



Roche AI Ethics Principles

Adarsh Srivastava

Head of Data, Analytics & Research Quality Assurance

Roche AI Ethics Principles paper



Why?

Given the ethical complexities of AI, there is a need to establish **foundational ethical principles that align to our organizational values** and govern our design, development, and Deployment of AI at Roche.



What?

Aim of this document is to **raise awareness of ethical behavior throughout the design, development, and deployment of AI**.



How?

Through **benchmarking and collaboration** a guiding set of 11 AI ethics principles were formulated.

Following creation of the principles document, the content was **reviewed internally by diverse group of experts and leaders** from Pharma, Diabetes Care, Diagnostics, FMI, etc. in various specialty fields.

Roche AI Ethics principles deep-dive



Roche has established and adopted **11 AI Ethics Principles** to guide responsible AI use in research and healthcare, maximising benefits, minimising harm, and promoting trust, fairness, transparency, and sustainability in AI applications.



Ethical Use: Maximizing the benefit of AI and minimizing harm.



Transparency: We are transparent about the use of AI to build trust and credibility with patients, and society at large.



Explainability: The ability of humans to understand and interpret AI solutions.



Human Control: Humans can override decisions made by AI systems and machine autonomy can be restricted.



Empowering People: Enhancing the shared decision-making of patients, carers, and healthcare providers.



Accountability: Monitoring, feedback, and governance actions are operationalized.

Roche AI Ethics principles deep-dive [Cont..]



Roche has established and adopted **11 AI Ethics Principles** to guide responsible AI use in research and healthcare, maximising benefits, minimising harm, and promoting trust, fairness, transparency, and sustainability in AI applications.



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



Privacy: Data for AI models are privacy-protected.



Security: AI systems are designed to be secure.



Safety by Design: Safety is embedded in the design of AI systems.



Fairness & Minimization of Bias: AI tools should be inclusive, equitable, and seek to support the mission of responding to the needs of all patients.

Fairness & Minimization of Bias from the Ethics Lens

- **Fairness** is an ethical principle that is linked to **equity** (NB note this is not *equality*)
- **Fairness** is also linked to the ethical principle of **justice**
- **Bias** conflicts with fairness, equity, and justice so we need to avoid it whenever possible
- We consider the **contextual features** of cases/situations in order to promote fairness



If we work with data with a biased mindset, and biased training sets (for AI), then the outputs will be biased and this conflicts with our Roche aim for patient centricity.



What do we do?



Promoting Fairness & Minimization of Bias in AI

Be reflective (ethics is not a hurdle, it's a strength).....often we aim for speed without also being reflective. Reflection is an ethics tool that gives time and space to analyze what we are doing and why we are doing it, as well as the consequences. It considers the bigger picture. **AI is not just technology – there are social variables.**



Ask yourself, is my work with data and AI building trust with my colleagues and the users?



Have you included diversity and inclusive data practices in your work? If not, you might have introduced bias.



Seek help from an ethicist – it's their expertise.



Doing now what patients need next