

Why GAO Did This Study

Developing and bringing a new drug to market is lengthy and expensive. Drug developers study the benefits and risks of new compounds before seeking Food and Drug Administration (FDA) approval. Only about one out of 10,000 chemical compounds initially tested for drug potential makes it through the research and development pipeline, and is then determined by FDA to be safe and effective and approved for marketing in the United States. Machine learning is enabling new insights in the field.

GAO was asked to conduct a technology assessment on the use of AI technologies in drug development with an emphasis on foresight and policy implications. This report discusses (1) current and emerging AI technologies available for drug development and their potential benefits; (2) challenges to the development and adoption of these technologies; and (3) policy options to address challenges to the use of machine learning in drug development.

GAO assessed AI technologies used in the first three steps of the drug development process—drug discovery, preclinical research, and clinical trials; interviewed a range of stakeholder groups including, government, industry, academia, and nongovernmental organizations; convened a meeting of experts in conjunction with the National Academies; and reviewed key reports and scientific literature. GAO is identifying policy options in this report.

View [GAO-20-215SP](#). For more information, contact Timothy M. Persons, PhD, at 202-512-6888 or personst@gao.gov.

ARTIFICIAL INTELLIGENCE IN HEALTH CARE

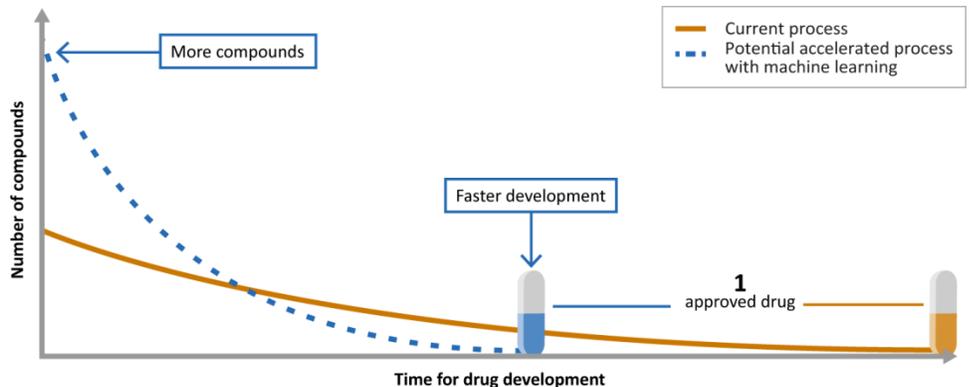
Benefits and Challenges of Machine Learning in Drug Development

What GAO Found

Machine learning—a field of artificial intelligence (AI) in which software learns from data to perform a task—is already used in drug development and holds the potential to transform the field, according to stakeholders such as agency officials, industry representatives, and academic researchers. Machine learning is used throughout the drug development process and could increase its efficiency and effectiveness, decreasing the time and cost required to bring new drugs to market. These improvements could save lives and reduce suffering by getting drugs to patients in need more quickly, and could allow researchers to invest more resources in areas such as rare or orphan diseases.

Machine learning could accelerate drug development

This set of technologies could screen more chemical compounds and zero in on promising drug candidates in less time than the current process.



Source: GAO. | GAO-20-215SP

Examples of machine learning in the early steps of drug development include:

- **Drug Discovery:** Researchers are identifying new drug targets, screening known compounds for new therapeutic applications, and designing new drug candidates, among other applications.
- **Preclinical Research:** Researchers are augmenting preclinical testing and predicting toxicity before testing potential drugs in humans.
- **Clinical Trials:** Researchers are beginning to improve clinical trial design, a point where many drug candidates fail. Their efforts include applying machine learning to patient selection, recruitment, and stratification.

GAO identified several challenges that hinder the adoption and impact of machine learning in drug development. Gaps in research in biology, chemistry, and machine learning limit the understanding of and impact in this area. A shortage of high-quality data, which are required for machine learning to be effective, is another challenge. It is also difficult to access and share these data because of costs, legal issues, and a lack of incentives for sharing. Furthermore, a low supply of skilled and interdisciplinary workers creates hiring and retention challenges for drug companies. Lastly, uncertainty about regulation of machine learning used in drug development may limit investment in this field.

GAO developed six policy options in response to these challenges. Five policy options are centered around research, data access, standardization, human capital, and regulatory certainty. The last is the status quo, whereby policymakers—federal agencies, state and local governments, academic and research institutions, and industry, among others—would not intervene with current efforts. See below for details of the policy options and relevant opportunities and considerations.

Policy Options to Address Challenges to the Use of Machine Learning in Drug Development

	Opportunities	Considerations
<p>Research (report page 60)</p> <p>Policymakers could promote basic research to generate more and better data and improve understanding of machine learning in drug development.</p>	<ul style="list-style-type: none"> • Could result in increased scientific and technological output by solving previously challenging problems. • Could result in the generation of additional high-quality, machine readable data. 	<ul style="list-style-type: none"> • Basic research is generally considered a long term investment and its potential benefits are uncertain. • Would likely require assessment of available resources and may require reallocation of resources from other priorities.
<p>Data Access (report page 61)</p> <p>Policymakers could create mechanisms or incentives for increased sharing of high-quality data held by public or private actors, while also ensuring protection of patient data.</p>	<ul style="list-style-type: none"> • Could shorten the length of the drug development process and reduce costs. • Could help companies identify unsuccessful drug candidates sooner, conserving resources. 	<ul style="list-style-type: none"> • Would likely require coordination between various stakeholders and incur setup and maintenance costs. • Improper data sharing or use could have legal consequences. • Cybersecurity risks could increase, and those threats would likely take additional time and resources to mitigate. • Organizations with proprietary data could be reluctant to participate.
<p>Standardization (report page 62)</p> <p>Policymakers could collaborate with relevant stakeholders to establish uniform standards for data and algorithms.</p>	<ul style="list-style-type: none"> • Could improve interoperability by more easily allowing researchers to combine different data sets. • Could help efforts to ensure algorithms remain explainable and transparent, as well as aid data scientists with benchmarking. 	<ul style="list-style-type: none"> • Could be time- and labor-intensive because standards development typically requires consensus from a multitude of public and private-sector stakeholders. This process can result in standards development taking anywhere from 18 months to a decade to complete and require multiple iterations.
<p>Human Capital (report page 63)</p> <p>Policymakers could create opportunities for more public and private sector workers to develop appropriate skills.</p>	<ul style="list-style-type: none"> • Could provide a larger pool of skilled workers for agencies, companies, and other research organizations, allowing them to better leverage advances in the use of machine learning in drug development. • Interdisciplinary teamwork could improve as workers with different backgrounds learn to better communicate with one another. 	<ul style="list-style-type: none"> • Data science-trained workers could exit the drug development field in search of higher-paying opportunities. • Would likely require an investment of time and resources. Companies and agencies will need to decide if the opportunities and challenges justify the investment or shifting of existing resources and how best to provide such training.
<p>Regulatory Certainty (report page 64)</p> <p>Policymakers could collaborate with relevant stakeholders to develop a clear and consistent message regarding regulation of machine learning in drug development.</p>	<ul style="list-style-type: none"> • Could help increase the level of public discourse surrounding the technology and allow regulators and the public to better understand its use. • Drug companies could better leverage the technology if they have increased certainty surrounding how, if at all, regulators will review or approve the machine learning algorithms used in drug development. 	<ul style="list-style-type: none"> • Would likely require coordination within and among agencies and other stakeholders, which can be challenging and require additional time and costs. • If new regulations are promulgated, compliance costs and review times could be increased.
<p>Status Quo (report page 65)</p> <p>Policymakers could maintain the status quo (i.e., allow current efforts to proceed without intervention).</p>	<ul style="list-style-type: none"> • Challenges may be resolved through current efforts. • Companies are already using machine learning and may not need action from policymakers to continue expanding its use. 	<ul style="list-style-type: none"> • The challenges described in this report may remain unresolved or be exacerbated.

Source: GAO.