Responsible AI dimensions, definitions, and examples

Source: AI Index, 2025 | Table: 2025 AI Index report

Responsible AI dimensions	Definition	Example
Privacy	An individual's right to confidentiality, anonymity, and security protections of their personal data, including the right to consent and be informed about data usage, coupled with an organization's responsibility to safeguard these rights when handling personal data.	Patient data is handled with strict confidentiality, ensuring anonymity and protection. Patients consent to whether their data can be used to train a tumor detection system.
Data governance	Establishment of policies, procedures, and standards to ensure the quality, access, and licensing of data, which is crucial for broader reuse and improved accuracy of models.	Policies and procedures are in place to maintain data quality and permissions for reuse of a public health dataset. There are clear data quality pipelines and specification of use licenses.
Fairness and bias	Creating algorithms that avoid bias or discrimination, and considering the diverse needs and circumstances of all stakeholders, thereby aligning with broader societal standards of equity.	A medical AI platform designed to avoid bias in treatment recommendations, ensuring that patients from all demographics receive equitable care.
Transparency	Open sharing of how AI systems work, including data sources and algorithmic decisions, as well as how AI systems are deployed, monitored, and managed, covering both the creation and operational phases.	The development choices, including data sources and algorithmic design decisions are openly shared. How the system is deployed and monitored is clear to health care providers and regulatory bodies.
Explainability	The capacity to comprehend and articulate the rationale behind the outputs of an AI system in ways that are understandable to its users and stakeholders.	The AI platform can articulate the rationale behind its treatment recommendations, making these insights understandable to doctors and patients to increase trust in the AI system.
Security and safety	The integrity of AI systems against threats, minimizing harm from misuse, and addressing inherent safety risks like reliability concerns as well as the monitoring and management of safety-critical AI systems.	Measures are implemented to protect against cyber threats and to ensure the system's reliability, minimizing risks from misuse and safeguarding patient health and data.