

The Use of Artificial Intelligence in Human Genetics and Genomics Research

ARTIFICIAL INTELLIGENCE (AI) AND HUMAN GENETICS

New AI tools (e.g., large language models and deep learning systems) are transforming science and healthcare. The AI used in research and clinical settings differs greatly from the AI people encounter in daily life. These tools are built for highly specialized scientific purposes, and can analyze large amounts of data, identify patterns, and support scientific and healthcare decision-making. In human genetics and genomics research, AI can now help researchers and clinicians interpret complex biological data at a scale and speed that had previously been impossible. AI tools are designed to support, but not replace, human expertise, and have exciting potential to help accelerate discoveries that are able to transform and benefit public health.

THE ROLE OF AI IN GENOMICS



A single human genome contains approximately **3.2 billion base pairs** of DNA.

Advances in genetics and genomics have generated **vast amounts of data**.

- For example, the NIH *All of Us* Research Program has collected genetic data from more than 747,000 participants, enabling discovery across both rare and common diseases.
- AI is able to synthesize and process large datasets in microseconds compared to human analysis, which can help researchers analyze this data quickly, accelerating the pace at which new breakthroughs can be made.

AI is currently being **used in many ways** across clinical care and research, including to:

- Automate analysis of large datasets to decode the functions of specific genes and variants in genetic diseases.¹
- Identify genetic risk factors for rare and common diseases.²
- Support clinical diagnosis of rare diseases by rapidly analyzing multiple data formats such as imaging and electronic health records.³
- Advance drug discovery, pharmacogenomics, and gene therapies.⁴
- Enable rapid interpretation of whole-genome sequencing results for critically ill newborns.⁵
- Facilitate clinical trial matching by identifying patients whose genomic profiles align with study criteria.⁶

REFERENCES ashg.org/advocacy/fact-sheets/

LOOKING FORWARD

As researchers and clinicians continue to incorporate AI into their workflows, **AI will be able to help:**

- Speed diagnosis of rare and inherited diseases, reducing the years-long diagnostic delays many patients currently experience
- Improve disease prevention and treatment, including cancer risk screening
- Enhance precision medicine by accurately identifying new targets for gene therapy while minimizing off-target effects
- Strengthen public health through expanded screening programs and population-level insights
- Improve patient care through increased efficiency and throughput, resulting in decreased wait times to see specialist providers.

KEY POLICY CONSIDERATIONS

Privacy and Data Protection: genetic data is inherently sensitive and can uniquely identify individuals and their biological relatives. Health privacy laws do not cover many genomic and consumer data uses, leaving significant gaps. Strong safeguards are needed to prevent misuse, reidentification, and unauthorized secondary use of genetic information, along with remedy for harms if safeguards fail. However, it is also vital to ensure that regulations do not obstruct collaborative, open science.

Transparency and Oversight: many AI models operate as “black boxes,” making it difficult for clinicians and researchers to understand how conclusions are reached. Human oversight is essential to ensure trustworthy research and clinical decisions.

Training and Workforce Strength: while AI offers powerful capabilities for large-scale data analysis, it cannot replace the need for human expertise and oversight. Therefore, it is essential for institutions to sustain strong training pipelines, ensuring that the next generation of scientists can both effectively leverage AI tools and rigorously validate their outputs. Continued investment in early-career training is vital to maintaining a deep talent pool, supporting scientific excellence, and preserving the United States’ long-term global competitiveness.

Bias and Representation: systems trained on non-representative genetic datasets may perform poorly for populations who are underrepresented in these datasets, which could lead to biased results. Using high-quality, representative genetic data helps to avoid reinforcing existing health disparities and increase the likelihood of meaningful biomedical discoveries that will improve precision medicine for all.

Validation and Reliability: AI tools must be rigorously tested and validated before use in clinical care to ensure accuracy and safety, and should be continually monitored to maintain a high quality of care across various settings.

HOW CAN CONGRESS SUPPORT THIS RESEARCH?

✓ **Maintain robust and sustained federal investment** in biomedical research, including research on the ethical, legal, and social implications of emerging technologies.

✓ **Support building and sustaining expansive and representative genomic data sets** through international collaboration and programs like the *All of Us* Research Program to accelerate medical breakthroughs and our understanding of the human genome.

✓ **Work with scientific experts to promote responsible AI use** that keeps pace with the rapidly evolving technology and regulatory landscape, as well as to develop clear guidance for benchmarking AI outputs to ensure standardized and reliable output for implementation in the clinical space.

✓ **Establish clear, comprehensive privacy, governance, and regulatory frameworks** for the use of AI in genetics that address data protection, consent, transparency, and accountability across the data lifecycle. This includes how data is collected, used, shared, stored, and reused.