Roche Artificial Intelligence (AI) Ethics Principles

Principles to guide ethical AI use

Table of Contents

Roche Artificial Intelligence (AI) Ethics Principles

AI is being developed and utilised across all major industrial sectors, including healthcare, where there is vast potential to improve the health, wellbeing, and lives of patients, caregivers, providers, and society at large. AI presents significant opportunities in the areas of preventative medicine and treatment, as well as healthcare administration, diagnosis and delivery systems, and resource stewardship.

A broad range of techniques can be used to create AI solutions to gather insights and carry out healthcare tasks in clinical settings. Given the increasing complexity of care, research, and the speed of innovation, AI will be an adjunct tool for researchers, caregivers, and decision makers. It can equip researchers and healthcare professionals to more rapidly and effectively interpret complex data sets and allow more focus on meaningful interactions with patients. AI can also provide greater automation of manual tasks to aid efficiency and standardisation. With AI, society can potentially benefit from advancements in evidence-based treatment, higher-quality interactions with healthcare professionals, and more personalised care plans.

Roche recognizes this potential for AI in life science research and healthcare, and is focused on developing a range of AI-related solutions to be deployed in medical settings (such as AI-enabled diagnostics applications), for use in the development of medicinal products (i.e. to optimise and accelerate Research & Development), for use in different areas of scientific decision-making, and commercial applications that compliantly enable patient and customer experience. There is currently no harmonised definition of AI. The OECD defines an AI system as “as a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that [can] influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.” (more)

In early 2023, Roche adopted 13 public-facing Roche Data Ethics Principles to guide our data handling behaviour. Similarly, Roche has developed a clear set of principles for the design, development, deployment, and monitoring of AI solutions. Together, the Roche Data Ethics Principles and AI Ethics Principles align with our corporate values of integrity, passion, and courage, our corporate goals, as well as the principles of the WHO guidance on Ethics & Governance of Artificial Intelligence for Health and the International Federation of Pharmaceutical Manufacturers and Associations, of which Roche is a member.

1. Ethical Use

Maximising the benefit of AI and minimising harm

AI should be used to maximise the health and well-being of individuals by optimising and accelerating innovation, to improve our understanding of disease, optimise healthcare, and to contribute to the sustainable development of healthy societies.

Promoting ethical behaviour, Roche provides training and awareness programs to educate employees on AI ethics and best practices. To minimise harm, Roche is responsive to addressing concerns about safety and effectiveness, as well as the broader social impacts of employing AI. For example, Roche is helping to develop medical software using AI with the aim to rapidly and accurately identify people who are at increased risk of developing lung cancer. Additionally, Roche is using AI technology to investigate the chemical structure of potential antibiotics to determine which ones have the ability to bypass the pathogen's membrane.

2. Transparency

We are transparent about the use of AI to build trust and credibility with patients, healthcare providers, and society at large

Disclosures should be made about which functions and features of our products are AI-enabled and how humans should ultimately be involved in decision-making when AI is used. In creating AI systems, Roche should use data only from appropriate sources, as well as have the ability to preserve data provenance as required.

At Roche we take appropriate measures to promote accurate, accessible information about the assumptions and limitations of the technology, which data were used, and the properties of the data (e.g., data collection, processing). While AI systems and data may not be fully transparent for proprietary reasons, at Roche, appropriate consent and robust documentation are key components of traceability. The provenance of data, processes, and artefacts involved in the production of associated AI models are linked to the quality of our work.

3. Explainability

The ability of humans to understand and interpret AI solutions

Roche aims for processes and methods that allow users to understand, interpret, and trust our AI solutions, allowing them to exercise meaningful control.

AI is based on machine generated algorithms in a context that has the potential for direct and significant impact on individuals. Roche applies extra focus on transparency with concern to human agency and oversight, as well as elimination of bias, especially in situations where the complexity of the algorithm makes understanding of the algorithm difficult. Fostering trust, we are committed to developing and using AI appropriately, and monitoring our AI-enabled products and services for safety and effectiveness.

4. Human Control

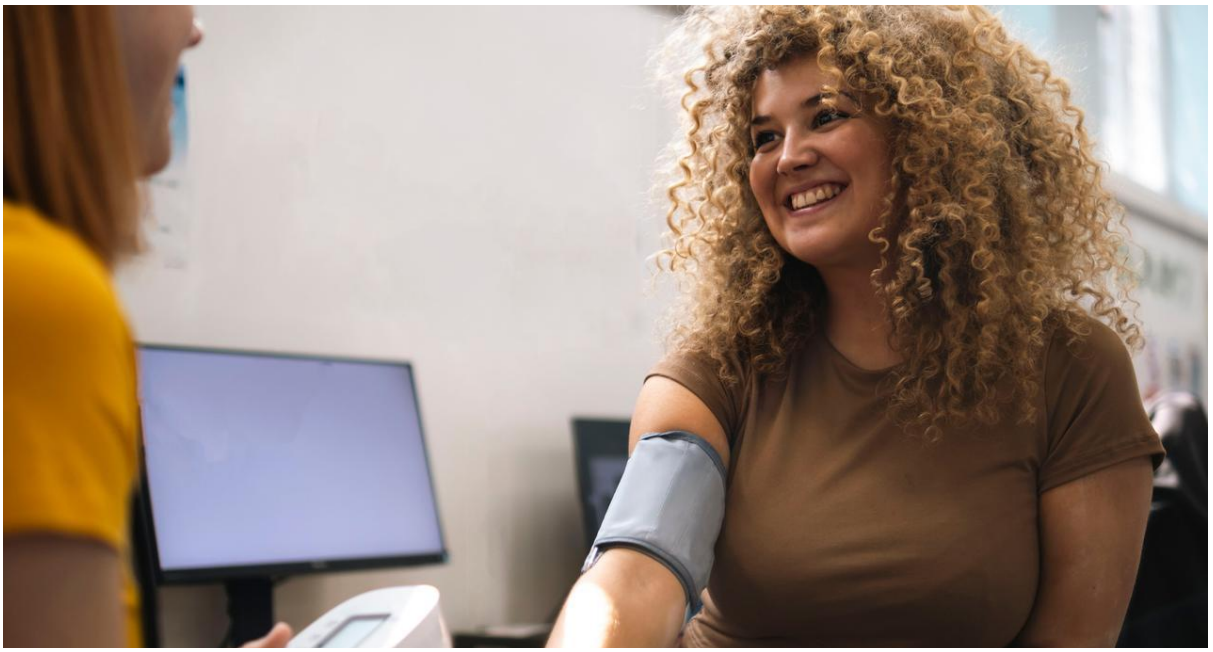
Humans can override recommendations made by AI systems and machine autonomy can be restricted

Because of the powerful impact that AI may have on individuals and society, AI shall not be given unrestricted autonomy. Medicine is a practice that enables our collective humanity, and retaining agency for patients, families, caregivers, and healthcare providers is paramount.

While allowing for necessary closed-loop systems, AI systems developed by Roche allow humans to maintain control of decisions. AI technology is leveraged to support humans to accomplish complex, analytic tasks, as well as for automation of tasks.

5. Empowering People

Enhancing the shared decision-making of patients, carers, and healthcare providers



Roche endeavours to develop and use AI solutions so as to empower those involved in healthcare to make better decisions through providing insights and interpretation of data. This aligns to the data ethics principle of autonomy and supports the best-practice of shared decision-making in AI design and development in healthcare.

For example, Genentech, a member of the Roche Group, is exploring the use of natural language processing to characterise speech and language patterns of patients with Alzheimer's Disease in a non-burdensome manner that could allow carers additional insight into their family members' condition.

6. Accountability

Monitoring, feedback, and governance actions are operationalized

There should be accountability for the responsible use of AI systems, including those developed by third parties, throughout the lifecycle of AI solutions developed or deployed by or on behalf of Roche.

Noting the fast-paced nature of AI technology, Roche incorporates appropriate governance and risk mitigation, deployment of impact-based controls, and incorporation of strategies to manage unintended consequences of AI systems, including contracting, adherence to regulatory requirements, auditing, as well as continual monitoring and feedback loops as AI evolves over time.

7. Fairness & Minimization of Bias

AI tools should be inclusive, equitable, and seek to support the mission of responding to the needs of all patients

Datasets and the assumptions used in the design of AI systems should avoid and minimise bias and maximise fairness through the proper selection of data and understanding of assumptions that are used in the design of AI-enabled systems.

Bias monitoring, detection, and prevention are essential for ethical AI systems. Roche will monitor and adapt AI systems to correct for bias and strive for inclusion throughout the AI lifecycle. Roche aims for transparency when known biases are present in AI solutions. Additionally, Roche pursues diversity among the designers and developers of AI. AI with Roche does community-driven science that aims to be patient-centric to promote diverse and inclusive datasets.

8. Privacy

Data for AI models are privacy-protected

With data at the heart of AI-enabled systems, Roche places privacy at the core of our development processes.

At Roche, in addition to ensuring lawfulness whenever personal data is used for the creation of AI systems, we protect privacy through the use of anonymization, pseudonymization, and encryption techniques, as well as synthetic data sets. This reduces the risk of personal identification when AI models are created. AI with Roche formed a collaborative partnership with an intent to resolve barriers for patients to share their de-identified data and engage in rare disease research. Roche's Data Ethics Principles and Code of Conduct also reflect commitment to data privacy.

9. Security

AI systems are designed to be secure

Strong cybersecurity standards are required, including processes to identify, assess, and mitigate risks of breaches in security, and notifying individuals when their data is breached.

AI can potentially create new avenues in manipulation and attack methods. Roche prioritises security during the design and development of AI systems and has teams of local and global data security experts to ensure that AI itself is secure and trustworthy.

10. Safety by Design

Safety is embedded in the design of AI systems

A proactive approach to AI system design should be used to fulfil user requirements as much as possible, while also ensuring they are verified and validated for effectiveness and safety, and thoroughly monitored to protect users from malfunctions or unintended consequences.

At Roche, our AI systems are designed with user experience considerations, as well as appropriate protection against threats to human health, human life, and the environment.

11. Sustainability

We think about the future impact of AI systems on our global society and planet

As stewards of both data and our planet, Roche is committed to reducing the negative impact of our AI systems on the environment.

Roche considers the long-term social, environmental, and economic impact of AI systems and ensures they are developed and operated in a sustainable manner. Roche's platforms curate large datasets for algorithm development, and this library of data fosters broad scientific collaboration across many facets of medical research. Aligning to Roche's sustainability goals, electricity from sustainable sources contributes to Roche's total electricity consumption. This electricity comes from specialist providers of energy from solar panels, windmills or hydropower plants, and it helps power computers and data servers (among other equipment). Referencing the Roche Data Ethics Principles, data is processed efficiently across the life cycle to ensure a small carbon footprint.

This position statement was created by Roche's Data & AI Ethics Workstream on 28 June 2023. It was reviewed by Roche Corporate Executive Committee and adopted on 20 December 2023.