Review

AI in Variant Analysis: Fast Track to Genetic Diagnoses

Elizabeth J. Wilk †, Sasha Taluri †, Timothy C. Howton, Anthony B. Crumley, Michal Mrug, Brittany N. Lasseigne *

Posted Date: 10 April 2026

doi: 10.20944/preprints202603.1025.v2

Keywords: diagnostic odyssey; variant reclassification; VUS; clinical genomics; black-box; artificial intelligence

 doi:10.20944/preprints202603.1025.v2

Review

AI in Variant Analysis: Fast Track to Genetic Diagnoses

Elizabeth J. Wilk [1,†], Sasha Taluri [1,†], Timothy C. Howton [1], Anthony B. Crumley [1], Michal Mrug [2,3] and Brittany N. Lasseigne [1,*]

[1] The Department of Cell, Developmental and Integrative Biology, Heersink School of Medicine, The University of Alabama at Birmingham, Birmingham, Alabama, USA

[2] The Department of Medicine, Heersink School of Medicine, The University of Alabama at Birmingham, Birmingham, Alabama, USA

[3] Department of Veterans Affairs Medical Center, Birmingham, Alabama, USA

* Correspondence: bnp0001@uab.edu

† Co-first authors.

Abstract

While falling costs have expanded access to genomic sequencing, clinical utility is frequently hindered by the challenge of interpreting complex genetic data. Although advances in genetic variant classification have improved diagnostic precision, they have also increased the identification of variants of uncertain significance (VUSs), widening the interpretation gap between data generation and clinical actionability. The high prevalence of VUSs can lead to false reassurance or psychological distress, as patients and non-expert clinicians may misinterpret inconclusive results. We propose that artificial intelligence (AI) is a critical clinical decision-support tool for bridging this gap, offering a scalable framework to optimize variant interpretation and shorten the diagnostic odyssey. We advocate integrating AI throughout the genetic diagnostic workflow–from initial phenotyping to variant prioritization–to facilitate data-driven, personalized treatment. We outline current AI-assisted approaches and discuss anticipated challenges in this pursuit, such as privacy, training data bias and quality, model explainability, and the necessity of a total product life cycle for validation. To address these challenges, we provide recommendations to ensure AI tools meet the highest standards of precision, reproducibility, and transparency. By standardizing AI across the variant analysis pipeline, we can fast-track the path to genetic diagnoses, effectively bridging the interpretation gap and enabling rapid delivery of personalized medical interventions.

Keywords: diagnostic odyssey; variant reclassification; VUS; clinical genomics; black-box; artificial intelligence

Introduction

The average time to receive a genetic diagnosis across high-income countries ranges from 4 to 19 years(Phillips et al. 2024; Faye et al. 2024), producing $86,000 to $516,000 in avoidable costs per patient(Lewin Group 2023). Current practices force patients with genetic diseases into a 'diagnostic odyssey,' subjecting them to rounds of unnecessary clinic visits, procedures, and medications (**Figure 1**). This process closes or narrows their window of intervention, enabling disease progression and long-term disease damage. High-throughput genetic testing, critical for addressing the diagnostic odyssey, has become widely accessible and cost-effective(Kris A. Wetterstrand 2019). Even when paid out of pocket, sequencing costs are a fraction of the overall odyssey's costs.

Approximately 30 million Americans have a genetic disease (~ 1 in 10)(Lewin Group 2023; Wan et al. 2023). Therefore, early disease identification and therapeutic intervention should be the norm. However, physicians report limitations in their genetics training(Peabody et al. 2015; Rasouly et al.

 doi:10.20944/preprints202603.1025.v2

2023; Kneifati-Hayek et al. 2024), and many express reduced interest in genetic screening due to the rarity of genetic conditions(Pasquier et al. 2022; Wan et al. 2023).

Variant analysis—the rate-limiting step in genetic testing(Tagliafico et al. 2018)—classifies variants by pathogenicity to guide clinical decision-making. Inaccurate interpretation at this stage fundamentally alters patient management, preventing the use of targeted therapies, initiating surveillance, or performing preventive procedures(Agaoglu et al. 2022). These errors also extend to the family, obscuring the need for cascade screening or preimplantation genetic diagnosis(McNeill 2022). Consequently, misinterpreted variants contribute to avoidable morbidity and mortality through missed preventative interventions, while simultaneously inflicting psychological harm via false reassurance or unnecessary anxiety(Campeau 2022). The standard of care uses ACMG/AMP and/or ESHG guidelines(Richards et al. 2015; Houge et al. 2022) for variant interpretation, but the process as a whole remains labor-intensive and relies heavily on experts. Results can be inconsistent and often yield variants that lack sufficient evidence to be classified as benign or pathogenic(Agaoglu et al. 2022; Lin et al. 2023; Zukin et al. 2023), complicating patient care. However, automation that leverages all available clinical, molecular, and population data in a standardized, reproducible manner could help reduce these issues. Artificial intelligence (AI), tools with "human-like reasoning" built from a variety of machine learning (ML) models and/or large language models (LLMs) (reviewed in(Nichols et al. 2019; Koteluk et al. 2021; Russell and Norvig 2021; Janiesch et al. 2021)), can optimize labor- and knowledge-intensive steps throughout the genetic testing process.

In this perspective (**Table 1**), we highlight opportunities, challenges, and recommendations for incorporating AI into variant analysis to support clinical genetic testing and research.

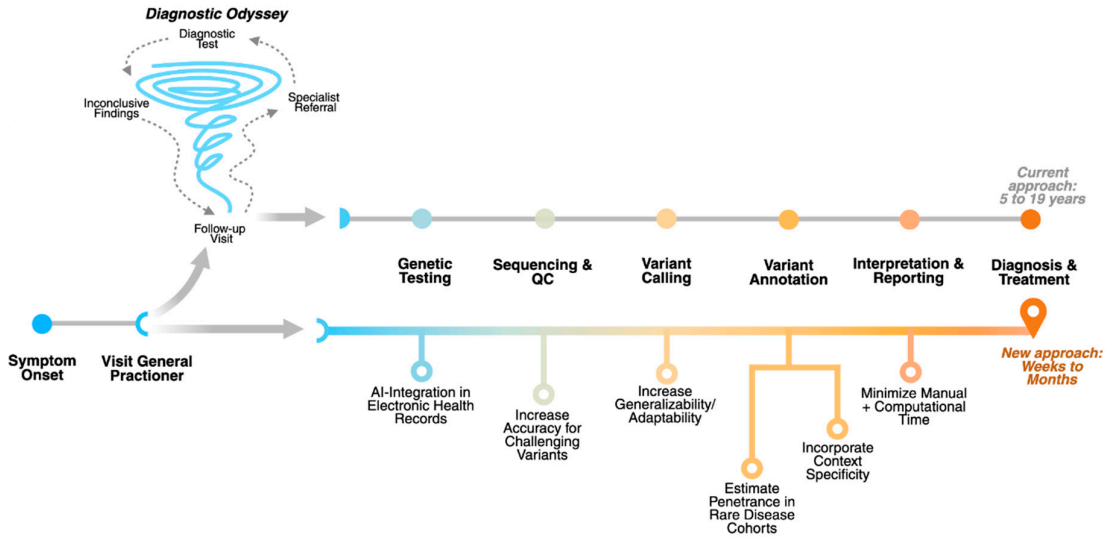

Figure 1. Schematic contrasting the current (filled circles) cyclic ("Diagnostic Odyssey" tornado) approach with targeted AI opportunities (unfilled circles) that improve variant analysis and shorten time to diagnosis (Created in BioRender. Wilk and Taluri (2026) https://BioRender.com/6xxko8q).

Descriptive caption: Diagram showing the steps involved in reaching a genetic diagnosis. After the first two steps: "symptom onset" and "visit general practitioner," the path diverges to compare the "Current approach" and the "New approach." The "Current approach," after "visit general practitioner" includes the diagnostic odyssey, represented as a tornado labeled at points around the tornado: "Follow-up Visit," "Specialist Referral," "Diagnostic Test," "Inconclusive Findings," leading back again to "Follow-up Visit." The eventual continuation after the Diagnostic Odyssey shows the typical steps from genetic testing to reaching a diagnosis and treatment. The "New approach" divergent path lists recommendations at each step: "AI-Integration in Electronic Health Records" for Genetic Testing, "Increase Accuracy for Challenging Variants" for Sequencing & QC, "Increase

 doi:10.20944/preprints202603.1025.v2

Generalizability/Adaptability" for Variant Calling, "Estimate Penetrance in Rare Disease Cohorts" and "Incorporate Context Specificity" for Variant Annotation, "Minimize Manual + Computational Time" for Interpretation & Reporting. After "Diagnosis & Treatment" the timeframes are noted as 5-19 years for the "Current approach" compared to "New approach" with AI recommendations taking weeks to months.

Approach to Variant Analysis

AI is emerging at a time when clinical genetics faces its greatest gap between knowledge and practice.

To prevent the diagnostic odyssey (**Figure 1**), physicians must first recognize patients who would benefit from genetic testing. Genetic diseases typically present with a constellation of signs (e.g., dysmorphism, early-onset, and/or multi-system involvement); therefore, AI can assist in determining when genetic testing may be appropriate (e.g., FACE2GENE(Gurovich et al. 2019)). For instance, AI-integrations in electronic health records (EHRs) could detect potential patients, even those with subtle clinical presentations(Yang et al. 2022; Ye et al. 2024). AI can also support physicians' continuing education through adaptive educational modules that account for each individual's time constraints, goals, and baseline knowledge(Hajek et al. 2022).

After sequencing, variant analysis processes the data in four key steps: variant calling, annotation, prioritization, and interpretation (**Figure 1**). AI/ML tools have already streamlined variant calling by reducing manual filtering and improving scalability. Examples of this include Google's DeepVariant(Poplin et al. 2018), DNAscope(Freed et al. 2022; Hu et al. 2025), DeepTrio(Hu et al. 2022), Clair3(Su et al. 2022), Medaka(Nagy et al. 2026), and HELLO(Ramachandran et al. 2021). These tools offer speed and generalizability across sequencing platforms(Abdelwahab et al. 2023; Brand et al. 2024; Abdelwahab and Torkamaneh 2025). Following variant identification, variant annotation contextualizes a patient's variants using sequence data, conservation, population frequency, and functional impact. This step requires synthesizing information across diverse databases. LLMs, a subtype of AI models that process and generate human language(2024), excel at automating this process. Mining resources like ClinVar and gnomAD (i.e., large databases of patient variants) have been assessed in the context of their genetic sequences to predict a variant's consequences on the primary structure (e.g., SpliceAI, AlphaMissense, and Evo2)(Tordai et al. 2024; Brixi et al. 2026). Other ML models have enhanced variant annotation through feature-based learning (e.g., REVEL, CADD, PrimateAI-3D)(Kircher et al. 2014; Ioannidis et al. 2016; Gao et al. 2023).

Full-stack variant analysis pipelines, including AI-MARRVEL(Mao et al. 2024), Qiagen's Franklin(2025a), Illumina's Emedgene(Meng et al. 2023), and Nostos Genomics(2025b), have already automated variant interpretation and prioritization. Despite these advances, variants of uncertain significance (VUSs) remain the most common variant classification, accounting for ~35-37% of variants associated with rare diseases and cancer(Balmaña et al. 2016; Fowler and Rehm 2024; Zawar et al. 2025). This ambiguity presents a critical clinical challenge; non-experts may misinterpret a VUS as 'normal' (false reassurance) or as a definitive diagnosis (unnecessary anxiety), leading to inappropriate care(Campeau 2022). Reclassification is inherently difficult, as assigning a variant to benign, likely benign, likely pathogenic, or pathogenic annotations requires ≥90% certainty of its clinical relevance(Richards et al. 2015). This threshold is challenging to meet, especially when context-specific data are limited and/or when considering non-coding (e.g., regulatory sequences(Avsec et al. 2021) and splice sites(Jaganathan et al. 2019)), low-penetrance, or hypomorphic variants(Richards et al. 2015; Fiorini et al. 2023). Emerging tools aim to address this, such as DYNA, a disease-specific LLM that compares context-specific networks to score the pathogenicity of coding and non-coding variants(Zhan and Zhang 2024). In a study of >17k cardiomyopathy VUSs from ClinVar, DYNA reclassified ~9% as pathogenic, likely pathogenic, benign, or likely benign(Zhan and Zhang 2024). Another promising approach to improving classification is to estimate penetrance. In rare diseases, small cohorts make it difficult, or even impossible, to calculate penetrance using traditional methods.

 doi:10.20944/preprints202603.1025.v2

However, Forrest et al.(Forrest et al. 2025), developed disease-specific ML models to calculate disease probability and penetrance using EHR and genetic data.

AI-assisted variant analysis can clarify genetic test results (e.g., AI-enabled ACMG scoring within EHR and clinical trial eligibility screening(Jin et al. 2024)), enabling clinicians to weigh genomic evidence alongside clinical findings. With data-driven rationales to support clinical diagnostics, clinicians are better-equipped to make more efficient and accurate decisions. Clinicians can thereby reduce trial-and-error prescribing by linking variants to targeted therapies and trials. Ultimately, AI-assistance will increase genetic screening rates, preventing delays in care.

Challenges and Recommendations

Integrating AI into clinical genetics shows great promise, but we expect challenges ahead (**Table 1**).

Table 1. AI integration challenges and recommendations in clinical genetics.

Challenge	Recommendation	Impact
Privacy & Safety	Adhere to *GDPR, *HIPPA, *ISO/IEC, and *GINA; use secure data handling practices	Protect sensitive information and maintain patient trust
Data Quality & Bias	Use high-quality, representative datasets; avoid "big data hubris"	Reduce bias, improve prediction accuracy, and ensure fairness
Model Transparency	Incorporate explainable AI (XAI) methods; ensure models are auditable	Improve trust, interpretability, and ethical accountability
Validation & Life Cycle	Implement post-market testing and total product life cycle monitoring	Ensure ongoing efficacy and safety of AI tools

The table summarizes potential challenges using our proposed AI-assisted approach, along with recommended solutions and their expected clinical impact. *General Data Protection Regulation (GDPR), *Health Insurance Portability and Accountability Act (HIPPA), *International Organization for Standardization (ISO), *International Electrotechnical Commission (IEC), *Genetic Information Nondiscrimination Act of 2008 (GINA).

Trust in scientists is declining in the US(Kennedy and Tyson 2023), and global opinion toward AI remains cautious(Poushter et al. 2025). To restore public confidence, developers should collaborate with patients and clinicians when designing AI tools, leveraging their domain-specific expertise to improve model performance and ensure relevance(Erikson 2018; Tomašev et al. 2020).

Genetic data has historically raised significant legal, ethical, and privacy concerns due to its uniquely identifiable nature. Using this data with AI could raise additional concerns; therefore, training data and software must comply with national/international laws and standards(Office for Civil Rights (OCR); Ruiz; 2008; European Union 2016; International Organization for Standardization, International Electrotechnical Commission 2022; Sokhansanj and Rosen 2025). Models for variant analysis should also adhere to established clinical standards from reputable organizations, such as Human Genome Variation Society (HGVS)(den Dunnen et al. 2016), ACMG, AMP, CAP(Richards et al. 2015), and ESHG(Houge et al. 2022, 2024).

 doi:10.20944/preprints202603.1025.v2

A major shortcoming of many AI tools stems from the data they are trained on. Overreliance on large, uncurated datasets can introduce bias, inaccuracies, and outdated information, leading to large errors in predictions(Lazer et al. 2014; Kessler et al. 2016; Ross and Swetlitz 2018; Xing et al. 2025; Fieldhouse 2025) and AI "hallucinations"(Beutel et al. 2023). Instead, datasets should be reliable and representative of the affected patient population(Tomašev et al. 2020; Nakayama et al. 2022; Daneshjou et al. 2022; Delgado et al. 2022; 2023; Center for Devices and Radiological Health 2025a; McCoy et al. 2025). This is especially critical in biomedical applications, where underrepresentation can perpetuate disparities(Larson et al. 2016; Diaz et al. 2018; Dastin 2018; Obermeyer et al. 2019; Nakayama et al. 2022; Daneshjou et al. 2022; Delgado et al. 2022). However, implementing retrieval-augmented generation (RAG) systems (curated knowledge bases) has already aided biomedical applications and reduced AI hallucinations(Lee et al. 2024; Leiser et al. 2025).

ML/AI models offer powerful capabilities for streamlining variant analysis by integrating multimodal data (e.g., genetic sequences, EHRs, biomedical knowledge graphs, and large-scale text mining) but often at the cost of interpretability, with many functioning as a "black box"(Ruiz; Gosiewska et al. 2021). To ensure fairness and accuracy, especially in clinical contexts, models must be auditable and explainable. An auditable model acts as a "glass box," where processes can be systematically examined and traced (e.g., by logging decision logic(Sina Gräupner et al. 2023) and data sources used as evidence(Mercurio et al. 2022; Meng et al. 2023; Allot et al. 2023; 2025b) (May et al. 2022; Sina Gräupner et al. 2023). Explainable AI (XAI) techniques further enable users to dissect models and their predictions to assess the influence of individual features. Numerous XAI approaches are currently available–even for complex LLMs–despite their scale of parameters and training(Zhao et al. 2024; Chen et al. 2024; Peng et al. 2025). Some AI-assisted variant analyses and workflows already incorporate explainable AI (XAI) methods, such as scoring and ranking the importance of features that drive their predictions(Meng et al. 2023; 2025c; Forrest et al. 2025)(Lundberg and Lee 2017).

Confirming the correctness and translatability of AI-prioritized variants requires multi-tiered validation and continual monitoring. Models must be benchmarked and tested against high-quality, expert-curated datasets (e.g., ClinVar or specific disease cohorts) to ensure high sensitivity (>90%) in real-world scenarios(2025c), and predictions should be verified through orthogonal biological tests. Potential orthogonal evidence-based methods include segregation analysis(Kim et al. 2019), confirming variant tracks with phenotypes in a family, and *in vivo* or *in vitro* functional assays(Kim et al. 2019; Agaoglu et al. 2022), providing experimental evidence supporting variant damage to a gene product. AI tools should follow a full product lifecycle approach, including international predetermined change control plans (PCCPs) for ML-enabled medical devices(Center for Devices and Radiological Health 2025b), with real-world performance tracked for safety and efficacy. As models evolve, outputs may change and even contradict earlier reports; this should be expected and documented so clinicians and patients can modify care as needed(Center for Devices and Radiological Health 2025b).

Conclusions

Incorporating ML and AI into variant analysis can transform and expedite the genetic testing process with actionable clinical intelligence, enabling earlier diagnostics and potentially life-saving interventions. When designed with transparency and community engagement, these tools accelerate variant interpretation without compromising clinical judgment or patient trust. By prioritizing ethical design, high-quality data, and explainable models, AI-assisted genomics advances the principle of beneficence by improving accuracy and efficiency, ultimately improving long-term patient outcomes.

Funding: This work was funded by the UAB Pilot Center for Precision Animal Modeling (C-PAM) (U54-OD030167), the UAB Childhood Cystic Kidney Disease Center (UAB-CCKDC) - Informatic and Data Analytics Resource (U54-DK126087), and M.M. was in part supported by the U.S. Department of Veterans Affairs (1-I01-BX006266-01).

 doi:10.20944/preprints202603.1025.v2

Competing Interests: MM reports additional research support from Otsuka, Sanofi, Vertex, and AbbVie, and has been a member of advisory boards for Sanofi, Santa Barbara Nutrients, and PKD Foundation. MM has provided consultancy for Otsuka, Sanofi, Vertex, AbbVie, and Regulus.

Author Contributions: This work was conceptualized by BNL, MM, EJW; literature search and analysis by EJW and ST; original manuscript drafted by EJW and ST; visualization by ST with help from EJW; critical manuscript revisions by TCH, ABC, MM, and BNL; manuscript edits by EJW, ST, and BNL; supervision and project management under BNL and EJW, funding acquisition by BNL and MM. All authors read and approved the final manuscript.

Acknowledgments: We thank the Lasseigne Lab members, including Tabea M. Soelter and Jaclyn M. Freeman, for their thoughtful input and discussion, as well as Brandon M. Wilk. Additionally, we are grateful to our funders for this work: the UAB Pilot Center for Precision Animal Modeling (C-PAM) (U54-OD030167), the UAB Childhood Cystic Kidney Disease Center (UAB-CCKDC) - Informatic and Data Analytics Resource (U54-DK126087), and the U.S. Department of Veterans Affairs (1-I01-BX006266-01).

References

Abdelwahab O, Belzile F, Torkamaneh D (2023) Performance analysis of conventional and AI-based variant callers using short and long reads. BMC Bioinformatics 24:472. https://doi.org/10.1186/s12859-023-05596-3

Abdelwahab O, Torkamaneh D (2025) Artificial intelligence in variant calling: a review. Front Bioinform 5:1574359. https://doi.org/10.3389/fbinf.2025.1574359

Agaoglu NB, Unal B, Akgun Dogan O, et al (2022) Consistency of variant interpretations among bioinformaticians and clinical geneticists in hereditary cancer panels. Eur J Hum Genet 30:378–383. https://doi.org/10.1038/s41431-022-01060-7

Allot A, Wei C-H, Phan L, et al (2023) Tracking genetic variants in the biomedical literature using LitVar 2.0. Nat Genet 55:901–903. https://doi.org/10.1038/s41588-023-01414-x

Avsec Ž, Agarwal V, Visentin D, et al (2021) Effective gene expression prediction from sequence by integrating long-range interactions. Nat Methods 18:1196–1203. https://doi.org/10.1038/s41592-021-01252-x

Balmaña J, Digiovanni L, Gaddam P, et al (2016) Conflicting interpretation of genetic variants and cancer risk by commercial laboratories as assessed by the Prospective Registry of Multiplex Testing. J Clin Oncol 34:4071–4078. https://doi.org/10.1200/JCO.2016.68.4316

Beutel G, Geerits E, Kielstein JT (2023) Artificial hallucination: GPT on LSD? Crit Care 27:148. https://doi.org/10.1186/s13054-023-04425-6

Brand F, Guski J, Krawitz P (2024) Extending DeepTrio for sensitive detection of complex de novo mutation patterns. NAR Genom Bioinform 6:lqae013. https://doi.org/10.1093/nargab/lqae013

Brixi G, Durrant MG, Ku J, et al (2026) Genome modelling and design across all domains of life with Evo 2. Nature 1–13. https://doi.org/10.1038/s41586-026-10176-5

Campeau PM (2022) An all-encompassing variant classification system proposed. Eur J Hum Genet 30:139. https://doi.org/10.1038/s41431-021-00992-w

Center for Devices, Radiological Health (2025a) Good Machine Learning Practice for Medical Device Development: Guiding Principles. In: U.S. Food and Drug Administration. https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles. Accessed 7 Oct 2025

Center for Devices, Radiological Health (2025b) Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices: Guiding Principles. In: U.S. Food and Drug Administration. https://www.fda.gov/medical-devices/software-medical-device-samd/predetermined-change-control-plans-machine-learning-enabled-medical-devices-guiding-principles. Accessed 7 Oct 2025

Chen X, Wang L, You M, et al (2024) Evaluating and enhancing large language models' performance in domain-specific medicine: Development and usability study with DocOA. J Med Internet Res 26:e58158. https://doi.org/10.2196/58158

Daneshjou R, Vodrahalli K, Novoa RA, et al (2022) Disparities in dermatology AI performance on a diverse, curated clinical image set. Sci Adv 8:eabq6147. https://doi.org/10.1126/sciadv.abq6147

 doi:10.20944/preprints202603.1025.v2

Dastin J (2018) Insight - Amazon scraps secret AI recruiting tool that showed bias against women. Reuters

Delgado J, de Manuel A, Parra I, et al (2022) Bias in algorithms of AI systems developed for COVID-19: A scoping review. J Bioeth Inq 19:407–419. https://doi.org/10.1007/s11673-022-10200-z

Diaz M, Johnson I, Lazar A, et al (2018) Addressing age-related bias in sentiment analysis. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA

Erikson SL (2018) Cell phones ≠ self and other problems with big data detection and containment during epidemics. Med Anthropol Q 32:315–339. https://doi.org/10.1111/maq.12440

European Union (2016) Regulation (EU) 2016/679 General Data Protection Regulation (GDPR). Official Journal of the European Union

Faye F, Crocione C, Anido de Peña R, et al (2024) Time to diagnosis and determinants of diagnostic delays of people living with a rare disease: results of a Rare Barometer retrospective patient survey. Eur J Hum Genet 32:1116–1126. https://doi.org/10.1038/s41431-024-01604-z

Fieldhouse R (2025) Too much social media gives AI chatbots "brain rot." Nature. https://doi.org/10.1038/d41586-025-03542-2

Fiorini MR, Dilliott AA, Farhan SMK (2023) Evaluating the utility of REVEL and CADD for interpreting variants in amyotrophic lateral sclerosis genes. Hum Mutat 2023:8620557. https://doi.org/10.1155/2023/8620557

Forrest IS, Vy HMT, Rocheleau G, et al (2025) Machine learning-based penetrance of genetic variants. Science 389:eadm7066. https://doi.org/10.1126/science.adm7066

Fowler DM, Rehm HL (2024) Will variants of uncertain significance still exist in 2030? Am J Hum Genet 111:5–10. https://doi.org/10.1016/j.ajhg.2023.11.005

Freed D, Pan R, Chen H, et al (2022) DNAscope: High accuracy small variant calling using machine learning. bioRxiv 2022.05.20.492556

Gao H, Hamp T, Ede J, et al (2023) The landscape of tolerated genetic variation in humans and primates. Science 380:eabn8153. https://doi.org/10.1126/science.abn8197

Gosiewska A, Kozak A, Biecek P (2021) Simpler is better: Lifting interpretability-performance trade-off via automated feature engineering. Decis Support Syst 150:113556. https://doi.org/10.1016/j.dss.2021.113556

Gurovich Y, Hanani Y, Bar O, et al (2019) Identifying facial phenotypes of genetic disorders using deep learning. Nat Med 25:60–64. https://doi.org/10.1038/s41591-018-0279-0

Hajek C, Hutchinson AM, Galbraith LN, et al (2022) Improved provider preparedness through an 8-part genetics and genomic education program. Genet Med 24:214–224. https://doi.org/10.1016/j.gim.2021.08.008

Houge G, Bratland E, Aukrust I, et al (2024) Comparison of the ABC and ACMG systems for variant classification. Eur J Hum Genet 32:858–863. https://doi.org/10.1038/s41431-024-01617-8

Houge G, Laner A, Cirak S, et al (2022) Stepwise ABC system for classification of any type of genetic variant. Eur J Hum Genet 30:150–159. https://doi.org/10.1038/s41431-021-00903-z

Hu J, Freed D, Feng H, et al (2025) Accelerated, Accurate, Hybrid Short and Long Reads Alignment and Variant Calling. Bioinformatics

Hu X, Feng C, Zhou Y, et al (2022) DeepTrio: a ternary prediction system for protein-protein interaction using mask multiple parallel convolutional neural networks. Bioinformatics 38:694–702. https://doi.org/10.1093/bioinformatics/btab737

International Organization for Standardization, International Electrotechnical Commission (2022) ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements. ISO, Geneva

Ioannidis NM, Rothstein JH, Pejaver V, et al (2016) REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet 99:877–885. https://doi.org/10.1016/j.ajhg.2016.08.016

Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al (2019) Predicting splicing from primary sequence with deep learning. Cell 176:535–548.e24. https://doi.org/10.1016/j.cell.2018.12.015

Janiesch C, Zschech P, Heinrich K (2021) Machine learning and deep learning. Electron Mark 31:685–695. https://doi.org/10.1007/s12525-021-00475-2

Jin Q, Wang Z, Floudas CS, et al (2024) Matching patients to clinical trials with large language models. Nat Commun 15:9074. https://doi.org/10.1038/s41467-024-53081-z

 doi:10.20944/preprints202603.1025.v2

Kennedy B, Tyson A (2023) Americans' Trust in Scientists, Positive Views of Science Continue to Decline. In: Pew Research Center. https://www.pewresearch.org/science/2023/11/14/confidence-in-scientists-medical-scientists-and-other-groups-and-institutions-in-society/. Accessed 6 Nov 2025

Kessler MD, Yerges-Armstrong L, Taub MA, et al (2016) Challenges and disparities in the application of personalized genomic medicine to populations with African ancestry. Nat Commun 7:12521. https://doi.org/10.1038/ncomms12521

Kim YE, Ki CS, Jang MA (2019) Challenges and considerations in sequence variant interpretation for Mendelian disorders. Ann Lab Med 39:421–429. https://doi.org/10.3343/alm.2019.39.5.421

Kircher M, Witten DM, Jain P, et al (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46:310–315. https://doi.org/10.1038/ng.2892

Kneifati-Hayek JZ, Zachariah T, Ahn W, et al (2024) Bridging the gap in genomic implementation: Identifying user needs for precision nephrology. Kidney Int Rep 9:2420–2431. https://doi.org/10.1016/j.ekir.2024.05.032

Koteluk O, Wartecki A, Mazurek S, et al (2021) How do machines learn? Artificial intelligence as a New Era in medicine. J Pers Med 11:32. https://doi.org/10.3390/jpm11010032

Kris A. Wetterstrand MS (2019) DNA Sequencing Costs: Data. In: Genome.gov. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data. Accessed 5 Nov 2025

Larson J, Angwin J, Kirchner L, Mattu S (2016) How We Analyzed the COMPAS Recidivism Algorithm. In: #creator. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm. Accessed 7 Oct 2025

Lazer D, Kennedy R, King G, Vespignani A (2014) Big data. The parable of Google Flu: traps in big data analysis. Science 343:1203–1205. https://doi.org/10.1126/science.1248506

Lee J, Cha H, Hwangbo Y, Cheon W (2024) Enhancing large language model reliability: Minimizing hallucinations with dual retrieval-augmented generation based on the latest diabetes guidelines. J Pers Med 14:1131. https://doi.org/10.3390/jpm14121131

Leiser F, Guse R, Sunyaev A (2025) Large language model architectures in health care: Scoping review of research perspectives. J Med Internet Res 27:e70315. https://doi.org/10.2196/70315

Lewin Group T (2023) The cost of delayed diagnosis in rare disease: a health economic study. EveryLife Foundation for Rare Diseases

Lin L, Pan H, Qi Y, et al (2023) Reasons and Resolutions for Inconsistent Variant Interpretation. Human Mutation 2023:1–11. https://doi.org/10.1155/2023/4955235

Lundberg S, Lee S-I (2017) A unified approach to interpreting model predictions. arXiv [cs.AI]

Mao D, Liu C, Wang L, et al (2024) AI-MARRVEL - A knowledge-driven AI system for diagnosing Mendelian disorders. NEJM AI 1.: https://doi.org/10.1056/aioa2300009

May W, Berghoff C, Böddinghaus J, et al (2022) May 2022 Towards Auditable AI Systems From Principles to Practice. https://www.semanticscholar.org/paper/Whitepaper-%7C-May-2022-Towards-Auditable-AI-Systems-May-Berghoff/2af03e944a0c0e905ccf1b24e795f53cc780ea36#related-papers. Accessed 22 Oct 2025

McCoy LG, Bihorac A, Celi LA, et al (2025) Building health systems capable of leveraging AI: applying Paul Farmer's 5S framework for equitable global health. BMC Glob Public Health 3:39. https://doi.org/10.1186/s44263-025-00158-6

McNeill A (2022) A new system for variant classification? Eur J Hum Genet 30:137–138. https://doi.org/10.1038/s41431-021-01032-3

Meng L, Attali R, Talmy T, et al (2023) Evaluation of an automated genome interpretation model for rare disease routinely used in a clinical genetic laboratory. Genet Med 25:100830. https://doi.org/10.1016/j.gim.2023.100830

Mercurio SA, Chunn LM, Khursigara G, et al (2022) ENPP1 deficiency: A clinical update on the relevance of individual variants using a locus-specific patient database. Hum Mutat 43:1673–1705. https://doi.org/10.1002/humu.24477

Nagy D, Pennetta V, Rodger G, et al (2026) Nanopore long-read-only genome assembly of clinical Enterobacterales isolates is complete and accurate. Microb Genom 12:001631. https://doi.org/10.1099/mgen.0.001631

 doi:10.20944/preprints202603.1025.v2

Nakayama LF, Kras A, Ribeiro LZ, et al (2022) Global disparity bias in ophthalmology artificial intelligence applications. BMJ Health Care Inform 29:e100470. https://doi.org/10.1136/bmjhci-2021-100470

Nichols JA, Herbert Chan HW, Baker MAB (2019) Machine learning: applications of artificial intelligence to imaging and diagnosis. Biophys Rev 11:111–118. https://doi.org/10.1007/s12551-018-0449-9

Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. Science 366:447–453. https://doi.org/10.1126/science.aax2342

Office for Civil Rights (OCR) Summary of the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html. Accessed 16 Mar 2022

Pasquier L, Minguet G, Moisdon-Chataigner S, et al (2022) How do non-geneticist physicians deal with genetic tests? A qualitative analysis. Eur J Hum Genet 30:320–331. https://doi.org/10.1038/s41431-021-00884-z

Peabody J, DeMaria L, Tamandong-LaChica D, et al (2015) Low rates of genetic testing in children with developmental delays, intellectual disability, and autism spectrum disorders. Glob Pediatr Health 2:2333794X15623717. https://doi.org/10.1177/2333794X15623717

Peng M, Guo X, Chen X, et al (2025) LC-LLM: Explainable lane-change intention and trajectory predictions with Large Language Models. Communications in Transportation Research 5:100170. https://doi.org/10.1016/j.commtr.2025.100170

Phillips C, Parkinson A, Namsrai T, et al (2024) Time to diagnosis for a rare disease: managing medical uncertainty. A qualitative study. Orphanet J Rare Dis 19:297. https://doi.org/10.1186/s13023-024-03319-2

Poplin R, Chang P-C, Alexander D, et al (2018) A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol 36:983–987. https://doi.org/10.1038/nbt.4235

Poushter J, Fagan M, Corichi M (2025) How People Around the World View AI. In: Pew Research Center. https://www.pewresearch.org/global/2025/10/15/how-people-around-the-world-view-ai/. Accessed 6 Nov 2025

Ramachandran A, Lumetta SS, Klee EW, Chen D (2021) HELLO: improved neural network architectures and methodologies for small variant calling. BMC Bioinformatics 22:404. https://doi.org/10.1186/s12859-021-04311-4

Rasouly HM, Balderes O, Marasa M, et al (2023) The effect of genetic education on the referral of patients to genetic evaluation: Findings from a national survey of nephrologists. Genet Med 25:100814. https://doi.org/10.1016/j.gim.2023.100814

Richards S, Aziz N, Bale S, et al (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405–424. https://doi.org/10.1038/gim.2015.30

Ross C, Swetlitz I (2018) IBM's Watson supercomputer recommended "unsafe and incorrect" cancer treatments, internal documents show. In: STAT. https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/. Accessed 7 Oct 2025

Ruiz J Machine learning and the right to explanation in GDPR. In: Open Rights Group. https://www.openrightsgroup.org/blog/machine-learning-and-the-right-to-explanation-in-gdpr/. Accessed 7 Oct 2025

Russell S, Norvig P (2021) Artificial intelligence: A modern approach, global edition, 4th edn. Pearson Education, London, England

Sina Gräupner O, Pawlaszczyk D, Hummert C (2023) Basics of Auditable AI Systems. In: 2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE). IEEE, pp 2355–2362

Sokhansanj BA, Rosen GL (2025) Regulating genome language models: navigating policy challenges at the intersection of AI and genetics. Hum Genet 144:949–970. https://doi.org/10.1007/s00439-025-02768-4

Su J, Zheng Z, Ahmed SS, et al (2022) Clair3-trio: high-performance Nanopore long-read variant calling in family trios with trio-to-trio deep neural networks. Brief Bioinform 23:bbac301. https://doi.org/10.1093/bib/bbac301

Tagliafico E, Bernardis I, Grasso M, et al (2018) Workload measurement for molecular genetics laboratory: A survey study. PLoS One 13:e0206855. https://doi.org/10.1371/journal.pone.0206855

Tomašev N, Cornebise J, Hutter F, et al (2020) AI for social good: unlocking the opportunity for positive impact. Nat Commun 11:2468. https://doi.org/10.1038/s41467-020-15871-z

 doi:10.20944/preprints202603.1025.v2

Tordai H, Torres O, Csepi M, et al (2024) Analysis of AlphaMissense data in different protein groups and structural context. Sci Data 11:495. https://doi.org/10.1038/s41597-024-03327-8

Wan EL, Elkaim Y, Gao W, Yoon R (2023) Zebras among us: Advocating for the 30 million Americans living with rare disease. Med Sci Educ 33:1239–1242. https://doi.org/10.1007/s40670-023-01856-2

Xing S, Hong J, Wang Y, et al (2025) LLMs can get "Brain Rot"! arXiv [cs.CL]

Yang X, Chen A, PourNejatian N, et al (2022) A large language model for electronic health records. NPJ Digit Med 5:194. https://doi.org/10.1038/s41746-022-00742-2

Ye J, Woods D, Jordan N, Starren J (2024) The role of artificial intelligence for the application of integrating electronic health records and patient-generated data in clinical decision support. AMIA Summits Transl Sci Proc 2024:459–467

Zawar A, Manoj G, Nair PP, et al (2025) Variants of uncertain significance: At the crux of diagnostic odyssey. Gene 962:149587. https://doi.org/10.1016/j.gene.2025.149587

Zhan H, Zhang Z (2024) DYNA: Disease-specific language model for variant pathogenicity. arXiv [q-bio.GN]

Zhao H, Chen H, Yang F, et al (2024) Explainability for large language models: A survey. ACM Trans Intell Syst Technol 15:1–38. https://doi.org/10.1145/3639372

Zukin E, Culver JO, Liu Y, et al (2023) Clinical implications of conflicting variant interpretations in the cancer genetics clinic. Genet Med 25:100837. https://doi.org/10.1016/j.gim.2023.100837

(2024) Machine Learning vs Deep Learning vs LLMs vs GenAI: Explained and How are they Different from Each Other? In: Cloud4C. https://www.cloud4c.com/blogs/genai-vs-machine-learning-vs-deep-learning-vs-llms. Accessed 14 Nov 2025

(2025a) Franklin - Bioinformatics Software. In: Bioinformatics Software | QIAGEN Digital Insights. https://digitalinsights.qiagen.com/franklin/. Accessed 14 Oct 2025

(2025b) Enhancing Rare Disease Diagnostics: Updated Performance Evaluation of the AION AI-Driven Variant Interpretation Platform. In: https://www.nostos-genomics.com. https://www.nostos-genomics.com/news/enhancing-rare-disease-diagnostics-updated-performance-evaluation-of-the-aion-ai-driven-variant-interpretation-platform. Accessed 30 Mar 2026

(2008) Genetic Information Nondiscrimination Act of 2008 (GINA)

(2023) Annual Report 2023. Partners in Health

(2025c) AI & Variant Interpretation: From Data Tsunami to Diagnostic Clarity. In: https://www.nostos-genomics.com. https://www.nostos-genomics.com/news/ai-and-variant-interpretation-from-data-tsunami-to-diagnostic-clarity. Accessed 7 Oct 2025