

Understand Responsible AI

10 minutes

At Microsoft, AI software development is guided by a set of six principles, designed to ensure that AI applications provide amazing solutions to difficult problems without any unintended negative consequences.

Fairness

AI systems should treat all people fairly. For example, suppose you create a machine learning model to support a loan approval application for a bank. The model should predict whether the loan should be approved or denied without bias. This bias could be based on gender, ethnicity, or other factors that result in an unfair advantage or disadvantage to specific groups of applicants.

Azure Machine Learning includes the capability to interpret models and quantify the extent to which each feature of the data influences the model's prediction. This capability helps data scientists and developers identify and mitigate bias in the model.

Another example is Microsoft's implementation of [Responsible AI with the Face service](#), which retires facial recognition capabilities that can be used to try to infer emotional states and identity attributes. These capabilities, if misused, can subject people to stereotyping, discrimination or unfair denial of services.

For more details about considerations for fairness, watch the following video.

<https://www.microsoft.com/en-us/videoplayer/embed/RE4vqfa?postJsllMsg=true>

Reliability and safety

AI systems should perform reliably and safely. For example, consider an AI-based software system for an autonomous vehicle; or a machine learning model that diagnoses patient symptoms and recommends prescriptions. Unreliability in these kinds of systems can result in substantial risk to human life.

AI-based software application development must be subjected to rigorous testing and deployment management processes to ensure that they work as expected before release.

For more information about considerations for reliability and safety, watch the following video.

<https://www.microsoft.com/en-us/videoplayer/embed/RE4vvll?postJsllMsg=true> 

Privacy and security

AI systems should be secure and respect privacy. The machine learning models on which AI systems are based rely on large volumes of data, which may contain personal details that must be kept private. Even after the models are trained and the system is in production, privacy and security need to be considered. As the system uses new data to make predictions or take action, both the data and decisions made from the data may be subject to privacy or security concerns.

For more details about considerations for privacy and security, watch the following video.

<https://www.microsoft.com/en-us/videoplayer/embed/RE4voJF?postJsllMsg=true> 

Inclusiveness

AI systems should empower everyone and engage people. AI should bring benefits to all parts of society, regardless of physical ability, gender, sexual orientation, ethnicity, or other factors.

For more details about considerations for inclusiveness, watch the following video.

<https://www.microsoft.com/en-us/videoplayer/embed/RE4vl9v?postJsllMsg=true> 

Transparency

AI systems should be understandable. Users should be made fully aware of the purpose of the system, how it works, and what limitations may be expected.

For more details about considerations for transparency, watch the following video.

<https://www.microsoft.com/en-us/videoplayer/embed/RE4vqfb?postJsllMsg=true> 

Accountability


People should be accountable for AI systems. Designers and developers of AI-based solutions should work within a framework of governance and organizational principles that ensure the solution meets ethical and legal standards that are clearly defined.


For more details about considerations for accountability, watch the following video.

<https://www.microsoft.com/en-us/videoplayer/embed/RE4vvlk?postJsllMsg=true> 

The principles of responsible AI can help you understand some of the challenges facing developers as they try to create ethical AI solutions.

Further resources

For more resources to help you put the responsible AI principles into practice, see <https://www.microsoft.com/ai/responsible-ai-resources> .

To see these policies in action you can read about [Microsoft's framework for building AI systems responsibly](#) .

Module complete:

Unlock achievement
