A systematic approach to identifying AI risks examines each category of risk in each business context.

Transparency and explainability.

A lack of transparency around how a model was developed (such as how data sets feeding into a model were combined) or the inability to explain how a model arrived at a particular result can lead to issues, not the least of which is potentially running afoul of legal mandates. For example, if a consumer initiates an inquiry into how his or her data were used, the organization using the data will need to know into which models the data were fed.

Safety and performance.

AI applications, if not implemented and tested properly, can suffer performance issues that breach contractual guarantees and, in extreme cases, pose threats to personal safety. Suppose a model is used to ensure timely updates of machinery in manufacturing or mining; a failure of this model could constitute negligence under a contract and/or lead to employee harm.

Third-party risks.

The process of building an AI model often involves third parties. For example, organizations may outsource data collection, model selection, or deployment environments. The organization engaging third parties must know and understand the risk-mitigation and governance standards applied by each third party, and it should independently test and audit all high-stakes inputs.

In our experience, most AI risks map to at least one of the overarching risk types just described, and they often span multiple types. For example, a model-extraction attack, in which a model is stolen based on a sample set of outputs, compromises both the privacy and security of the model.

Therefore, organizations should ask if each category of risk could result from each AI model or tool the company is considering or already using.Pinpointing the context in which these risks can occur can help provide guidance as to where mitigation measures should be directe

Transparency and explainability.

A lack of transparency around how a model was developed (such as how data sets feeding into a model were combined) or the inability to explain how a model arrived at a particular result can lead to issues, not the least of which is potentially running afoul of legal mandates. For example, if a consumer initiates an inquiry into how his or her data were used, the organization using the data will need to know into which models the data were fed.

Safety and performance.

AI applications, if not implemented and tested properly, can suffer performance issues that breach contractual guarantees and, in extreme cases, pose threats to personal safety. Suppose a model is used to ensure timely updates of machinery in manufacturing or mining; a failure of this model could constitute negligence under a contract and/or lead to employee harm.

Third-party risks.

The process of building an AI model often involves third parties. For example, organizations may outsource data collection, model selection, or deployment environments.

The organization engaging third parties must know and understand the risk-mitigation and governance standards applied by each third party, and it should independently test and audit all high-stakes inputs. In our experience, most AI risks map to at least one of the overarching risk types just described, and they often span multiple types. For example, a model-extraction attack, in which a model is stolen based on a sample set of outputs, compromises both the privacy and security of the model. Therefore, organizations should ask if each category of risk could result from each AI model or tool the company is considering or already using