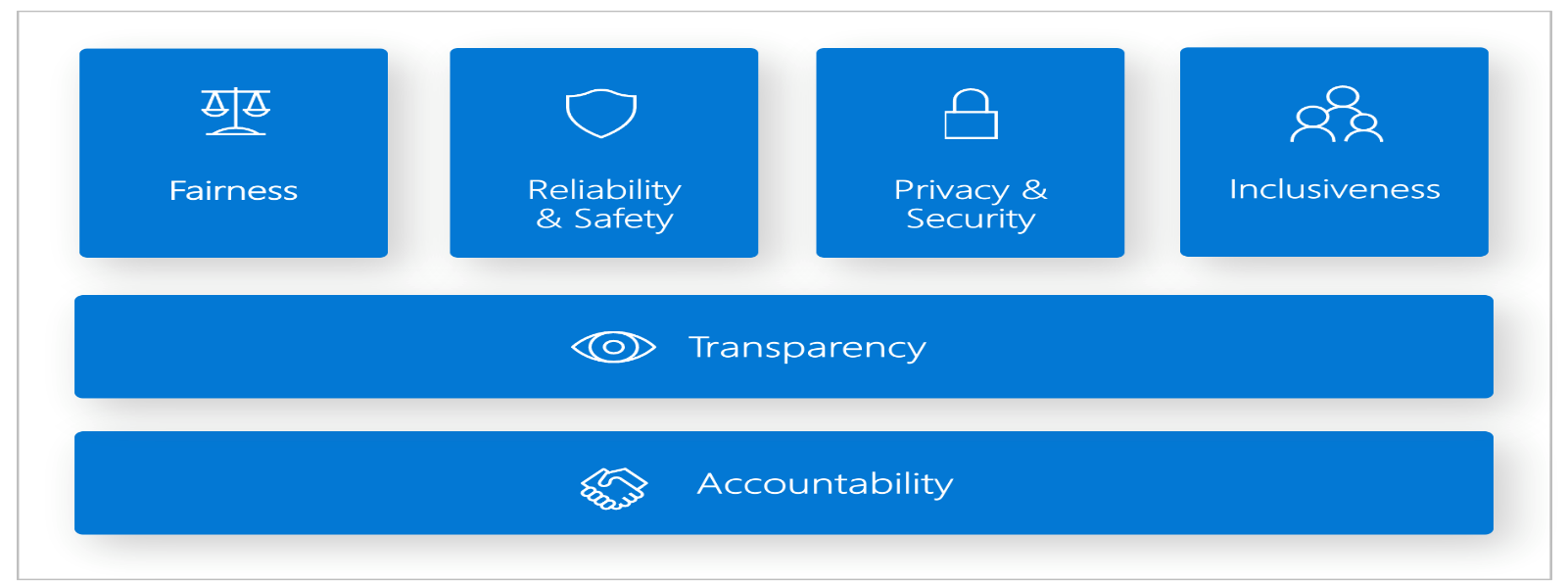
**Responsible AI principles**

Responsible Artificial Intelligence (Responsible AI) is an approach to developing, assessing, and deploying AI systems in a safe, trustworthy, and ethical way.



Six key principles define responsible AI:

* **Fairness**:

AI systems should treat everyone equally and provide the same recommendations to all individuals. Fairness in AI systems prevents discrimination based on personal characteristics.

* **Inclusiveness**: AI systems should empower and engage everyone. Inclusive design practices can help AI system developers understand and address potential exclusion barriers in a product or service. Inclusiveness fosters innovation and helps design experiences that benefit everyone.

**Example**, when AI systems provide guidance on medical treatment, loan applications, or employment, they should make the same recommendations to everyone who has similar symptoms, financial circumstances, or professional qualification

**Fairness and inclusiveness in Azure Machine Learning**: The [fairness assessment](https://learn.microsoft.com/en-us/azure/machine-learning/concept-fairness-ml?view=azureml-api-2) component of the [Responsible AI dashboard](https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai-dashboard?view=azureml-api-2) enables data scientists and developers to assess model fairness across sensitive groups defined in terms of gender, ethnicity, age, and other characteristics.

* **Reliability and safety**: AI systems must operate reliably, safely, and consistently under various conditions to help build trust.

These systems should be able to operate as they were originally designed, respond safely to unanticipated conditions, and resist harmful manipulation. How they behave and the variety of conditions they can handle reflect the range of situations and circumstances that developers anticipated during design and testing.

**Reliability and safety in Azure Machine Learning**: The [error analysis](https://learn.microsoft.com/en-us/azure/machine-learning/concept-error-analysis?view=azureml-api-2) component of the [Responsible AI dashboard](https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai-dashboard?view=azureml-api-2) enables data scientists and developers to:

* Get a deep understanding of how failure is distributed for a model.
* Identify cohorts (subsets) of data with a higher error rate than the overall benchmark.
* **Privacy and security**: AI systems should respect privacy and maintain security by protecting private and confidential information. They should also resist attacks and attempts to corrupt or compromise the system.

With AI, privacy and data security require close attention because access to data is essential for AI systems to make accurate and informed predictions and decisions about people. AI systems must comply with privacy laws that:

* Require transparency about the collection, use, and storage of data.
* Mandate that consumers have appropriate controls to choose how their data is used.

**Privacy and security in Azure Machine Learning**: Azure Machine Learning enables administrators and developers to [create a secure configuration that complies](https://learn.microsoft.com/en-us/azure/machine-learning/concept-enterprise-security?view=azureml-api-2) with their companies' policies. With Azure Machine Learning and the Azure platform, users can:

* Restrict access to resources and operations by user account or group.
* Restrict incoming and outgoing network communications.
* Encrypt data in transit and at rest.
* Scan for vulnerabilities.
* Apply and audit configuration policies.

Microsoft also created two open-source packages that can enable further implementation of privacy and security principles:

* [SmartNoise](https://github.com/opendifferentialprivacy/smartnoise-core): Differential privacy is a set of systems and practices that help keep the data of individuals safe and private. In machine learning solutions, differential privacy might be required for regulatory compliance. SmartNoise is an open-source project (co-developed by Microsoft) that contains components for building differentially private systems that are global.
* [Counterfit](https://github.com/Azure/counterfit/): Counterfit is an open-source project that comprises a command-line tool and generic automation layer to allow developers to simulate cyberattacks against AI systems. Anyone can download the tool and deploy it through Azure Cloud Shell to run in a browser or deploy it locally in an Anaconda Python environment. It can assess AI models hosted in various cloud environments, on-premises, or in the edge. The tool is agnostic to AI models and supports various data types, including text, images, or generic input.
* **Transparency**: AI systems should be transparent and understandable. AI systems can inform decisions that can deeply affect people's lives, so it's crucial for individuals to understand how the system makes these decisions.

**Example**, a bank might use an AI system to decide whether a person is creditworthy. A company might use an AI system to determine the most qualified candidates to hire.

A crucial part of transparency is *interpretability*: the useful explanation of the behavior of AI systems and their components. Improving interpretability requires stakeholders to comprehend how and why AI systems function the way they do. The stakeholders can then identify potential performance issues, fairness issues, exclusionary practices, or unintended outcomes.

**Transparency in Azure Machine Learning**: The [model interpretability](https://learn.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability?view=azureml-api-2) and [counterfactual what-if](https://learn.microsoft.com/en-us/azure/machine-learning/concept-counterfactual-analysis?view=azureml-api-2) components of the [Responsible AI dashboard](https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai-dashboard?view=azureml-api-2) enable data scientists and developers to generate human-understandable descriptions of the predictions of a model.

The model interpretability component provides multiple views into a model's behavior:

* + *Global explanations*. For example, what features affect the overall behavior of a loan allocation model?
  + *Local explanations*. For example, why was a customer's loan application approved or rejected?
  + *Model explanations for a selected cohort of data points*. For example, what features affect the overall behavior of a loan allocation model for low-income applicants?
* **Accountability**: AI systems and their developers should be accountable and answerable.

Organizations should draw upon industry standards to develop accountability norms. These norms can ensure that AI systems aren't the final authority on any decision that affects people's lives. They can also ensure that humans maintain meaningful control over otherwise highly autonomous AI systems.

**Accountability in Azure Machine Learning**: [Machine learning operations (MLOps)](https://learn.microsoft.com/en-us/azure/machine-learning/concept-model-management-and-deployment?view=azureml-api-2) is based on DevOps principles and practices that increase the efficiency of AI workflows. Azure Machine Learning provides the following MLOps capabilities for better accountability of your AI systems:

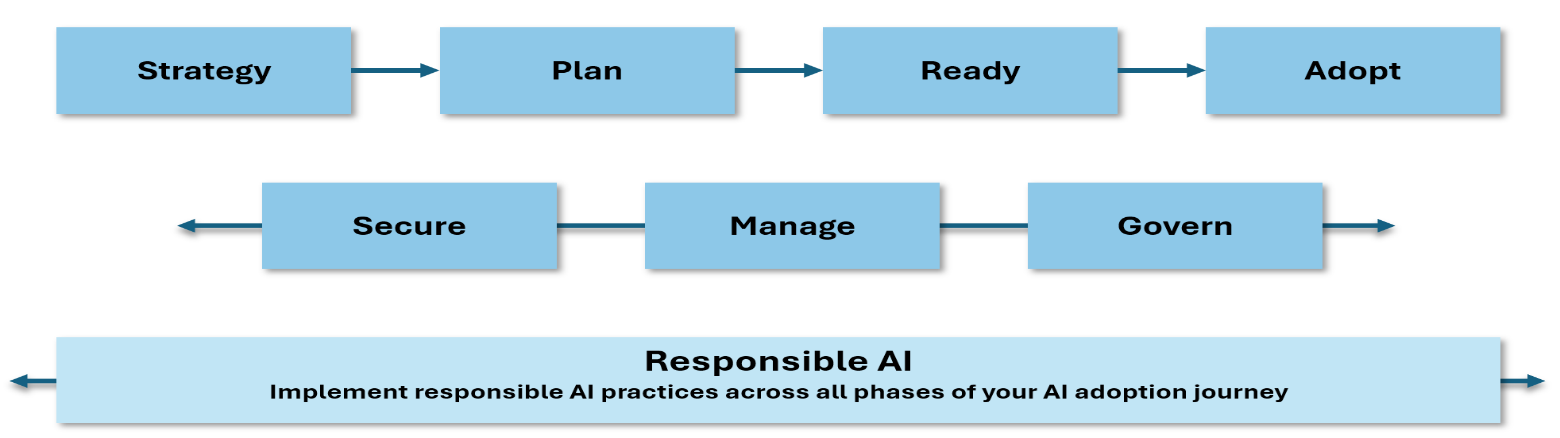
* Register, package, and deploy models from anywhere. You can also track the associated metadata that's required to use the model.
* Capture the governance data for the end-to-end machine learning lifecycle. The logged lineage information can include who is publishing models, why changes were made, and when models were deployed or used in production.
* Notify and alert on events in the machine learning lifecycle. Examples include experiment completion, model registration, model deployment, and data drift detection.
* Monitor applications for operational issues and issues related to machine learning. Compare model inputs between training and inference, explore model-specific metrics, and provide monitoring and alerts on your machine learning infrastructure.

Besides the MLOps capabilities, the [Responsible AI scorecard](https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai-scorecard?view=azureml-api-2) in Azure Machine Learning creates accountability by enabling cross-stakeholder communications. The scorecard also creates accountability by empowering developers to configure, download, and share their model health insights with their technical and non-technical stakeholders about AI data and model health. Sharing these insights can help build trust.

The machine learning platform also enables decision-making by informing business decisions through:

* **Data-driven insights**, to help stakeholders understand causal treatment effects on an outcome, by using historical data only. For example, "How would a medicine affect a patient's blood pressure?" These insights are provided through the [causal inference](https://learn.microsoft.com/en-us/azure/machine-learning/concept-causal-inference?view=azureml-api-2) component of the [Responsible AI dashboard](https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai-dashboard?view=azureml-api-2).
* **Model-driven insights**, to answer users' questions (such as "What can I do to get a different outcome from your AI next time?") so they can take action. Such insights are provided to data scientists through the [counterfactual what-if](https://learn.microsoft.com/en-us/azure/machine-learning/concept-counterfactual-analysis?view=azureml-api-2) component of the [Responsible AI dashboard](https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai-dashboard?view=azureml-api-2).

Incorporate responsible AI principles throughout your AI adoption journey, from strategy and planning to implementation. Ensure that you apply these principles when you secure, manage, and govern your AI initiatives.



**The importance of responsible AI**

Responsible AI helps to mitigate the following risks:

* **Unintended consequences**: Plan and oversee your responsible AI implementation to reduce the risk of unforeseen effects that have ethical implications.
* **Evolving threats**: Novel threats emerge regularly as AI technology evolves. To help mitigate and stay ahead of these threats, adhere to the principles of responsible AI.
* **Bias**: Bias mitigation in AI can be challenging but is necessary to ensure that AI systems are fair and unbiased. Use the responsible AI principles to help guide you.
* **Sensitive technologies**: Technologies like facial recognition can be considered sensitive technology because of the risk to fundamental freedoms and human rights. Consider the implications of these technologies to ensure that you use them responsibly.