & Accountability

Role	Responsibility
Model Developer / AI Team	Technical development, matrix construction, internal validation, maintenance
Data Scientist	Algorithm design, data source integration, performance analysis
Product Manager	Minor update approval; user requirement management
Clinical Expert	Clinical risk review; model retraining approval; major/minor update sign-off
Compliance Officer	Regulatory compliance oversight; GDPR/IVDR coordination
Medical/Scientific Advisor	Clinical performance interpretation; guideline alignment
Chief Technology Officer	Model retraining co-approval alongside Clinical Expert

10. Version History

Version	Description of changes	Approval date
1	Initial Model Card creation	2026-04-29
2	Update to dataset identifiers (§4 and §5) and risks (§6)	See approval date

Table of signatures	Name / Function	Date	Signature
Written by	Jiri Ruzicka Data Engineer	2026-09-25	
Reviewed by	Nicolas Duforet Datascience manager	2026-09-25	Nicolas Duforet
Approved by	Amélie Martinez AIGO	2026-09-25	Amélie Martinez

Appendix A: HPO Term Selection Guide

Principle: More specific is better than more numerous. Aim for 5-15 well-chosen, granular HPO terms.

Examples of Good HPO Terms:

- HP:0001263 — Global developmental delay
- HP:0001250 — Seizures
- HP:0002650 — Scoliosis
- HP:0000252 — Microcephaly

Less Optimal (too broad):

- HP:0000118 — Phenotypic abnormality (avoid unless no more specific term applies)

Tips:

- Start broad, then narrow down using the HPO browser
- Include temporal information when available (e.g. onset age, progression)
- Include informative negative findings
- Use modifiers when relevant (severity, progression rate)

Common Pitfalls to Avoid:

- Mixing syndromic diagnoses with individual phenotypes
- Using overlapping parent and child HPO terms simultaneously
- Omitting key discriminating phenotypes
- Including too many non-specific terms that dilute the signal

Appendix B: Glossary

ACMG	American College of Medical Genetics and Genomics
AI System	Combination of algorithms, models, knowledge bases, and infrastructure that processes phenotypic and genomic inputs to generate variant prioritization recommendations
AIMS	AI Management System as defined in ISO/IEC 42001:2023
Causal Variant	The genetic variant responsible for the patient's phenotype
ClinVar	Public archive of reports of relationships among human variations and phenotypes
Compound Heterozygous	Two different pathogenic variants in the same gene, one from each parent
EU IVDR	EU In Vitro Diagnostic Regulation 2017/746
GDPR	General Data Protection Regulation (EU) 2016/679
HIPAA	Health Insurance Portability and Accountability Act (US)
HPO	Human Phenotype Ontology — standardized vocabulary of phenotypic abnormalities (hpo.jax.org)
ISO 42001	ISO/IEC 42001:2023 — AI Management Systems standard
Model Drift	Degradation in model performance over time due to changes in data distribution or evolving knowledge
OMIM	Online Mendelian Inheritance in Man — database of human genes and genetic disorders
Orphanet	European rare disease and orphan drug database (orphadata.org)
PanelApp	Genomics England gene panel database with clinical confidence ratings

Phenotype	Observable characteristics of an organism resulting from genetic and environmental factors
SHAP	SHapley Additive exPlanations — method for attributing prediction contributions to input features
Variant Prioritization	Process of ranking genetic variants by likelihood of being disease-causing
VCF	Variant Call Format — standard file format for storing gene sequence variations
VUS	Variant of Uncertain Significance — genetic variant whose relationship to disease is unclear