AI Model Card: DiagAI Score v1.0

Compliance Framework: ISO/IEC 42001:2023

1. General Information

1.1. Basic Information

AI System Name	DiagAI Score
Model Identifier	AIMC-03
Model Version	1.0
Release Date	November 2024 (SeqOne Platform v1.37)
Model Type	Linear regression combining 10 features
Development Team	SSC, SeqOne Nicolas Duforet-Frebourg | Jiri Ruzicka | Alexandre Boulat
Maintenance Status	Active
Related Products	UP2 (Universal Pathogenicity Predictor), DiagAI HPO, PhenoGenius, 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

DiagAI Score is an AI-powered variant prioritization model that integrates multiple evidence streams to produce a single composite score (0-100) reflecting the probability that a genetic variant is causally responsible for a patient's clinical phenotype. It combines pathogenicity evidence, phenotypic relevance, gene-panel alignment, variant quality, and family context into a single ranked prioritization signal within the SeqOne Platform.

Primary Functions:

- Integrate eight feature inputs (pathogenicity, phenotypic similarity, panel relevance, variant quality, transmission mode, and family structural features) into a single 0-100 prioritization score.
- Rank candidate variants within a patient case, placing the most likely causal variant at the top.
- Power the DiagAI Smart Picks (score >=90, ~90% precision) and DiagAI Short List (score >=71, ~95% sensitivity) features.
- Provide transparent score decomposition breaking down the contribution of each of the 10 input features.
- Support clinicians in reducing the diagnostic search space for rare disease genomic analysis.

Clinical Role:

DiagAI Score is a clinical decision support tool intended to augment — not replace — expert clinical judgement. It must always be used alongside ACMG/AMP variant classification guidelines and thorough review of primary clinical and genetic evidence.

2.2. Target Users

Primary Users:

Board-certified clinical geneticists, molecular geneticists, genetic counselors, and laboratory professionals performing genomic diagnostic analysis.

Indirect Users:

- Patients receiving genomic diagnostic services
- Healthcare systems relying on diagnostic outputs
- Regulatory authorities overseeing IVD device performance

2.3. Intended Use Environment

Setting	Clinical genomics and molecular diagnostics laboratories
Platform	SeqOne Platform (cloud-based, GxP-compliant)
Integration	Used downstream of variant calling, annotation, and UP2 pathogenicity scoring
Regulatory Context	EU IVDR (2017/746) — Class IIa In Vitro Diagnostic Device
CE Marking	CE-IVD certified under EU IVDR
Geographic Scope	France (primary), European Union, international authorized markets

2.4. Out-of-Scope Uses

- Standalone clinical diagnosis or variant classification without expert review
- Direct patient reporting without expert interpretation and genetic counseling
- Automated clinical decision-making or reporting without human sign-off
- Somatic/cancer variant prioritization (unless separately validated)
- Pharmacogenomic variant interpretation
- Complex structural variant or repeat expansion disorder analysis
- Forensic, legal, or employment screening purposes
- Use outside the SeqOne Platform or by users who have not completed required training

3. Model Architecture & Technical Specifications

3.1. Algorithm Overview

Algorithm Type	Linear regression
Input Features	10 composite features
Output	Continuous score 0-100; higher = more likely causal
Interpretability	per-feature contribution breakdown
Uncertainty	Score confidence interval displayed in UI
Computation	Real-time within the SeqOne Platform pipeline

3.2. Input Features

DiagAI Score integrates 10 evidence-based features, each contributing a weighted component to the final composite score:

Feature	Description	Source
UP2 Score	Universal Pathogenicity Predictor v2 — ML pathogenicity estimate for the variant (XGBoost, ~75 inputs)	SeqOne UP2 model
PhenoGenius Score	Phenotypic similarity between patient HPO terms and gene-disease phenotypes; quantifies match quality	PhenoGenius (SeqOne)
PanelApp Gene Relevance	Relevance of the gene to the patient's clinical panel, based on PanelApp gene confidence levels	Genomics England PanelApp, PanelApp Australia
Variant Quality Score	Technical quality of the variant call (sequencing quality, genotype confidence, coverage metrics)	SeqOne QC pipeline
Transmission Mode	Compatibility of the variant's genotype with the expected disease transmission for the gene	OMIM, GenCC, PanelApp England, PanelApp Australia
Homozygosity Flag	Whether the variant is in a homozygous or hemizygous state, relevant for AR/XL disease	Variant annotation
De Novo Status	Confirmation or strong prediction that the variant arose de novo (not inherited from either parent)	Family data
Family Co-segregation	Evidence of co-segregation of the variant with disease phenotype across family members	Familydata

3.3. Output Specification

Score Range	0 (benign) to 100 (strong pathogenic causal candidate)
Score Thresholds	Smart Picks: >=90 (~90% precision) | Short List: >=71 (~95% sensitivity) | Without HPO Short List: >=65
Score Display	Numeric score + color-coded gauge
Evidence Breakdown	Contribution of each of the 10 input features
Conflict Flags	Automatic detection and display of conflicting evidence across sources

3.4. Score Interpretation Scale

Score Range	Interpretation	Clinical Guidance
0 - 9	Very strong evidence for benign	Consider benign classification; apply ACMG criteria
10 - 30	Strong evidence for benign	Likely benign; complete ACMG evaluation
31 - 49	Moderate evidence for benign	Uncertain — additional evidence needed
50 - 69	Moderate evidence for pathogenic	Uncertain — additional evidence needed
70 - 89	Strong evidence for pathogenic	Likely pathogenic; complete ACMG evaluation
90 - 100	Very strong evidence for pathogenic	Consider pathogenic classification; apply ACMG criteria

ℹ️ Always review the confidence interval width. Wide intervals (e.g. 40-80) indicate high uncertainty and require increased caution. ACMG/AMP classification must integrate all available evidence.

3.5. Infrastructure

Deployment	SeqOne Platform (cloud-based, GxP environment)
Docker Image	654654589101.dkr.ecr.eu-west-3.amazonaws.com/seqone-registry/tools/classify-ml
API	Integrated within SeqOne Platform pipeline; not externally exposed as standalone API
Data Flow	Variant annotations from upstream pipeline -> Feature extraction -> Linear regression -> Score output -> UI display

3.6. Frameworks and Libraries

The model was developed using the following libraries:

Library / Framework	Version
Python	3.8.x
numpy	1.26.4
pandas	2.1.4
scikit-learn	1.4.1

4. Training Data

4.1. Dataset Overview

Dataset Type	Proprietary clinical diagnostic dataset — pseudonymized
Total Variants	46 million variants
Total Cases	559 patient cases
Causal Variants	678 confirmed causal variants
Geography	France
Collection Period	2021 - 2023
Data Custodian	SeqOne
Reference Genome	GRCh37

4.2. Data Characteristics

Case Mix:

- Rare genetic disease cases submitted through French clinical genetics centres
- Includes trio and singleton analyses across a broad range of rare disease categories
- Variant types: SNVs, small indels, CNVs
- Inheritance modes represented: autosomal dominant (AD), autosomal recessive (AR), X-linked dominant (XLD), X-linked recessive (XLR), de novo

Labelling:

- Ground truth established from confirmed clinical diagnoses and expert variant classifications
- 678 causal variants used as positive training labels
- Remaining variants in cases serve as negative/non-causal examples

Causal Variant Types:

Variant Type	Count
Missense variant	402
Frameshift variant	102
Stop gained	83
Splice donor variant	23
Inframe deletion	23
Splice acceptor variant	18
Splice region variant	13
Splice polypyrimidine tract variant	4
Synonymous variant	4
Start lost	2
Splice donor region variant	2
Splice donor 5th base variant	1
Intron variant	1
Inframe insertion	3

Disease Categories (training cohort):

Disease Category	N Cases
Nephropathie indeterminee avec insuffisance renale	194
Prenatal - Signes d'appel echographiques	88
Nephrogenitique	85
Anomalies du developpement - sans hypothese diagnostic	71
Nephropathie kystique autosomique dominante y compris polykystose AD	48
DI syndromique	47
Deficience intellectuelle - DI sans hypothese diagnostic	35
Prenatal SAE	33
Oncogenetique	33
Autres (genetique)	24
Cancers du sein : Predisposition hereditaire aux cancers	24
DI sans malformation	22
Anomalies du developpement - avec hypothese diagnostic	22
Foetopathologie - Syndrome malformatif	14
Foetopathologie	14
Pancreatite	13
Hepato-gastro	13
Lithiase ou nephrocalcinose	12
Pancreatite hereditaire et idiopathique	11
Nephropathie tubulo-interstitielle autosomique dominante (ADTKD)	10
Proteinurie et syndrome nephrotique cortico-resistant	10
Epilepsies	10
Maladie auto-inflammatoire sans hypothese diagnostic	9
Neonatale	8
Maladies auto-inflammatoire	8
Nephropathie hematurique familiale	8
Tubulopathie	8

Inclusion/Exclusion Criteria:

Variants were called by SeqOne in-house pipeline. Risk factor variants (such as APOL1 variants) were excluded. Only single nucleotide variants (SNVs) and short indels (<300 bp) were included.

Annotation/Labeling Process:

Ground truth for the proprietary dataset was established by dual-reading by two independent experts, with a third adjudicating any conflicts. All variants not reported by experts are labeled with a score of 0. Variants reported by experts are labeled based on the ACMG classification given:

- VUS classification: label of 50
- Likely pathogenic classification: label of 80
- Pathogenic classification: label of 90

Dataset Breakdown:

The total combined dataset of 46 million variants with 678 causal variants was partitioned into independent, patient-level-stratified sets:

- Training Set: 26 million variants, with 397 causal variants from 307 samples, used to train the model parameters.
- Validation Set: 20 million variants with 281 diagnostic variants from 252 samples, used for validation of DiagAI score prediction.
- Testing Cohort JT: First multicentric cohort of 316 samples with known diagnostic variants and phenotypes, used to calibrate thresholds for smart picks and short lists.
- Testing Cohort LM: Second cohort of 196 samples with known diagnostic variants and phenotypes from the same hospital, with nephrologic diseases.

4.3. Data Governance

Privacy Framework	GDPR (EU) compliant; patient data fully pseudonymized before use
Data Access	Restricted to authorized SeqOne staff under data processing agreements
Retention Policy	Governed by SeqOne data governance framework and applicable regulations
Bias Monitoring	Dataset reviewed for demographic and disease-type representation

4.4. Known Limitations of Training Data

- Dataset is predominantly French (European ancestry); performance may be reduced for variants from underrepresented ancestry groups
- Demographic data (age, sex, ethnic background) not systematically available — limits targeted bias analysis
- Known bias toward well-studied genes with extensive published evidence (e.g. COL4A3, PKD1, CFTR)
- Cases with ambiguous or conflicting diagnoses were excluded — may limit generalizability to complex diagnostic scenarios
- Training data does not include somatic variants, pharmacogenomic variants, or mitochondrial variants

4.5. Data Preprocessing

The following pre-processing steps were applied consistently across all datasets:

- Feature Extraction: Bioinformatic annotations (e.g., OMIM, GenCC) were extracted from genomic databases.
- Imputation: Missing numerical values were imputed using a linear regression model. Features were imputed as 0 if the given pattern is not present or not applicable, corresponding to biological standards of the features.
- Normalization: Binary features were already distributed between 0 and 1. For continuous features, UP2 score was rescaled using min-max normalization and PhenoGenius raw score was rescaled using mean normalization.
- Data Augmentation (Training set only): Due to the heavily imbalanced dataset towards non-causal variants, variants of interest were augmented to have greater impact during training.
- Feature Engineering: Several features were engineered to correspond to expert knowledge rules, including inheritance transmission patterns, family inheritance from parents, and variant quality call, allowing distinction between false positive variants and patient variants.

4.6. Training Process

The final DiagAI score (ranging from 0-100) results from a supervised learning approach trained on a diverse cohort of diagnosed cases and their associated causal variants.

Training Algorithm:

- Algorithm: Linear regression
- Loss Function: Residual sum of squares between observed targets in the dataset and targets predicted by the linear approximation.

Training Methodology:

- Objective Function: Ordinary least squares
- Optimization: Non-Negative least squares optimization

Subsampling and Weighting:

- Subsampling of 80% of variants not passing the compatibility logic filter of SeqOne to remove large numbers of negatives. The Compatibility filter is based on the transmission modes associated with each variant and their genotype calling.
- Weight importance set for positive variants: 3000x for variants with target 50; 4000x for variants with target 80; 4000x for variants with target 90; 5000x for variants with target 95.

Split-Validation:

- Strategy: Stratified split (307 samples per training set, 252 per validation set)
- Stratification by: HPO term presence and family analyses

Computational Resources:

- Training duration: ~1 hour
- Hardware: 1 CPU
- Framework: Python 3.x, scikit-learn, pandas

5. Performance Metrics & Evaluation

5.1. Evaluation Methodology

Test Dataset	Hold-out test set from same pseudonymized clinical cohort (temporal and random split)
Cross-Validation	5-fold stratified cross-validation during development; stratification by pathogenicity class and variant type
Primary Metric	Rank of causal variant among all candidates in a case (Rank 1 = top-ranked)
Secondary Metrics	Top-N sensitivity (% cases where causal variant is in top N), comparison with external tools

5.2. Overall Performance — SNV/Indel

Performance Detail — With HPO Terms:

Metric	Validation Set	Test Set
N	**34**	**196**
Top 1	**24 (70.6%)**	**144 (73.4%)**
Top 3	**32 (94.1%)**	**172 (87.8%)**
Top 10	**32 (94.1%)**	**187 (95.4%)**

Performance Detail — Without HPO Terms:

Metric	Validation Set	Test Set
N	**176**	**196**
Top 1	**67 (38.1%)**	**82 (41.8%)**
Top 3	**124 (70.5%)**	**155 (79.1%)**
Top 10	**153 (86.9%)**	**179 (91.3%)**

5.3. Overall Performance — CNV

Metric	Value
Top-1 sensitivity (CNV)	83.3%

5.4. Performance by Variant Type (SNV/Indel, with HPO)

Detailed Performance by Effect (with HPO):

Effect	N	Median Rank	Top 1	Top 3	Top 10
Frameshift variant	41	2.0	19	32	37
Inframe deletion	8	1.5	4	8	8
Inframe insertion	3	2.0	1	2	3
Missense variant	109	2.0	4	78	94
Splice acceptor variant	5	5.0	1	2	3
Splice donor 5th base variant	1	3.0	0	1	1
Splice donor region variant	1	3.0	0	1	1
Splice donor variant	3	3.0	1	2	3
Splice region variant	5	3.0	1	3	4
Stop gained	34	1.0	21	27	31

5.5. Performance by Score Thresholds

Distribution of causal variants across score thresholds — With HPO:

Score Threshold	Validation Sensitivity	Test Sensitivity	Predicted Clinical Interpretation
≥90 (Very High)	21 (61.8%)	141 (58.5%)	Strong evidence for pathogenicity
≥70 (High)	13 (38.2%)	87 (36.1%)	Likely pathogenic range
30–70 (Moderate)	0 (0.0%)	9 (3.7%)	Uncertain significance
≤30 (Low)	0 (0.0%)	3 (1.2%)	Likely benign range
≤10 (Very Low)	0 (0.0%)	1 (0.4%)	Strong evidence for benign

Distribution of causal variants across score thresholds — Without HPO:

Score Threshold	Validation Sensitivity	Test Sensitivity	Predicted Clinical Interpretation
≥90 (Very High)	8 (4.5%)	0 (0.0%)	Strong evidence for pathogenicity
≥70 (High)	168 (95.5%)	213 (88.4%)	Likely pathogenic range
30–70 (Moderate)	0 (0.0%)	24 (10.0%)	Uncertain significance
≤30 (Low)	0 (0.0%)	0 (0.0%)	Likely benign range
≤10 (Very Low)	0 (0.0%)	4 (1.7%)	Strong evidence for benign

5.6. Performance by Inheritance Mode

Validation Set — Detailed:

Inheritance	N	Median Rank	Top 1	Top 3	Top 10
Autosomal Dominant (AD)	144	2.0	64	111	134
Autosomal Recessive (AR)	151	2.0	68	111	129
X-Linked Dominant (XLD)	17	1.0	14	16	16
X-Linked Recessive (XLR)	17	1.0	14	16	16

Test Dataset — Detailed:

Inheritance	N	Median Rank	Top 1	Top 3	Top 10
Autosomal Dominant (AD)	179	1.0	137	163	170
Autosomal Recessive (AR)	169	1.0	123	145	157
X-Linked Dominant (XLD)	29	1.0	27	28	29
X-Linked Recessive (XLR)	28	1.0	27	28	28

5.7. Benchmark Comparison

Performance comparison on the internal test set (% of cases where causal variant ranked in top positions):

Method	Top-1 or Top Rank %	Notes
DiagAI Score (SeqOne)	73.9%	Integrated multi-feature scoring
Exomiser	62.4%	Phenotype-driven variant prioritization tool
MARRVEL	43.6%	Database-driven variant annotation tool

ℹ️ DiagAI Score achieves 73.9% top-ranked causal variant placement vs. 62.4% for Exomiser and 43.6% for MARRVEL on the internal test set, representing a significant improvement in diagnostic yield efficiency.

5.8. Clinical Decision Thresholds

Threshold	Score Cutoff	Performance Target	Use Case
DiagAI Smart Picks	>=90	~90% precision	High-confidence rapid review — highest priority variants
DiagAI Short List (with HPO)	>=70	~95% sensitivity	Comprehensive shortlist — minimize missed causal variants
DiagAI Short List (no HPO)	>=66	~95% sensitivity	Used when HPO phenotype data unavailable

5.9. Performance Monitoring

- Quarterly analysis of score distributions, calibration, and user-reported discrepancies
- Annual comprehensive performance re-evaluation against updated test sets

References : Tenon cohort analysis, Validation analysis.

6. Risk Assessment

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-83	Annotation Quality Dependency / Score quality heavily dependent on completeness of upstream variant annotations	Incorrect pathogenic classification when annotations are incomplete or low-quality	2	2	Low	Annotation quality checks for some annotations and warnings built into pipeline; Score is explainable to users (SR-252 to SR-257)	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-85	Conflicting Evidence / Contradictory signals across ClinVar, ACMG, and in silico predictors may produce unreliable scores	Incorrect results - Misinterpretation of data	2	2	Low	All information is displayed in the UI. Score is explainabled to users (SR-252 to SR-257). SHAP breakdown displayed for ClinVar and in silico predictors	Low
RSK-86	"False negative resutls such as : - pathogenic variant scored below threshold, as Benign	Incorrect results — misinterpretation of data; missed pathogenic variant; delayed or incorrect	2	2	Low	Mandatory human review before clinical reporting; Score is explainable to users (SR-252 to SR-257).	Low

ID	Hazard Description	Consequence / Harm	P	S	Risk Level	Mitigation Strategy	Residual Risk
	- miss an article about the variant of interest"	diagnosis; potential patient harm				Module retraining according to users false negative report. Thresholds are calibrated and displayed via tooltip.	
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
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	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	Low

ID	Hazard Description	Consequence / Harm	P	S	Risk Level	Mitigation Strategy	Residual Risk
	after model training, and as genomic databases and evidence evolve post-training					missing an important annotation. Models are retrained at least annually	
RSK-94	Out-of-distribution (OOD) input / scope creep AI model used outside validated use cases	Unreliable outputs on unseen data patterns without detection	2	2	Low	Models are gatekept in variant specific pipelines.	Low
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-103	Feature extraction defect feeding the ML model An internal computation error in the feature-engineering step produces an incorrect input feature that is silently passed to a classification/scoring model (DiagAI Score, UP2), independently of any external database or annotation error.	Incorrect results - Misinterpretation of data	3	2	Medium	Automated unit and non-regression tests cover the feature-computation pipeline. Model evaluation with feature engineering pipeline before release Display of feature values in the explanability interface	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)	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

ID	Hazard Description	Consequence / Harm	P	S	Risk Level	Mitigation Strategy	Residual Risk
	is in place beyond general PMS activities. Gradual degradation of classification/scoring accuracy in the field is not detected in a timely manner						
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 Result

Question	Answer
High-impact individual decisions?	Yes — variant pathogenicity scoring directly influences clinical genomic interpretation and diagnostic decisions
Sensitive data processed?	Yes — genomic variant data; all training data de-identified; no PHI stored in model
Overall risk level	High (clinical diagnostic aid; patient safety implications)
Referenced AIIA document	AIIA-001_DiagAI

7. Ethical Considerations

7.1. Fairness & Bias

Identified Biases:

- Training data is predominantly French (European ancestry); performance may be systematically lower for variants from non-European populations
- Known bias toward well-studied genes with rich published evidence (e.g. COL4A3, PKD1, CFTR) — variants in poorly characterized genes may receive lower, less informative scores
- Demographic data (age, sex, ethnic background) was not systematically available during training — limits targeted bias analysis and correction

Mitigation Strategies:

- Broad disease category coverage during training to reduce disease-specific over-fitting
- Bias monitoring as part of quarterly performance reviews
- Performance stratification by variant type and inheritance mode to detect differential effects
- User guidance on known limitations for specific populations and gene classes
- Ongoing plans to expand training data diversity in future model versions

7.2. Transparency & Explainability

Score Decomposition	Predictor values calculated for each variant, showing contribution of all 10 input features to the final score
UI Visualization	Color-coded gauge (0-100 scale), numeric score, and feature-level breakdown displayed in SeqOne Platform
Evidence Attribution	Sources driving the score are displayed with individual feature scores, allowing clinician to assess evidence quality
Conflict Detection	Automatic detection and visual flagging of conflicting evidence across databases and predictors
Score Icons	Visual icons indicate score tier (Smart Pick, Short List candidate, etc.) for rapid clinical orientation

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 patient data pseudonymized before use in model development and validation
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 during development and validation. They must be communicated to all operators prior to deployment and upon each model update.**

Limitation	Impact	Communication Required?
Reduced accuracy for variants in genes with limited functional data	May yield inconclusive or incorrect scores for poorly characterised genes; requires expert review	Yes
Performance depends on quality and completeness of input annotations	Incomplete or erroneous annotations may produce unreliable scores; annotation pipeline quality must be ensured	Yes
Requires comprehensive variant annotation pipeline	Scores cannot be generated without a fully configured annotation pipeline; partial pipelines will cause errors or missing results	Yes
Scores are limited to monogenic diseases	Cannot be applied to polygenic or complex trait analyses; use is restricted to monogenic disease contexts	Yes
Scores are estimates, not definitive classifications	Scores must be interpreted in clinical context and cannot replace expert clinical judgement	Yes
Cannot replace comprehensive clinical evaluation	Clinical decisions must not be based solely on the score; full clinical workup remains mandatory	Yes
Limited performance on variants with conflicting evidence	Variants with contradictory database entries may receive unreliable scores; manual curation is recommended	Yes
Cannot account for genetic modifiers or oligogenic effects	Scores may under- or over-estimate pathogenicity in complex genetic contexts; oligogenic cases require additional interpretation	Yes
Does not consider environmental factors or epigenetics	Phenotypic outcomes influenced by non-genetic factors are not reflected in the score	Yes
Limited ability to predict phenotypic variability	Variants with variable expressivity may receive scores that do not reflect the full clinical spectrum	Yes
Cannot definitively determine causality without functional validation	Functional studies remain required to confirm pathogenicity for variants of uncertain significance	Yes
Training data biased toward well-studied genes and common diseases	Performance may be reduced for rare diseases or understudied gene regions; scores should be interpreted with caution in these contexts	Yes
Underrepresentation of certain ancestries may affect calibration	Score calibration may be less accurate for patients from underrepresented ancestral backgrounds	Yes
Dependent on accuracy of training data classifications	Errors in reference databases propagate into model predictions; known misclassifications in source data affect reliability	Yes
Historical biases in literature and databases may be reflected	Variants over-represented in historical studies may receive inflated pathogenicity scores	Yes
Training Data integrates a significant number of missing clinical information	Missing phenotype or clinical data in the training set may reduce model accuracy and calibration	Yes

8.2. 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
- Serious incident linked to AI output contributing to a significant adverse outcome
- Regulatory or policy framework change affecting this AI system

8.3. Model Drift Monitoring

Monitored Signal	Detection Method	Monitoring Frequency	Alert Threshold	Response Action
Concept drift (output performance)	Performance metrics tracked vs. validated baseline	Quarterly	ΔTop-5 accuracy > 5%	Formal re-evaluation triggered — §8.2 retraining criteria
Output anomalies (score distribution)	Statistical monitoring of prediction score distribution	Quaterly	shortlist lenghts increase above a median of 25	Manual inspection — escalate if patient impact confirmed

8.4. 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 AI system 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	2025-12-05
2	Update to risks (§6), and sources (§3.2)	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	Amélie Martinez	2026-09-25

Appendix A: Score Quick Reference

Score Range	Interpretation	Confidence Level	Clinical Action
0 - 9	Very strong evidence for benign	High	Consider benign classification; apply ACMG criteria
10 - 30	Strong evidence for benign	Moderate-High	Likely benign; complete ACMG evaluation
31 - 49	Moderate evidence for benign	Moderate	Uncertain; additional evidence needed
50 - 69	Moderate evidence for pathogenic	Moderate	Uncertain; additional evidence needed
70 - 89	Strong evidence for pathogenic	Moderate-High	Likely pathogenic; complete ACMG evaluation
90 - 100	Very strong evidence for pathogenic	High	Consider pathogenic classification; apply ACMG criteria

Appendix B: Glossary

ACMG/AMP	American College of Medical Genetics / Association for Molecular Pathology — variant interpretation guidelines
AUC-ROC	Area Under the Receiver Operating Characteristic Curve — measure of classification performance
Calibration	Alignment between predicted probabilities and observed outcome frequencies
CE-IVD	CE marking for In Vitro Diagnostic devices under EU IVDR
DiagAI Short List	SeqOne feature displaying variants with DiagAI Score >=70 (or >=66 without HPO) — ~95% sensitivity
DiagAI Smart Picks	SeqOne feature displaying variants with DiagAI Score >=90 — ~90% precision
ECE	Expected Calibration Error — measure of probability calibration quality
EU IVDR	EU In Vitro Diagnostic Regulation 2017/746
GDPR	General Data Protection Regulation (EU) 2016/679
gnomAD	Genome Aggregation Database — large population allele frequency reference
HIPAA	Health Insurance Portability and Accountability Act (US)
HPO	Human Phenotype Ontology — standardized phenotype terminology
ISO 42001	ISO/IEC 42001:2023 — AI Management Systems standard
PanelApp	Genomics England gene panel database with clinical confidence ratings
PhenoGenius	SeqOne phenotype similarity scoring tool using HPO terms
SHAP	SHapley Additive exPlanations — method attributing prediction contribution to each feature
SNV	Single Nucleotide Variant
CNV	Copy Number Variant
UP2	Universal Pathogenicity Predictor v2 — SeqOne XGBoost variant pathogenicity model
VUS	Variant of Uncertain Significance — variant that cannot be confidently classified as pathogenic or benign