Qrious

AI Principles Template
Qrious

Standard Template for AI Principles

[insert your company]
Artificial Intelligence Principles

NOTE: This is a standardised set of AI Principles we have developed, aimed at getting you started. Please customise to your organisation to ensure its fit for purpose.

About this document

Artificial Intelligence (AI) technologies are rapidly evolving and being implemented at scale. These technologies have increasingly sophisticated capabilities, some of which can directly impact individuals or influence their behaviours, opinions, and choices.

These principles represent a public commitment and set out the requirements for all personnel when AI technologies are designed, deployed, and operated within our organisation.

Our AI Principles

We will adopt a responsible and ethical approach to the design and operation of AI technologies. These principles will be monitored and evolved in line with the development of new tools and processes related to AI technologies.

Human-centred

In using AI technologies, we will prioritise people – considering their needs and respecting human rights, autonomy, and diversity.

Ethical design

All employees involved in the creation and design of AI technologies, including those leveraging third-party software, must:

- Consider the end-to-end ethical impacts of the technology and its intended purpose.
- Adhere to our Code of Ethics, conducting themselves with integrity in an open, honest, and transparent manner.

Diversity, inclusivity, and bias

AI technologies rely on data sets to detect patterns and train machine learning algorithms to predict outcomes. Data sets and algorithms that contain biases can lead to unfair outcomes. We will implement processes to identify and minimise bias in the data sets, AI technologies, and algorithms we use. This includes human review of outputs for potential bias before use.

Safety and reliability

Deployers of AI technologies are accountable for ensuring that AI technology is fit for purpose and operates safely, responsibly, reliably, and effectively throughout its lifecycle.

Users of AI technologies are accountable for:

- Using each AI technology as intended by the deployers and within guidance aligned to these AI Principles.
- Monitoring whether the AI technology produces valid results and outcomes, adapting or stopping usage to ensure appropriate performance.
- Reporting any unanticipated outcomes or non-performance to the deployer.

Privacy

When designing AI and automated solutions, we will adhere to our Privacy Policy and Privacy Values, ensuring the protection of personal information.

Informed human decision making

For decisions involving sensitive or impactful topics, an appropriately skilled human must be accountable for informed and responsible decision making. The limits of the AI solution must be clearly understood and articulated to the user. The data used, the automated decision-making process, and the logic behind recommendations should be transparent to the individual responsible for the decision and action.

Explicability and transparency

Our AI technologies and algorithms must be explicable and transparent. We must be able to clearly articulate and explain how the solution works and how automated decisions are made. AI technologies must have a means of logging or capturing data sets, automated decisions, and actions for human review and understanding of the automated processes and outcomes.

We will inform customers in advance if they are interacting with an AI technology rather than a human, and enable customers to request contact with a human if needed. We will also inform customers if AI has been used to generate content sent to them without human review.

Adherence to these Principles

Training and support We will ensure that employees using AI technologies receive appropriate training and support to use these technologies competently, for their intended purposes, and in alignment with these AI Principles as well as our values, policies, and legal obligations.

Review and oversight We regularly review AI technologies, prediction, and decision-making processes to ensure they function as designed, including reviewing for bias, accuracy, and performance, and addressing any issues found. Employees who feel an AI technology is not delivering the intended outcome or operating in accordance with these Principles should raise their concerns with their manager, the appropriate accountable team, or through the anonymous whistleblowing process.

Qrious