Delivering responsible
AI in the healthcare and
# life sciences industries

IBM.

Table of contents

Delivering responsible AI in the healthcare and life sciences industries

Technologists who serve industries like healthcare and life sciences (HLCS) must balance three goals:

1. Build technologies like AI to drive innovation, efficiency and competitive advantage.
2. Create solutions that promote responsibly curated, safe health equity.
3. Mitigate risk associated with security vulnerabilities and regulatory and IP infringement.

This is a challenge because it requires that human values be reflected in AI systems. AI should not lie or discriminate, and it should be safe for everyone to use. In HCLS, the stakes couldn't be higher.

The first step toward meeting this challenge involves understanding an organization's current state.
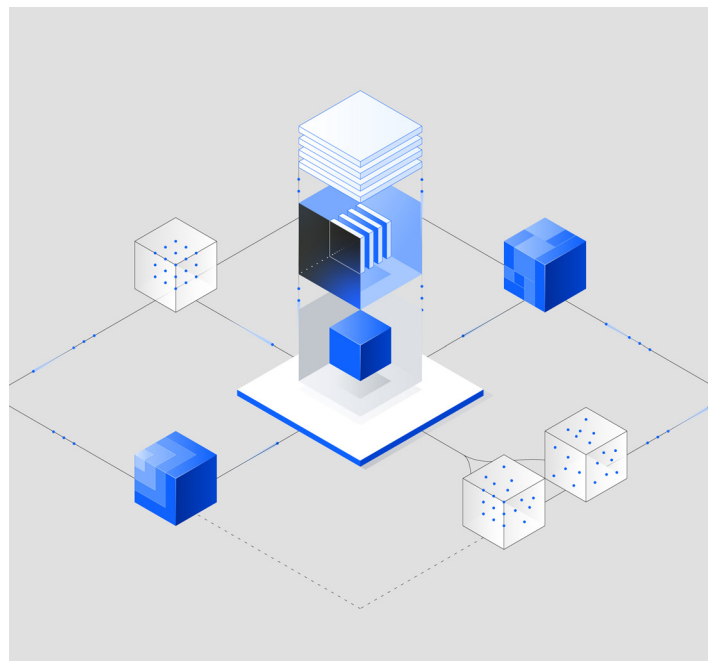
Assessing an organization's AI maturity

AI must be responsibly curated through a multidisciplinary and holistic approach. It begins with assessing organizations on a wide variety of enterprise capabilities that span multiple criteria, including:

– Vision and strategy
– AI operating model
– Data and technology
– AI engineering and operations
– Change management
– People management
– Regulatory understanding and readiness
– Cybersecurity

The first question often is, "Who is accountable for equitable and safe outcomes from AI?"

The three most common responses are quite concerning:

1. **"No one is responsible."**
   This is obviously concerning, because there can be no trust without accountability.
2. **"We don't use AI."**
   Employees are absolutely using AI, even if an organization doesn't realize it.
3. **"Everyone is responsible."**
   This is concerning because the work of AI governance requires power and a [funded mandate](). In other words, if everyone is responsible—including a fresh-out-of-school 22-year-old who is being pushed to meet a quota and a deadline—is anyone truly accountable?

AI governance

AI governance is composed of two key domains that often have separate stakeholders.

Organizational AI governance
Organizational AI governance explores the power structures within an organization as well as the culture, the investments in AI literacy and the overall understanding. Stakeholders necessarily have to be in the C-suite because individuals need to have enough power to align principles with an organization's brand. They need the power to shut down projects, stand up policy boards and roll out organization-wide AI literacy programs. They also need the power to have processes in place that measure the behaviors of personnel who will make models safer.

Organizations often have special advisory boards called AI ethics boards that explore the risks of specific use cases. Recently, the World Economic Forum (WEF) published a paper that stated, "An AI ethics council modelled on value-based principles is indispensable for any sized organization, no matter how small." The WEF also published a toolkit for fledgling AI ethics boards to consider, with guidance that covers a range of topics from whistleblower protections to rotating leadership models.

Automated AI model governance
Automated AI model governance uses tools to capture insights on how AI models behave. These models are critical to understanding whether AI models produce results that align to an organization's expectations and values. Watsonx.governance is an example of such a tool. Those who configure these tools will need training to know what to look for, as well as guidance from those doing the work of organizational AI governance to understand what is considered compliant.
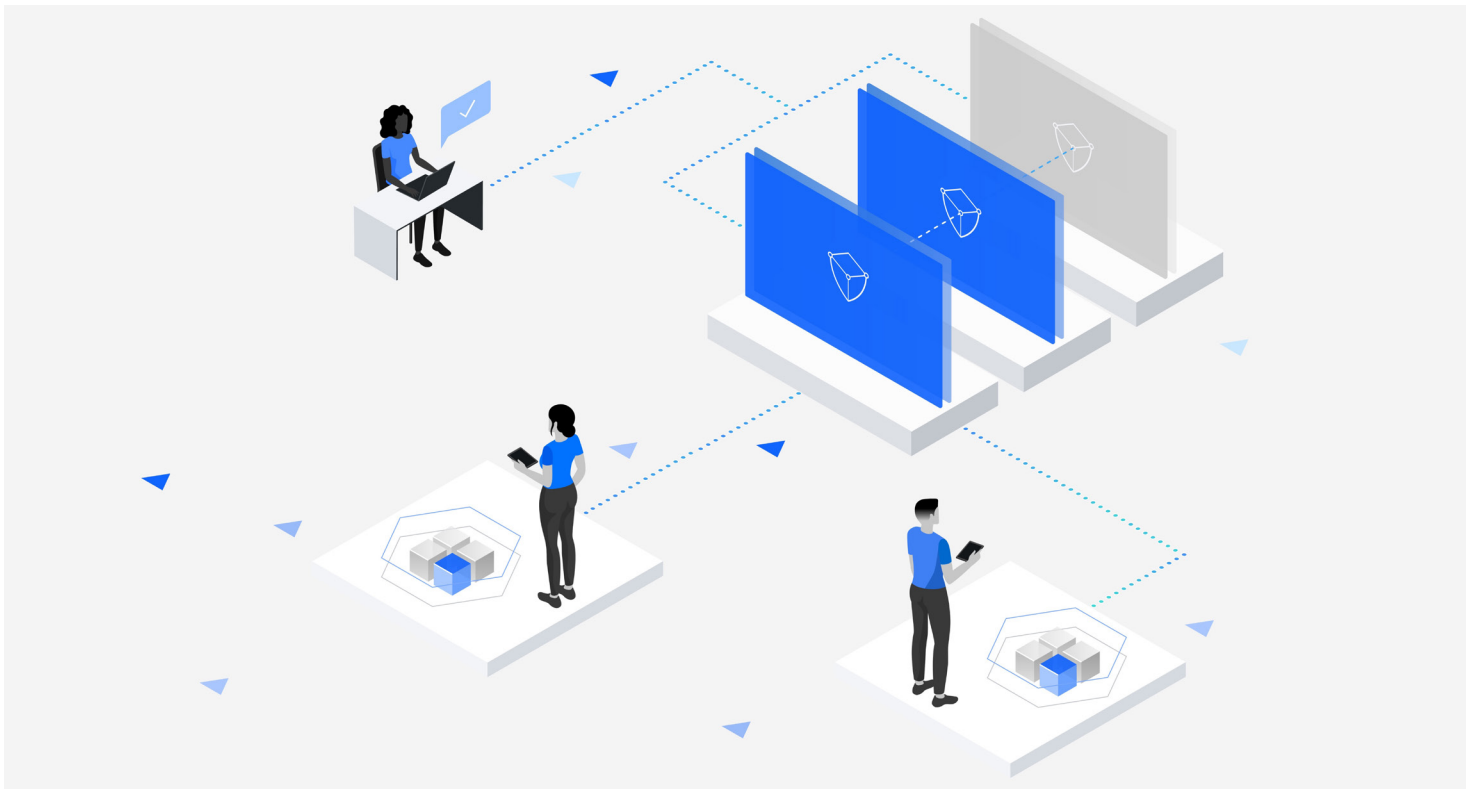
The challenge of health equity—and ensuring that the data used to train AI models represents the diverse communities that make up society

The COVID-19 pandemic revealed disturbing data about health inequity. In 2020, the National Institutes of Health (NIH) published a report stating that Black Americans died from COVID-19 at higher rates than White Americans, even though the former comprise a smaller percentage of the population. According to the NIH, these disparities were due to limited access to care, inadequacies in public policy, and a disproportionate burden of comorbidities, including cardiovascular disease, diabetes and lung diseases.

The NIH further stated that 47.5 million to 51.6 million Americans cannot afford the cost of a doctor visit. There is also a high likelihood that historically underserved communities might use a generative transformer—especially one that is embedded unknowingly into a search engine—to find medical advice. It is not inconceivable that individuals would go to a popular search engine with an embedded AI agent and query, "My dad can't afford the heart medication that was prescribed to him anymore. What is available over the counter that might work instead?"

According to researchers at Long Island University, ChatGPT is inaccurate 75% of the time, and according to CNN, the chatbot sometimes furnished dangerous advice, even approving the combination of two medications that could have serious adverse reactions. Given that generative transformers do not understand meaning and will have erroneous outputs, historically underserved communities that use this technology in place of professional help might be hurt at far greater rates than others.

Proactively invest in AI for more equitable and trustworthy outcomes

Today's new generative AI products make trust, security and regulatory issues top concerns for government healthcare officials and C-suite leaders who represent biopharmaceutical companies, health systems, medical device manufacturers and other organizations. Using generative AI requires AI governance, including conversations on appropriate use cases and guardrails around safety and trust (see the AI US Blueprint for an AI Bill of Rights, the EU AI ACT and the White House executive on AI).

Curating AI responsibly is a sociotechnical challenge that requires a holistic approach. Many elements are required to earn people's trust, including making sure that AI models are accurate, auditable, explainable, fair and protective of people's data privacy. Institutional innovation can help.

A historical note about institutional innovation
Institutional change is often preceded by a cataclysmic event. Consider the evolution of the US Food and Drug Administration (FDA), whose primary role is to make sure that food, drugs and cosmetics are safe for public use. While this regulatory body's roots can be traced back to 1848, monitoring drugs for safety was not a direct concern until 1937—the year of the elixir sulfanilamide disaster.

Created by a respected Tennessee pharmaceutical firm, elixir sulfanilamide was a liquid medication touted to dramatically cure strep throat. As was common for the time, the drug was not tested for toxicity before it went to market. This turned out to be a deadly mistake because the elixir contained diethylene glycol, a toxic chemical used in antifreeze. More than 100 people died from taking the poisonous elixir, which led to the FDA's Food, Drug and Cosmetic Act, which requires drugs to be labeled with adequate directions for safe use.

This milestone in FDA history made sure that physicians and their patients could fully trust in the strength, quality and safety of medications—an assurance we take for granted today. Similarly, institutional innovation is required to ensure equitable outcomes from AI.

Delivering responsible AI in the healthcare and life sciences industries | April 2024
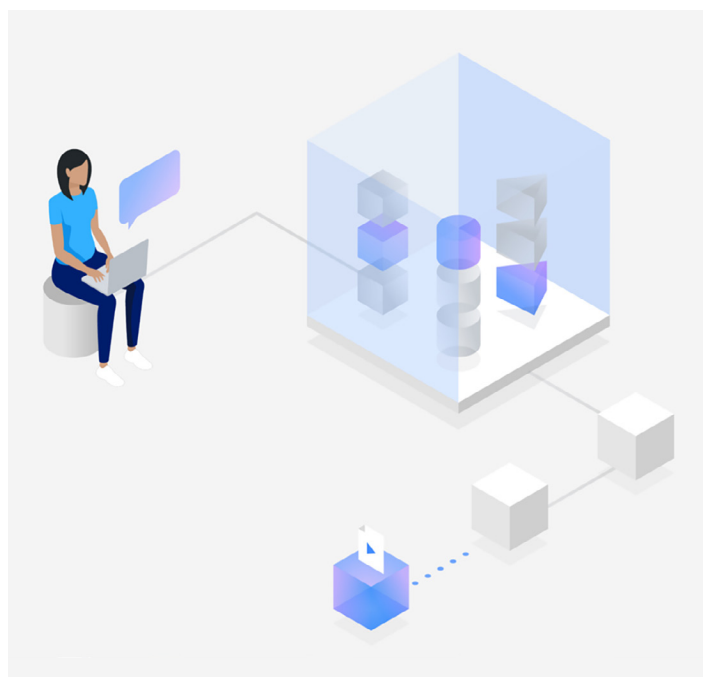
7 key steps to make sure generative AI supports the communities it serves

The use of generative AI in HCLS requires the same kind of institutional innovation that the FDA required during the elixir sulfanilamide disaster. The following recommendations can help make ensure that all AI solutions achieve more equitable and just outcomes for vulnerable populations.

1. **Get a holistic assessment of an organization's AI maturity.**
   Understand where any organization needs help to ensure that the right culture, AI governance processes, safeguard rails, and tools and AI engineering frameworks are in place so AI in the HCLS domain is responsibly built.

2. **Operationalize principles for trust and transparency.**
   AI models must be trained and audited to serve the most historically underserved populations. To earn the trust of the communities it serves, AI must have proven, repeatable, explained and trusted outputs that perform better than a human. AI must also be safe for people to use.

3. **Ensure safety from adversarial attacks.**
   Assessments about AI maturity are necessarily holistic because ensuring the safety of AI models—including safety from adversarial attacks—requires support from many different personas across an organization. Traditional cybersecurity operations will need to adapt by adopting new capabilities and developing new skill sets and expertise. Additionally, it is important to invite those with cybersecurity responsibilities to meetings about AI investments.

4. **Appoint individuals to be accountable for equitable outcomes from the use of AI.**
   Give these individuals the power and resources they need to perform the hard work that's required, including clinical, legal, regulatory, ethics compliance and security. Verify that these domain experts have fully funded mandates to do the work because without accountability, there is no trust. Someone must have the power, mindset and resources to do the work that's necessary for governance.

5. **Empower domain experts to curate and maintain the trusted sources of data that train models.**
   Trusted sources of data can offer content grounding for products that use large language models (LLMs) to provide variations on language for answers that come directly from a trusted source, like an ontology or semantic search.

6. **Mandate that outputs be auditable and explainable.**
   For example, some organizations are investing in generative AI that offers medical advice to patients or doctors. To encourage institutional change and protect all populations, these HCLS organizations should be subject to audits to ensure accountability and quality control. Outputs for these high-risk models should offer test-retest reliability. They should also be 100% accurate and detail data sources along with evidence.

7. **Require transparency.**
   As HCLS organizations integrate generative AI into patient care—for example, in the form of automated patient intake when checking into a US hospital or helping a patient understand what they can expect during a clinical trial—they should inform patients that a generative AI model is in use. Organizations should also offer interpretable metadata to patients that details the accountability and accuracy of the model, the source of the training data for the model and the audit results of the model. Metadata should also show how a user can opt out of using the model and get the same service elsewhere. As organizations use and reuse synthetically generated text in a healthcare environment, people should be informed about data that has been synthetically generated.

HCLS leaders can and must learn from the FDA's approach as they institutionally innovate and transform operations with AI. The journey to earning people's trust starts with making systemic changes that make sure AI better reflects the communities it serves.

Through collaborative partnerships, rigorous oversight and continuous ethical reflection, the HCLS industries can confidently navigate the complexities of traditional as well as gen AI deployment, paving the way for a more equitable, safe and patient-centered healthcare ecosystem.

For more information on leading practices or to connect with the IBM AI Ethics Board, contact your IBM representative or visit ibm.com/impact/ai-ethics.

About the authors

Bill Van Antwerp
Partner
IBM Consulting
Bill.VanAntwerp@ibm.com

Bill leads the payer practice at IBM Consulting and focuses on helping commercial and government health plans address business and technical challenges. His focus areas include commercial payer solutions, hybrid cloud transformation, Medicare Advantage growth and quality, provider/payer partnership optimization, and digital solutions. He provides his clients a full suite of payer solutions, including ERP, CRM, FHIR, HL7 support, and pharmacy benefits management (PBM) system transition, as well as digital, analytics, and business process improvement solutions.

Tamkeen Rasool
Partner
IBM Consulting
Tamkeen.Rasool@ibm.com

Tamkeen leads the healthcare platform services offering for IBM Consulting and manages a portfolio of large health plans as well as commercial and provider clients interested in data, advanced analytics, and healthcare interoperability. She is part of the IBM Consulting Healthcare Center of Excellence and has over 15 years of healthcare experience, developing many client relationships as a healthcare industry advisor, providing strategic support for long-term vision planning, managing programs, and identifying new business offerings.

About the authors

Brendan Fowkes
IBM Global Technology Leader, Healthcare
BJFowkes@us.ibm.com

Brendan's 25-plus years of healthcare experience across payers, providers and governments provides a unique viewpoint on the industry. His previous roles include a progression from electronic medical record development and implementation to revenue cycle business process optimization to analytics consulting. This allows him to easily bridge the gap between the executive suite and the technical solution design and deployment of any healthcare business challenge. As a member of the global healthcare team, Brendan brings innovative and proven use cases to his clients to help accelerate their time to value, as well as analytics-driven initiatives, including examples in conversational AI, real-time ICU monitoring and predictive modeling to prevent postsurgical complications.

Phaedra Boinodiris
IBM Consulting, Global Leader, Trustworthy AI
PBoinodi@us.ibm.com

A fellow with the London-based Royal Society of Arts, Phaedra has focused on inclusion in technology since 1999. She leads the IBM Consulting Trustworthy AI practice, is the AI Ethics Board focal for all consulting and leads the IBM Trustworthy Center of Excellence. She is the co-author of the book "AI for the Rest of Us." Phaedra is a co-founder of the Future World Alliance, a 501(c)(3) organization dedicated to curating K–12 education in AI ethics. In 2019, she won the United Nations Woman of Influence in STEM and Inclusivity Award and was recognized by Women in Games International as one of the Top 100 Women in the Games Industry as she began one of the first scholarship programs in the United States for women to pursue degrees in game design and development.

IBM