

Review

Not peer-reviewed version

Artificial Intelligence in the Diagnosis of Pediatric Rare Diseases: From Real World Data Towards a Personalized Medicine Approach

[Nikola Ilic](#) * and [Adrijan Sarajlija](#)

Posted Date: 17 July 2025

doi: 10.20944/preprints202507.1405.v1

Keywords: artificial intelligence (AI); pediatric rare diseases; genomic diagnostics; personalized medicine; large language models (LLMs); real-world data; ethical considerations



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Artificial Intelligence in the Diagnosis of Pediatric Rare Diseases: From Real World Data Towards a Personalized Medicine Approach

Nikola Ilić^{1,*} and Adrijan Sarajlija^{1,2,3}

¹ Clinical Genetics Outpatient Clinic, Mother and Child Health Care Institute of Serbia "Dr Vukan Cupic", 11070 Belgrade, Serbia

² Department of Pediatrics, Faculty of Medicine, University of Belgrade, 11000 Belgrade, Serbia

³ Faculty of Medicine, University of Eastern Sarajevo, 73300 Foča, Bosnia and Herzegovina

* Correspondence: nikola.ilic@imd.org.rs

Abstract

Background: Artificial intelligence (AI) is increasingly used in the diagnosis of pediatric rare diseases, helping to improve the speed, accuracy, and accessibility of genetic interpretations. This development supports the ongoing shift toward personalized medicine in clinical genetics. **Objective:** This review explores current applications of AI in pediatric rare disease diagnostics, with an emphasis on real-world data integration and implications for individualized care. **Methods:** A narrative review was conducted covering AI tools for variant prioritization, phenotype-genotype correlation, large language models (LLMs), and ethical considerations. Literature was identified through PubMed, Scopus, and Web of Science up to July 2025. **Results:** AI platforms offer promising support for genomic interpretation, especially in structured diagnostic workflows. Tools integrating HPO-based inputs and LLMs enable phenotype matching and support reverse phenotyping. Real-world data enhance AI's applicability in complex, heterogeneous cases. However, challenges remain regarding data standardization, interpretability, workflow integration, and bias. **Conclusion:** AI has the potential to support earlier and more personalized diagnostics for children with rare diseases. To fully realize this, multidisciplinary collaboration and careful attention to clinical, technical, and ethical considerations are essential.

Keywords: artificial intelligence (AI); pediatric rare diseases; genomic diagnostics; personalized medicine; large language models (LLMs); real-world data; ethical considerations

1. Introduction

Rare diseases affect millions of children worldwide. Although each individual condition is uncommon, together they form a significant clinical and public health challenge. For many families, the path to diagnosis, often referred to as the diagnostic odyssey, is long, uncertain, and emotionally draining [1]. According to a 2017 EURORDIS survey, the average diagnostic journey for rare disease patients spans 5–7 years, often involving multiple misdiagnoses and fragmented care [2]. In pediatric practice, rare diseases frequently manifest as unexplained developmental delays, multisystem anomalies, or atypical syndromes. Over 70% have a genetic basis, and most begin in early childhood. Timely and accurate diagnosis is essential not only for therapeutic decisions but also for genetic counseling, family planning, and prognosis [3].

Yet, even with the widespread availability of next-generation sequencing technologies, such as whole-exome (WES) and whole-genome sequencing (WGS), interpreting the massive volume of genetic data remains a critical bottleneck. The shortage of trained clinical geneticists, variability in phenotypic documentation, and fragmented health information systems further delay diagnosis or

lead to missed opportunities. These challenges are especially pronounced in low-resource settings or outside tertiary centers [4].

In response, artificial intelligence (AI) has emerged as a promising tool to support rare disease diagnostics. From variant interpretation to phenotype-genotype correlation, modern AI systems—particularly machine learning platforms and large language models—offer new ways to prioritize findings, generate differential diagnoses, and reduce the cognitive burden on clinicians. Their scalability and adaptability make them attractive both in high-resource institutions and in settings with limited specialist access [5].

This review explores the current landscape of AI applications in the diagnosis of pediatric rare diseases, with a main focus on their integration into personalized medicine. We highlight recent comparative evidence in the diagnostic performance of AI vs human experts—both from structured studies and real-world clinical applications. In doing so, we examine not only the theoretical potential of AI in genomic interpretation and phenotype-genotype correlation, but also how these tools function under real-world conditions. Furthermore, we discuss the practical, ethical, and technical challenges associated with implementing AI technologies in pediatric care, particularly in settings where infrastructure, training, and data standardization remain variable.

Ultimately, we aim to explore how AI can serve not as a replacement but as a valuable ally in achieving faster, more accurate, and personalized care for children with rare conditions.

2. Materials and Methods

This narrative review was conducted to synthesize current knowledge on the application of AI in the diagnosis of pediatric rare diseases. We conducted a narrative review of English-language publications from 2018 to July 2025 using the keywords 'artificial intelligence', 'rare diseases', 'pediatrics', 'genomic interpretation', and 'LLM(s)'. Studies were included if they directly addressed AI applications in pediatric diagnosis or variant interpretation workflows. Priority was given to recent systematic reviews, original research articles, and landmark studies regarding the aim of this review. Additional references were identified through manual searches of cited literature in key articles. No formal systematic review methodology was applied, as the aim was to provide a comprehensive narrative synthesis rather than an exhaustive systematic analysis.

3. Artificial Intelligence in the Diagnosis of Pediatric Rare Diseases

3.1. The Role of AI in Genomic Data Interpretation

The advent of NGS has revolutionized the diagnostic landscape of genetic medicine, particularly in the field of rare diseases. By enabling high-throughput parallel sequencing, NGS allows simultaneous analysis of hundreds or thousands of genes. Consequently, NGS has greatly enhanced diagnostic efficiency and helped expand the horizons of personalized medicine. Diagnostic yields have increased significantly, with reported rates ranging from 40% to over 70%, depending on the specific patient cohort and clinical context. As a result, gene panels and whole-exome sequencing (WES) have become the standard of care in many tertiary centers worldwide [6,7].

Crucially, NGS has shifted the diagnostic paradigm from a traditional linear phenotype-to-genotype workflow toward a more integrated and bidirectional approach. Molecular findings can help to refine or even redefine clinical hypotheses, and clinical context shapes the interpretation of genetic data. This process, often termed reverse phenotyping (RF), has proven particularly valuable in rare diseases characterized by atypical, evolving, or overlapping features where classical syndromic recognition may fall short of establishing a definitive diagnosis [8].

However, despite its transformative power, NGS is not without limitations. The technology frequently generates variants of uncertain significance (VUS), uncovers incidental or secondary findings, and may detect mutations in genes not previously linked to the suspected phenotype. Furthermore, certain categories of pathogenic alterations (including deep intronic variants, structural rearrangements, epimutations, and complex copy number variants) may be missed by exome-based

approaches. This brings the necessity for the application of whole-genome sequencing (WGS) for more comprehensive coverage. Yet, the clinical implementation of WGS introduces additional layers of complexity, including data interpretation challenges and ethical considerations [9,10].

Importantly, even when a pathogenic or likely pathogenic variant is identified, assigning the correct diagnostic label remains a demanding task. Accurate interpretation requires molecular confirmation integrated with expert assessment of the patient's clinical features, laboratory data, imaging studies and global developmental trajectory or familial context. This is especially challenging in pediatric populations, where phenotypic expression may be incomplete or age-dependent, making classical diagnostic patterns harder to recognize [8,11].

Given these complex circumstances, multidisciplinary collaboration has become the cornerstone of rare disease diagnostics. Optimal interpretation requires close cooperation between clinical geneticists, subspecialty clinicians, molecular biologists, radiologists, and genetic counselors to contextualize genomic findings within the broader clinical picture. However, this model is resource-intensive and not always accessible in under-resourced settings [11].

3.1.1. AI Tools for Variant Interpretation

In response to these challenges, there is a growing effort for the development of advanced decision-support tools capable of assisting clinicians in the complex task of genomic interpretation [12] (Table 1).

Table 1. Applications of AI in Genomic Data Interpretation for Pediatric Rare Diseases.

Application Area	Description	Example Tools/Platforms
Variant Prioritization	Automated ranking of genetic variants based on pathogenicity predictions, allele frequencies, and gene-disease associations	MOON (Diploid), Fabric Genomics, Emedgene, GEM
Phenotype-Genotype Matching	Linking patient phenotypic features (HPO terms) to known gene-disease relationships	Phenomizer, GEM
Reverse Phenotyping	AI-driven re-evaluation of clinical features based on unexpected or novel genetic findings	LLM-assisted reverse phenotyping workflows
NLP ¹	Extracting structured phenotypic information from unstructured clinical notes	NLP modules integrated within genomic AI pipelines
Clinical Summarization & Decision Support	Generating diagnostic hypotheses and literature-informed interpretations	ChatGPT (OpenAI), DeepSeek Medical AI

¹ Natural Language Processing. **Note:** This table presents major application areas of AI in the interpretation of genomic data for pediatric rare diseases, with representative tools that illustrate the diversity of functions—from variant prioritization to natural language processing and clinical decision support.

Modern AI platforms can automate variant prioritization by integrating pathogenicity predictions, population frequency data, inheritance patterns, and gene-disease associations in real time. For example, tools such as MOON (by Diploid), Fabric Genomics, Emedgene, and GEM utilize phenotype-driven algorithms that match Human Phenotype Ontology (HPO) terms to known gene-disease relationships. This process can effectively narrow down candidate variants that require further clinician review [13].

3.1.2. Role of Large Language Models (LLMs)

A particularly promising subset of AI applications involves large language models (LLMs), which are capable of processing natural language input and generating human-like responses. These models, such as OpenAI's ChatGPT and DeepSeek, have demonstrated the ability to simulate

diagnostic reasoning and generate differential diagnoses when provided with detailed clinical input [5].

In the context of pediatric rare disease diagnostics, LLMs can assist in hypothesis generation, reinterpretation of existing genomic data, and even suggest plausible genetic conditions based on minimal input. Some studies and institutional experiences have explored their role in reverse phenotyping, particularly in ambiguous cases or where traditional phenotype-driven workflows have failed to yield actionable leads [14].

However, their clinical utility remains largely experimental. Unlike dedicated variant interpretation tools, LLMs are not yet optimized for structured genomic inputs and lack formal integration into diagnostic pipelines [5]. A key concern is the phenomenon of hallucination—where models produce outputs that are syntactically valid but factually incorrect. This has been extensively documented in recent evaluations of generative AI tools in medical and non-medical domains [15]. This is particularly problematic in rare disease diagnostics, where subtle nuances determine clinical decisions. Moreover, LLMs may reflect biases present in their training data, leading to overrepresentation of well-documented conditions and underrepresentation of phenotypic variability. These concerns mirror well-documented disparities observed in algorithmic performance when trained on demographically skewed datasets [16,17].

The lack of transparency in model architecture and training data further complicates their validation. Without rigorous benchmarking against expert consensus and real-world data, LLMs cannot yet be considered reliable standalone tools. Nevertheless, their potential to complement expert-driven interpretation is significant, particularly as user interfaces evolve and integration with electronic health records becomes feasible [18]

Furthermore, critical questions remain regarding how AI outputs compare to multidisciplinary human expert judgment, especially in a rare disease diagnostic scenario. Rigorous benchmarking studies are needed to evaluate AI models against both each other and established clinical standards, in realistic, patient-centered contexts [19].

Ultimately, LLMs represent a novel and dynamic frontier in AI-assisted medicine—but their integration into pediatric rare disease diagnostics demands caution, validation, and continual human oversight [20].

3.2. *Phenotype-Genotype Integration Through Automated Tools*

The diagnostic yield of genomic sequencing is strongly influenced by the quality, completeness, and granularity of phenotypic data provided alongside molecular analyses. In the field of pediatric rare disease diagnostics, clinical presentations are frequently complex, evolving, or syndromically overlapping. This makes accurate phenotyping as critical as the sequencing itself. Unlike adult-onset conditions, many pediatric disorders display age-dependent features, incomplete penetrance, or subtle morphological signs that may initially escape recognition. This diagnostic complexity often results in delayed or missed diagnoses, especially for rare conditions with variable expressivity [21].

Traditional approaches to phenotyping rely heavily on the clinician's expertise to recognize distinctive patterns and manually match them to known disorders. While experienced dysmorphologists and geneticists can achieve impressive diagnostic accuracy, this process is inherently subjective and variable. It depends not only on individual knowledge, but also on precise clinical measurements, standardized terminology, and systematic recording of subtle physical or developmental features. Small inconsistencies, such as imprecise anthropometric data, incomplete family history, or missing of seemingly minor anomalies, can significantly alter diagnostic pathways and interpretations [22].

AI-driven platforms have introduced an incredible innovation by enabling structured, scalable, and automated phenotype-genotype correlation. Central to this development is the widespread adoption of the Human Phenotype Ontology (HPO), which standardizes clinical features into a hierarchical and computable terminology. By encoding patient phenotypes using HPO terms,

clinicians can now input structured phenotypic profiles into AI tools, facilitating computational matching with curated disease databases [23,24].

Once phenotypic data are encoded, AI platforms apply algorithms that prioritize candidate genes and associated disorders based on semantic similarity, probabilistic modeling, and evidence-based knowledge graphs. This approach minimizes the biases and inconsistencies inherent in purely manual interpretation while maximizing diagnostic spread, especially for conditions with non-classical or overlapping phenotypes [25].

A growing array of tools exemplifies this integration: Phenomizer calculates statistical matches between patient phenotype sets and known Mendelian conditions, enabling rapid narrowing of diagnostic possibilities based on similarity scoring. GEM (Genetic Evaluation Module) combines phenotype scoring with variant pathogenicity analysis to generate ranked, integrated diagnostic hypotheses. Face2Gene, leveraging deep learning and facial recognition algorithms, analyzes facial morphology to suggest syndromic diagnoses, effectively serving as a digital dysmorphologist that supports clinician assessments. AMELIE mines the scientific literature to prioritize candidate genes based on both phenotypic features and gene relevance, bridging genomic data with up-to-date knowledge in the medical literature [26,27] (Table 2).

Table 2. Leading AI Tools for Phenotype-Genotype Integration in Pediatric Genomics.

Tool	Function	Integration	Validation Status
MOON	Variant prioritization based on phenotype-genotype correlation	Standalone; requires manual HPO input	Used in clinical diagnostics; validated in internal benchmarking
GEM	AI-based interpretation and scoring of variants	Integrated with Fabric Genomics platform	Deployed in hospital settings; comparative benchmarking with human panels
Phenomizer	Suggests differential diagnoses from HPO terms	Standalone; research use	Open-access tool; used in academic projects
Face2Gene	Image-based facial phenotype recognition	Mobile/web platform	High accuracy in syndromic conditions; not validated for nonsyndromic cases
Emedgene	AI-supported variant analysis with automated reporting	Commercial clinical platform	Regulatory-cleared in some jurisdictions; limited open-access data
DeepPhen	Phenotype-driven gene ranking using ML	Research-use; experimental	Experimental validation in selected cohorts

Note: The tools listed vary in terms of accessibility, integration into clinical pipelines, and robustness of validation. Selection is based on recent literature and institutional experience in pediatric genomics.

These AI tools consistently outperform traditional keyword searches or manual OMIM queries, significantly reducing time to diagnosis and expanding the diagnostic reach of clinicians. They are particularly valuable in resource-constrained settings, regional centers without subspecialty expertise, or scenarios where rapid diagnostic triage is essential [28].

An additional and rapidly evolving layer of genotype-phenotype integration involves LLMs. Unlike structured phenotype-driven algorithms alone, LLMs can flexibly interpret natural language descriptions of symptoms, examination findings, and historical narratives [29] (Table 3).

Table 3. Comparison of AI Approaches: Phenotype-Driven Algorithms vs Large Language Models.

Feature	Phenotype-Driven Algorithms	Large Language Models (LLMs)
---------	-----------------------------	------------------------------

Primary Input Type	Structured data (HPO terms)	Natural language, unstructured text
Strengths	Precise gene-disease matching, standardized outputs	Flexible interpretation, literature summarization, clinical reasoning
Limitations	Dependence on structured inputs, limited in ambiguous cases	Potential hallucinations, interpretability concerns
Examples	MOON, GEM, Phenomizer	ChatGPT, DeepSeek Medical AI
Ideal Use Case	Variant prioritization with detailed phenotypic data	Complex differential diagnosis, summarizing patient histories

Note: This table summarizes key distinctions between phenotype-driven algorithms and LLMs with regard to input type, strengths, limitations, and ideal use cases in rare disease diagnostics.

They can further map them to HPO-like representations or relevant diagnostic concepts. This allows AI to function effectively even when phenotypic data are unstructured, incomplete, or non-standardized, that is a frequent reality in everyday pediatric clinical practice [16].

Moreover, by integrating LLM capabilities with existing phenotype-genotype tools, AI systems can help synthesize diverse information streams. Structured HPO terms, free-text clinical notes, imaging descriptions, and literature evidence can all be transformed into coherent, ranked diagnostic suggestions. This holistic approach enhances both diagnostic efficiency and accuracy, reducing the work load on clinicians and minimizing the risk of missed rare diagnoses [19].

Ultimately, the automation and augmentation of phenotype-genotype correlation through AI-driven tools reinforce the foundation of personalized medicine in pediatric genetics. These tools are empowering clinicians to move beyond the limitations of human memory or single-specialist expertise. This approach offers open access to advanced diagnostic insights and paves the way toward faster, more accurate, and equitable care for children with rare diseases [24].

3.3. Real-World Data: Opportunities and Challenges for AI-Assisted Rare Disease Diagnosis

The diagnostic journey for children with rare diseases is often complex, prolonged, and nonlinear. For many families, it involves years of uncertainty, repeated hospital visitations, and inconclusive investigations. This process is known as the “diagnostic odyssey” [1]. Rare diseases by their nature frequently defy the rigid frameworks of randomized controlled trials (RCTs). While RCTs remain the gold standard for evaluating therapeutic interventions, they are often impractical or unsuitable for answering diagnostic questions in the rare disease setting [30].

In this context, real-world data (RWD) has emerged as an invaluable source of insight. Unlike RCTs, which depend on strict inclusion criteria, controlled environments, and predefined outcomes, RWD reflects the authentic complexity of clinical practice. It is longitudinal, multimodal, and captures how diseases manifest, progress, and respond to interventions in actual patients, not idealized study populations. RWD is generated from diverse sources, including electronic health records (EHRs), patient registries, diagnostic and genomic databases, administrative health records, and increasingly from digital health tools such as mobile applications and wearable devices [31].

Importantly, RWD encompasses a broader spectrum of clinical variability. This corpus of data includes patients that are often excluded from clinical trials due to comorbidities, atypical presentations, or age restrictions. This makes it especially valuable for understanding the full phenotypic spectrum of rare diseases, capturing early signs, variable expressivity, and real-life treatment responses across diverse pediatric populations [32].

However, the organic nature of RWD presents significant challenges. These data are often unstructured and heterogeneous. They are stored in free-text clinical notes, radiology reports, laboratory systems, or fragmented across institutional archives. Terminology may be inconsistent, documentation incomplete, and phenotypic descriptions scattered without standardized

frameworks. Genetic findings are frequently stored separately from phenotypic data, limiting integrated analysis. Consequently, the analytical utility of RWD is often underexploited [33].

AI is uniquely positioned to overcome these limitations. Modern AI tools, including natural language processing (NLP) and LLMs, can automatically extract structured information from unstructured text, map clinical observations to standardized terminologies such as Human Phenotype Ontology (HPO) terms, and harmonize data across disconnected datasets. Beyond extraction, AI systems can impute missing information, identify patterns across time, cluster patients based on phenotypic similarity, and detect rare associations that may elude human review [29].

Critically, AI models can continuously learn from new patient data through adaptive feedback loops, refining diagnostic hypotheses and improving performance over time. Unlike traditional statistical methods that require clean, complete datasets and are often very sensitive in the face of incomplete/incoherent data, AI systems can operate effectively in these imperfect conditions. This characteristic is particularly valuable in rare disease contexts, where large, homogeneous, and standardized cohorts are rarely available. In this way, each individual case, no matter how incomplete, can contribute to a cumulative expansion of clinical knowledge [24].

Despite this promise, transforming RWD into structured, operable, and analytically usable databases presents significant challenges. Data interoperability remains a major barrier, with differing standards and architectures across healthcare systems limiting integration. Privacy concerns, especially for pediatric populations, add an essential ethical dimension requiring robust governance frameworks. Moreover, trust in AI-generated outputs must be carefully built. If models are trained on non-representative datasets, there is a risk of bias propagation, leading to unequal diagnostic accuracy across demographic groups [34].

To fully harness RWD's potential in rare disease diagnostics, a multilayered infrastructure is needed. This includes standardized data collection protocols, harmonized phenotype-genotype registries, validation studies, and transparent AI pipelines that integrate seamlessly into clinical workflows. Within this ecosystem, AI should not be viewed merely as an analytical tool but rather as an active partner across the data lifecycle—from acquisition and curation to analysis, interpretation, and clinical application [35].

By combining the scalability and pattern-recognition capabilities of AI with the authenticity and richness of real-world data, a new diagnostic paradigm can emerge in pediatric genetics. One that does not begin in idealized study settings, but is rooted in the day-to-day complexity of medicine as practiced. In this paradigm, AI becomes the engine that transforms real-world evidence into concrete diagnostic value—facilitating faster identification of rare diseases, enabling earlier intervention, and delivering more personalized and equitable care to children worldwide [33].

3.4. Comparative Diagnostic Performance of AI and Human Experts

Despite the accelerating momentum of AI technologies in pediatric rare disease diagnostics, several barriers hinder their seamless adoption into clinical practice. These challenges extend beyond technical limitations and involve issues related to data quality, clinical workflows, and regulatory uncertainty [24].

To provide a more structured overview, Table 4 categorizes these obstacles based on their origin, highlights how they affect diagnostic effectiveness, and suggests potential mitigation strategies. Understanding the scope and nature of these challenges is essential for responsible implementation and long-term sustainability of AI tools in pediatric settings.

Table 4. Challenges and Their Diagnostic Impact.

Challenge	Category	Impact on Diagnosis	Addressable by:
Unstructured EMR data	Data issue	Limits phenotypic precision; weakens AI inputs	NLP tools; structured phenotyping
Lack of interoperability	Data/workflow	Prevents integration with AI tools and databases	Cross-platform EMR integration

Clinician skepticism and unfamiliarity	Workflow/human factor	Delays adoption; mistrust of AI recommendations	Targeted training, demonstration studies
Hallucination risk in LLMs	Algorithmic/technical	Produces plausible but false diagnoses	Validation, hybrid expert oversight
Regulatory ambiguity	Legal/ethical	Unclear liability; discourages clinical use	Guidelines, legal frameworks
Bias in training data	Ethical/data quality	Overlooks underrepresented populations	Diverse datasets, fairness auditing

Note: This table outlines key challenges affecting the use of AI in diagnostics, categorized by their nature and corresponding impact on clinical utility. It also highlights potential strategies or tools that can help mitigate each issue.

Several recent studies have sought to address this by benchmarking AI-assisted diagnostics against expert-led evaluations using real-world clinical cases of rare diseases. Controlled investigations involving genetically confirmed disorders suggest that AI platforms, ranging from phenotype-driven algorithms such as GEM to LLMs like ChatGPT and DeepSeek Medical AI, exhibit moderate to high diagnostic accuracy under specific conditions. However, their performance is highly variable and context-dependent, influenced by factors such as disease type, phenotypic clarity, data structure, and the complexity of individual cases [20,36,37].

For instance, AI systems generally perform well in diagnosing common and syndromically well-defined conditions, such as achondroplasia, osteogenesis imperfecta, or Noonan syndrome. These disorders are characterized by distinctive, easily recognizable phenotypic features that align closely with curated training data and established genotype-phenotype associations embedded within AI knowledge bases. In such settings, AI tools can rapidly match phenotypic inputs to known diagnostic entities, suggesting accurate diagnoses with minimal human input and delivering results within seconds [38].

However, AI performance declines notably when faced with ultra-rare, genetically heterogeneous, or clinically ambiguous conditions. In these scenarios, real life clinical judgment becomes essential. Diagnostic accuracy depends not only on recognizing textbook features, but also on integrating subtle, context-specific clues and elements that may not be explicitly encoded within AI training datasets. Additionally, ultra-rare diseases by definition lack large-scale representation in public databases, limiting AI's ability to draw upon prior examples for pattern recognition [39].

One comparative study evaluated AI models and human experts using a dataset of pediatric rare bone disease cases with confirmed molecular diagnoses. Experienced human clinicians achieved diagnostic accuracies exceeding 80%, reflecting the depth of their specialized training, clinical reasoning, and capacity to synthesize disparate data points into coherent diagnostic hypotheses. In contrast, AI models in the same study achieved accuracies in the range of 60–65%, depending on the complexity of cases, formulation of the input data, and the type of AI system employed [40].

Interestingly, combined approaches, where outputs from multiple AI tools were integrated, often resulted in improved diagnostic yield. This finding highlights the potential of multi-model strategies, analogous to multidisciplinary clinical discussions where diverse perspectives converge to refine a diagnosis. Such combined AI models can leverage complementary strengths of different systems, combining deep learning models' semantic reasoning with phenotype-driven tools' structured ontology matching to achieve higher accuracy [41].

Importantly, when AI-generated diagnoses were correct, their reasoning pathways demonstrated high concordance with human diagnostic logic. Models produced overlapping differential diagnoses, prioritized similar key phenotypic features, and aligned with expert interpretations. This convergence suggests that AI systems, when provided with sufficient structured data, can mirror aspects of human diagnostic reasoning, particularly for straightforward or classic presentations [42].

In comparing AI-based diagnostic outputs with those of human experts, it is important to consider not only accuracy but also confidence levels associated with each decision. Previous studies have shown that while AI models often achieve comparable or even higher diagnostic accuracy in certain tasks, their confidence levels can vary widely depending on the complexity of the case and the specificity of training data. Similarly, human experts tend to exhibit calibrated self-assessment, generally showing higher confidence in correct diagnoses and lower confidence when uncertain. In contrast, AI systems may demonstrate either overconfidence or underconfidence due to limitations in probabilistic calibration [43].

However, AI models remain susceptible to certain limitations. Overgeneralization is a recurrent challenge, with models occasionally proposing broad syndromic categories without sufficiently discriminating among closely related entities. Misinterpretation of delicate phenotypic cues, such as subtle dysmorphisms, behavioral features, or growth patterns, also limits diagnostic precision, especially when inputs are incomplete, unstructured, or described in non-standardized language. Additionally, AI lacks the ability to contextualize findings within psychosocial, cultural, or family-specific frameworks, aspects often critical to holistic clinical assessments [18].

Despite these limitations, AI models consistently deliver faster turnaround times compared to traditional diagnostic workflows. They can process vast datasets within seconds to minutes, perform unbiased variant assessments, and generate reproducible outputs without fatigue or cognitive bias. This scalability makes them invaluable in triaging large numbers of cases, flagging potentially actionable findings, and generating early diagnostic hypotheses that guide subsequent expert review and targeted testing strategies [44].

Beyond efficiency, AI systems contribute to democratize availability of the expertise. In settings where access to experienced clinical geneticists is limited, AI tools can serve as surrogate diagnostic aids, elevating the diagnostic capacity of the physicians. This holds particular promise in under-resourced regions, where the shortage of trained personnel remains a significant barrier to timely rare disease diagnosis [45].

3.5. Challenges in Clinical Implementation

Despite the growing promise of AI in the diagnosis of pediatric rare diseases, integrating these technologies into clinical workflows remains a complex challenge. Several interrelated obstacles must be addressed before AI tools can be fully embedded into everyday pediatric care. These can be grouped into three main categories: data-related issues, workflow and clinician adoption, and regulatory or ethical barriers (Table 5) [41].

Table 5. Key Challenges in Clinical Implementation of AI Tools.

Challenge	Description	Potential Solutions
Data Interoperability	Lack of standardized EMR ¹ and genomic data integration	Harmonized data standards
Workflow Integration	AI tools functioning as standalone systems	Seamless integration into hospital information systems
Clinician Training and Trust	Limited familiarity with AI methodologies	Targeted educational programs-demonstration projects
Validation and Regulation	Lack of universal validation standards	Development of regulatory frameworks specific to AI diagnostics
Resource Constraints	Infrastructure and cost barriers in low-resource settings	Cloud-based AI platforms, tiered implementation models

¹ Electronic Medical Record. **Note:** This table summarizes key barriers to the clinical implementation of AI tools, including technical, organizational, and regulatory challenges. It also presents potential solutions aimed at improving integration, usability, and trust in real-world healthcare settings.

AI platforms require access to high-quality, structured data, but many healthcare systems still rely on unstructured electronic medical records (EMRs). Phenotypic descriptions are often buried in free-text notes, inconsistently documented, or incomplete, which limits the functionality of AI tools dependent on standardized inputs such as HPO terms. Data fragmentation across institutions and the lack of interoperability between EMRs and genomic databases further hampers integrated analysis and AI-assisted interpretation. Automated natural language processing (NLP) could mitigate this issue, but its clinical implementation and validation are still limited.

One of the most pressing challenges is data interoperability. For AI platforms to function optimally, they require access to structured, high-quality clinical and genomic data. However, many healthcare systems continue to rely on electronic medical records (EMRs) that are predominantly unstructured, with phenotypic information buried within free-text clinical notes. Documentation practices vary significantly across clinicians, specialties, and institutions, resulting in inconsistent terminology and incomplete phenotypic capture. This heterogeneity severely limits the utility of phenotype-driven AI tools, which depend on standardized inputs such as HPO terms to generate accurate diagnostic suggestions [46].

Automated natural language processing systems capable of extracting and mapping free-text descriptions to structured ontologies remain underutilized in clinical genomics. Their integration into routine workflows would be transformative, yet technical implementation, validation, and clinician trust in automated data extraction remain ongoing effort [47].

Another major challenge is the lack of standardization across AI platforms. Different tools employ distinct algorithms, knowledge bases, variant classification frameworks, and decision-support logic. This variability leads to discrepancies in outputs, with potentially conflicting diagnostic suggestions for the same patient. In the absence of universally accepted guidelines or consensus statements outlining which AI tools to use, clinicians may hesitate to rely on them in high-stakes diagnostic decisions [41].

The integration of AI tools into existing clinical workflows poses additional difficulties. Many current AI solutions function as standalone platforms, requiring separate logins, data uploads, and output retrieval processes. This digital solution disrupts diagnostic pipelines and adds operational burdens to already stretched clinical teams. True integration demands seamless interoperability between AI systems, sequencing laboratories, EMRs, and hospital information systems, enabling streamlined data flow without duplication of effort [39].

Moreover, the need for human oversight introduces a paradoxical complexity. While AI systems can accelerate variant interpretation and phenotype matching, their outputs still require clinical validation and contextual judgment. Ultimately, responsibility for diagnostic decisions remains with physicians, who must critically assess AI-generated recommendations before acting upon them. Without thoughtful implementation, this dual-layer model can inadvertently increase clinician workload rather than alleviate it, creating additional cognitive and operational burdens [48].

Training and trust represent another significant challenge. Many clinicians remain unfamiliar with AI methodologies, underlying model architectures, and potential limitations. The perception of AI as a “black box” system, with unclear reasoning pathways, fuels skepticism and reluctance to fully integrate its outputs into clinical decision-making. Targeted education on AI fundamentals, strengths, and weaknesses, as well as exposure to its practical use, is essential for the meaningful and informed adoption and avoidance of superficial or misapplied utilization [49].

Finally, regulatory frameworks have yet to keep pace with the speed of AI innovation. Questions surrounding accountability, liability, data governance, and validation standards for AI-driven medical recommendations remain largely unresolved. This is particularly critical in pediatrics, where the legal and ethical dimensions of diagnostic decisions are much more challenging given the patient vulnerability, parental consent dynamics, and the potential long-term impact of diagnostic labeling on a child’s life trajectory. Ensuring that AI models are trained on diverse and representative pediatric datasets is essential to avoid propagating biases that could exacerbate existing health disparities [45].

Additionally, concerns around data privacy and security are amplified in the context of genomic and phenotypic data, which are easily identifiable. Robust safeguards are necessary to maintain public trust and comply with evolving legal frameworks, such as national data protection regulations [50].

Addressing these complex challenges will require a coordinated, multidisciplinary effort. Developers, clinicians, bioinformaticians, hospital administrators, ethicists, and policymakers must collaboratively design implementation strategies. These strategies must prioritize technical robustness, clinical relevance, ethical integrity, and operational feasibility. Transparent model validation, explainability of AI outputs, user-friendly interfaces embedded within clinical workflows, and ongoing post-implementation evaluation will be essential components of success [24].

Ultimately, AI should not be viewed as a standalone technological solution, but as a tool that, when thoughtfully integrated, complements and augments human expertise. Its true potential will be realized only when it seamlessly fits into the complex ecosystem of pediatric rare disease care. This will empower clinicians to deliver faster, more accurate, and more reliable diagnoses for children and families navigating their challenging diagnostic journeys [44].

3.6. Ethical Considerations in Pediatric Settings

The integration of AI into pediatric rare disease diagnostics introduces a constellation of unique ethical and interpretational challenges, extending far beyond questions of technical performance and diagnostic accuracy. These challenges are deeply rooted in the inherent vulnerability of pediatric patients, the psychological and emotional complexities of parental decision-making, and the long-term implications that early-life genetic diagnoses may hold for children as they grow into adulthood (Table 6) [41].

Table 6. Ethical Considerations in AI-Assisted Pediatric Rare Disease Diagnostics.

Ethical Domain	Key Issues	Proposed Mitigations
Transparency and Explainability	“Black box” decision-making processes	Develop interpretable AI models- provide output rationales
Informed Consent	Complexity of explaining AI involvement to parents	Tailored consent forms detailing AI role, benefits, and limitations
Equity and Bias	Underrepresentation of certain ethnic groups in training datasets	Diversify training data- continuous model revalidation
Privacy and Data Security	Handling identifiable genomic and phenotypic data	Robust encryption- compliance with pediatric data protection laws
Psychosocial Impact	Emotional burden of AI-generated diagnoses	Ensure clinician-led communication with empathy and support

Note: This table highlights core ethical concerns in applying AI to pediatric rare disease diagnostics, focusing on transparency, consent, bias, privacy, and psychosocial impact. Suggested mitigations emphasize the need for human-centered, secure, and equitable implementation.

One central concern is the “black-box” nature of many AI systems. While some platforms provide traceable reasoning pathways, confidence scores, or ranked output explanations that allow clinicians to understand how a conclusion was reached, many advanced deep learning models operate with undisclosed algorithms. The internal logic of their decision-making processes is often inaccessible even to developers. This lack of transparency raises fundamental questions of accountability. When an AI system suggests a diagnosis that carries life-altering consequences, whether in terms of treatment decisions, reproductive planning, or psychosocial impact, clinicians and families must face with recommendations that may be difficult to interpret or contest [49].

This issue is further magnified in pediatric settings, where decisions are made not for oneself but on behalf of a child. Here, diagnostic labels extend beyond immediate medical management- they

shape family identity, influence educational and social opportunities, and carry potential for future discrimination. The introduction of AI into this delicate landscape adds another layer of interpretational difficulties. Clinicians may feel ethically conflicted about relying on algorithmic outputs they cannot fully explain, while parents may struggle to trust diagnoses perceived as generated by an impersonal machine rather than guided by human expertise and empathy [51].

The interpretation of uncertain or incidental findings becomes especially ethically charged in the pediatric context. AI-driven variant prioritization tools often highlight genes of uncertain clinical significance (VUS) or reveal associations with adult-onset conditions unrelated to the current diagnostic question. Deciding what to report, how to communicate these findings to families, and whether any clinical action is warranted demands a careful balance between thoroughness and the potential for harm through overdiagnosis or undue anxiety [11].

Moreover, informed consent processes must evolve to encompass these complexities. Parents consenting to genomic testing for their child should be made aware not only of the nature and scope of the sequencing itself, but also of the role AI algorithms play in interpreting these vast datasets. Explaining the limitations, uncertainties, and possible unintended findings of AI-assisted analysis in an accessible, non-technical manner is essential to maintaining transparency, and trust [51].

Another critical consideration is equity and bias within AI systems. Most AI algorithms are trained on datasets derived from populations with disproportionate representation of certain ethnicities, geographic regions, and socioeconomic groups. As a result, diagnostic accuracy may be lower in underrepresented populations, including ethnic minorities, children from low-income regions, or those with atypical phenotypic features. In the field of rare diseases, where each case is precious and delays in diagnosis can carry lifelong consequences, such biases risk exacerbating existing disparities in healthcare access, quality, and outcomes [52].

Furthermore, the longitudinal implications of early AI-assisted diagnoses demand careful ethical reflection. Labeling a child with a rare genetic disorder based on an AI-informed diagnosis can influence their entire medical trajectory, shape self-perception, and affect insurability or future reproductive decisions. As AI becomes increasingly integrated into newborn screening, early developmental assessments, and routine diagnostic decision-making, these implications extend from the individual child to familial and societal levels [51].

The potential psychological impact on families should not be underestimated. Receiving a diagnosis through AI-assisted processes may feel depersonalized or disempowering if not communicated within a framework of compassionate clinical care. It is critical that AI never becomes a substitute for the human connection, empathy, and meaningful conversation that families navigating rare disease diagnoses urgently need [1].

Finally, there is a growing recognition that ethical frameworks must keep pace with technological innovation. Developing guidelines and safeguards to ensure responsible, equitable, and child-centered use of AI will require broad collaboration. Clinicians, bioethicists, AI developers, policymakers, and data scientists must all work together to create standards for transparency and accountability. Equally essential is the engagement of patient advocacy groups, parents, and young people themselves in shaping these frameworks. Their experiences and perspectives provide invaluable insights into how AI can best serve patients, families, and society [51,53].

Ultimately, grounding AI innovation in ethical reflection is not a peripheral consideration- it is central to realizing its promise. Only by addressing these ethical, interpretational, and societal challenges with foresight, humility, and inclusivity can AI achieve its transformative potential in the personalized pediatric medicine [54].

4. Conclusion and Future Direction

Artificial intelligence holds immense promise for transforming the diagnostic landscape of pediatric rare diseases. By augmenting the clinician's capacity to interpret complex genomic and phenotypic data, AI tools—ranging from variant prioritization engines to large language models—can reduce diagnostic delays, identify atypical presentations, and support clinical decision-making

in resource-limited settings. In a field marked by heterogeneity, time sensitivity, and limited specialist availability, these systems have the potential to act as diagnostic accelerators and equity multipliers [44].

Throughout this review, we have highlighted the expanding role of AI in interpreting complex genomic data, aligning diverse and sometimes subtle phenotypic presentations with known genetic conditions, and supporting earlier, more personalized clinical decision-making (Table 7).

Table 7. Key Advantages and Limitations of AI in Pediatric Rare Disease Diagnostics.

Aspect	Advantages of AI Approach	Limitations of AI Approach
Speed	Rapid analysis of large-scale genomic and phenotypic datasets	Limited validation for ultra-rare and atypical cases
Accuracy	High precision for syndromically well-defined conditions (e.g. achondroplasia)	Reduced accuracy in genetically heterogeneous or phenotypically ambiguous disorders
Accessibility	Expands diagnostic capacity in settings lacking subspecialty expertise	Dependent on data quality and input standardization
Result Interpretability	Transparent algorithms in some platforms allow reasoning review	“Black box” models hinder interpretability and trust
Cost-effectiveness	Long-term reduction in diagnostic odyssey costs	Initial investment required for infrastructure and training
Ethical Considerations	Enables faster diagnosis and personalized therapies	Risks of bias propagation and unequal diagnostic accuracy across populations

Note: This table contrasts key aspects of AI-assisted diagnostics in rare diseases, outlining both the benefits and limitations across domains such as speed, accuracy, accessibility, interpretability, cost-effectiveness, and ethics.

However, this potential must be tempered by a clear understanding of the limitations and risks involved. Many AI tools, particularly those based on deep learning or natural language generation, remain experimental and lack rigorous clinical validation. Phenomenon such as model hallucinations, data bias, and non-transparency in algorithmic decision-making underscore the need for human oversight and robust benchmarking. Moreover, the ethical implications of AI—ranging from algorithmic accountability to health equity and informed consent—require careful navigation [41,50,51].

True integration of AI into rare disease diagnostics will depend not only on technological maturity, but also on regulatory clarity, institutional readiness, and interdisciplinary collaboration. Clinicians must be actively involved in shaping AI tools to ensure they align with the realities of practice. Educational initiatives targeting AI literacy, transparent performance metrics, and dynamic feedback loops will be essential in establishing trust and long-term utility [51].

Importantly, AI should not be viewed as a substitute for expert reasoning, but rather as a dynamic partner—amplifying human insight where data volume, complexity, or fragmentation would otherwise impede diagnosis. As the field evolves, we must remain vigilant that AI solutions do not outpace the clinical and ethical frameworks needed to govern their use [44].

In the end, success will be measured not by technological novelty alone, but by whether AI contributes to timely, equitable, and meaningful diagnoses for children and families navigating the diagnostic odyssey. We must not only ask what AI can do, but what it should do—and under whose guidance [44,45,51].

Author Contributions: N.I. and A.S.: conceptualization, methodology, literature search, writing the original manuscript draft, and illustration preparation; editing, reviewing, and finalizing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are contained within the article.

Acknowledgments: The authors acknowledge the use of artificial intelligence tools to support language refinement and preliminary structuring of this manuscript. Final content, interpretation, and critical revisions were conducted solely by the authors. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
LLM /LLMs	Large Language Model(s)
NGS	Next-Generation Sequencing
WES	Whole-Exome Sequencing
WGS	Whole-Genome Sequencing
HPO	Human Phenotype Ontology
RF	Reverse Phenotyping
VUS	Variant of Uncertain Significance
NLP	Natural Language Processing
EMR / EMRs	Electronic Medical Record(s)
EHR / EHRs	Electronic Health Record(s)
RWD	Real-World Data
RCT / RCTs	Randomized Controlled Trial(s)
OMIM	Online Mendelian Inheritance in Man

References

1. Bauskis A, Strange C, Molster C, Fisher C. The diagnostic odyssey: insights from parents of children living with an undiagnosed condition. *Orphanet J Rare Dis.* 2022 Jun 18;17(1):233.
2. The Voice of 12,000 Patients [Internet]. EURORDIS-Rare Diseases Europe. [cited 2025 Jul 13]. Available from: <https://www.eurordis.org/publications/the-voice-of-12000-patients/>
3. Aldharman SS, Al-Jabr KH, Alharbi YS, Alnajjar NK, Alkhanani JJ, Alghamdi A, et al. Implications of Early Diagnosis and Intervention in the Management of Neurodevelopmental Delay (NDD) in Children: A Systematic Review and Meta-Analysis. *Cureus.* 2023 May;15(5):e38745.
4. Bowling KM, Thompson ML, Amaral MD, Finnla CR, Hiatt SM, Engel KL, et al. Genomic diagnosis for children with intellectual disability and/or developmental delay. *Genome Medicine.* 2017 May 30;9(1):43.
5. Ao G, Chen M, Li J, Nie H, Zhang L, Chen Z. Comparative analysis of large language models on rare disease identification. *Orphanet J Rare Dis.* 2025 Apr 1;20:150.
6. Sánchez Fernández I, Loddenkemper T, Gaínza-Lein M, Sheidley BR, Poduri A. Diagnostic yield of genetic tests in epilepsy: A meta-analysis and cost-effectiveness study. *Neurology.* 2019 Jan 28;92(5):e418–28.
7. Sun Y, Peng J, Liang D, Ye X, Xu N, Chen L, et al. Genome sequencing demonstrates high diagnostic yield in children with undiagnosed global developmental delay/intellectual disability: A prospective study. *Hum Mutat.* 2022 May;43(5):568–81.
8. Best S, Yu J, Lord J, Roche M, Watson CM, Bevers RPJ, et al. Uncovering the burden of hidden ciliopathies in the 100 000 Genomes Project: a reverse phenotyping approach. *Journal of Medical Genetics.* 2022 Dec 1;59(12):1151–64.

9. Burdick KJ, Cogan JD, Rives LC, Robertson AK, Koziura ME, Brokamp E, et al. Limitations of exome sequencing in detecting rare and undiagnosed diseases. *Am J Med Genet A*. 2020 Jun;182(6):1400–6.
10. Abbasi A, Alexandrov LB. Significance and limitations of the use of next-generation sequencing technologies for detecting mutational signatures. *DNA Repair (Amst)*. 2021 Nov;107:103200.
11. Austin-Tse CA, Jobanputra V, Perry DL, Bick D, Taft RJ, Venner E, et al. Best practices for the interpretation and reporting of clinical whole genome sequencing. *NPJ Genom Med*. 2022 Apr 8;7:27.
12. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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*. 2015 May;17(5):405–24.
13. Alvarez-Costes S. Deciphering Genomic Complexity: The Role of Explainable AI in Evolutionary Genomics. *Methods Mol Biol*. 2025;2927:221–34.
14. Aster A, Laupichler MC, Rockwell-Kollmann T, Masala G, Bala E, Raupach T. ChatGPT and Other Large Language Models in Medical Education — Scoping Literature Review. *Med Sci Educ*. 2024 Nov 13;35(1):555–67.
15. Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of Hallucination in Natural Language Generation. *ACM Comput Surv*. 2023 Mar 3;55(12):248:1-248:38.
16. Sharma J, Goel P. The Use of AI for Phenotype-Genotype Mapping. *Methods Mol Biol*. 2025;2952:369–410.
17. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019 Oct 25;366(6464):447–53.
18. Deng J, Zubair A, Park YJ. Limitations of large language models in medical applications. *Postgrad Med J*. 2023 Nov 20;99(1178):1298–9.
19. Carbonari V, Veltri P, Guzzi PH. Decoding Rarity: Large Language Models in the Diagnosis of Rare Diseases [Internet]. arXiv; 2025 [cited 2025 Jun 7]. Available from: <http://arxiv.org/abs/2505.17065>
20. Iqbal U, Tanweer A, Rahmanti AR, Greenfield D, Lee LTJ, Li YCJ. Impact of large language model (ChatGPT) in healthcare: an umbrella review and evidence synthesis. *J Biomed Sci*. 2025 May 7;32:45.
21. Wilczewski CM, Obasohan J, Paschall JE, Zhang S, Singh S, Maxwell GL, et al. Genotype first: Clinical genomics research through a reverse phenotyping approach. *Am J Hum Genet*. 2023 Jan 5;110(1):3–12.
22. Smail C, Ge B, Keever-Keigher MR, Schwendinger-Schreck C, Cheung WA, Johnston JJ, et al. Complex trait associations in rare diseases and impacts on Mendelian variant interpretation. *Nat Commun*. 2024 Sep 18;15(1):8196.
23. Gargano MA, Matentzoglou N, Coleman B, Addo-Lartey EB, Anagnostopoulos AV, Anderton J, et al. The Human Phenotype Ontology in 2024: phenotypes around the world. *Nucleic Acids Res*. 2024 Jan 5;52(D1):D1333–46.
24. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J*. 2021 Jul;8(2):e188–94.
25. Garcelon N, Neuraz A, Salomon R, Bahi-Buisson N, Amiel J, Picard C, et al. Next generation phenotyping using narrative reports in a rare disease clinical data warehouse. *Orphanet Journal of Rare Diseases*. 2018 May 31;13(1):85.
26. Mishima H, Suzuki H, Doi M, Miyazaki M, Watanabe S, Matsumoto T, et al. Evaluation of Face2Gene using facial images of patients with congenital dysmorphic syndromes recruited in Japan. *J Hum Genet*. 2019 Aug;64(8):789–94.
27. Birgmeier J, Haeussler M, Deisseroth CA, Steinberg EH, Jagadeesh KA, Ratner AJ, et al. AMELIE speeds Mendelian diagnosis by matching patient phenotype and genotype to primary literature. *Sci Transl Med*. 2020 May 20;12(544):eaau9113.

28. Rao A, Joseph T, Saipradeep VG, Kotte S, Sivadasan N, Srinivasan R. PRIORI-T: A tool for rare disease gene prioritization using MEDLINE. *PLOS ONE*. 2020 Apr 21;15(4):e0231728.
29. Kafkas Ş, Abdelhakim M, Althagafi A, Toonsi S, Alghamdi M, Schofield PN, et al. The application of Large Language Models to the phenotype-based prioritization of causative genes in rare disease patients. *Sci Rep*. 2025 Apr 29;15(1):15093.
30. Tudur Smith C, Williamson PR, Beresford MW. Methodology of clinical trials for rare diseases. *Best Pract Res Clin Rheumatol*. 2014 Apr;28(2):247–62.
31. Liu J, Barrett JS, Leonardi ET, Lee L, Roychoudhury S, Chen Y, et al. Natural History and Real-World Data in Rare Diseases: Applications, Limitations, and Future Perspectives. *J Clin Pharmacol*. 2022 Dec;62(Suppl 2):S38–55.
32. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence — What Is It and What Can It Tell Us? *N Engl J Med*. 2016 Dec 8;375(23):2293–7.
33. Hampson G, Towse A, Dreitlein WB, Henshall C, Pearson SD. Real-world evidence for coverage decisions: opportunities and challenges. *Journal of Comparative Effectiveness Research*. 2018 Dec;7(12):1133–43.
34. Orsini LS, Berger M, Crown W, Daniel G, Eichler HG, Goetsch W, et al. Improving Transparency to Build Trust in Real-World Secondary Data Studies for Hypothesis Testing—Why, What, and How: Recommendations and a Road Map from the Real-World Evidence Transparency Initiative. *Value in Health*. 2020 Sep 1;23(9):1128–36.
35. Weiss A, Michels C, Burgmer P, Mussweiler T, Ockenfels A, Hofmann W. Trust in everyday life. *Journal of Personality and Social Psychology*. 2021;121(1):95–114.
36. hanjianwei. Everything About DeepSeek: Key Features, Usage, and Technical Advantages [Internet]. PopAi. 2025 [cited 2025 Jun 6]. Available from: <https://www.popai.pro/resources/everything-about-deepseek/>
37. Germain DP, Gruson D, Malcles M, Garcelon N. Applying artificial intelligence to rare diseases: a literature review highlighting lessons from Fabry disease. *Orphanet Journal of Rare Diseases*. 2025 Apr 17;20(1):186.
38. Wojtara M, Rana E, Rahman T, Khanna P, Singh H. Artificial intelligence in rare disease diagnosis and treatment. *Clin Transl Sci*. 2023 Aug 30;16(11):2106–11.
39. Wojtara M, Rana E, Rahman T, Khanna P, Singh H. Artificial intelligence in rare disease diagnosis and treatment. *Clinical and Translational Science*. 2023;16(11):2106–11.
40. Ilić N, Marić N, Cvetković D, Bogosavljević M, Bukara-Radujković G, Krstić J, et al. The Artificial Intelligence-Assisted Diagnosis of Skeletal Dysplasias in Pediatric Patients: A Comparative Benchmark Study of Large Language Models and a Clinical Expert Group. *Genes*. 2025 Jul;16(7):762.
41. Jandoubi B, Akhloufi MA. Multimodal Artificial Intelligence in Medical Diagnostics. *Information*. 2025 Jul;16(7):591.
42. Harada T, Shimizu T, Kaji Y, Suyama Y, Matsumoto T, Kosaka C, et al. A Perspective from a Case Conference on Comparing the Diagnostic Process: Human Diagnostic Thinking vs. Artificial Intelligence (AI) Decision Support Tools. *Int J Environ Res Public Health*. 2020 Aug 22;17(17):6110.
43. Ilić N, Marić N, Cvetković D, Bogosavljević M, Bukara-Radujković G, Krstić J, et al. The Artificial Intelligence-Assisted Diagnosis of Skeletal Dysplasias in Pediatric Patients: A Comparative Benchmark Study of Large Language Models and a Clinical Expert Group. *Genes*. 2025 Jul;16(7):762.
44. Brasil S, Pascoal C, Francisco R, Dos Reis Ferreira V, Videira PA, Valadão AG. Artificial Intelligence (AI) in Rare Diseases: Is the Future Brighter? *Genes (Basel)*. 2019 Nov 27;10(12):978.
45. Rubeis G, Dubbala K, Metzler I. “Democratizing” artificial intelligence in medicine and healthcare: Mapping the uses of an elusive term. *Front Genet*. 2022;13:902542.

46. Holmes JH, Beinlich J, Boland MR, Bowles KH, Chen Y, Cook TS, et al. Why Is the Electronic Health Record So Challenging for Research and Clinical Care? *Methods Inf Med*. 2021 May;60(1–02):32–48.
47. Kersloot MG, van Putten FJP, Abu-Hanna A, Cornet R, Arts DL. Natural language processing algorithms for mapping clinical text fragments onto ontology concepts: a systematic review and recommendations for future studies. *Journal of Biomedical Semantics*. 2020 Nov 16;11(1):14.
48. Holzinger A, Zatloukal K, Müller H. Is human oversight to AI systems still possible? *N Biotechnol*. 2025 Mar 25;85:59–62.
49. London AJ. Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *Hastings Cent Rep*. 2019 Jan;49(1):15–21.
50. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics*. 2021 Sep 15;22(1):122.
51. Rigby MJ. Ethical Dimensions of Using Artificial Intelligence in Health Care. *AMA Journal of Ethics*. 2019 Feb 1;21(2):121–4.
52. Saint James Aquino Y. Making decisions: Bias in artificial intelligence and data-driven diagnostic tools. *Aust J Gen Pract*. 2023 Jul;52(7):439–42.
53. Kolbinger FR, Veldhuizen GP, Zhu J, Truhn D, Kather JN. Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. *Commun Med (Lond)*. 2024 Apr 11;4:71.
54. Mennella C, Maniscalco U, De Pietro G, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon*. 2024 Feb 15;10(4):e26297.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.