

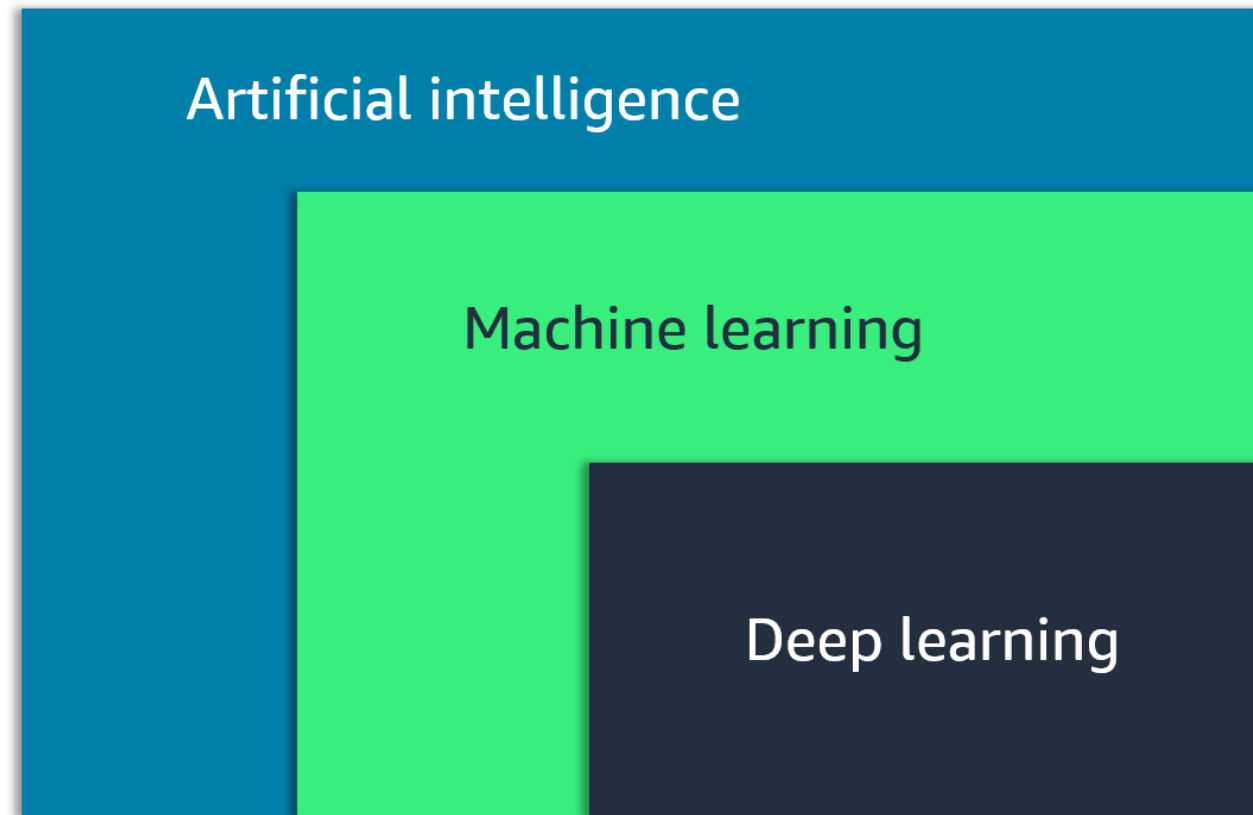
Building Machine Learning Pipelines

Outline

- The Machine Learning Workflow
- The Machine Learning Pipeline

The Machine Learning Workflow

Artificial Intelligence, Machine Learning, Deep Learning



Source: Amazon Web Services

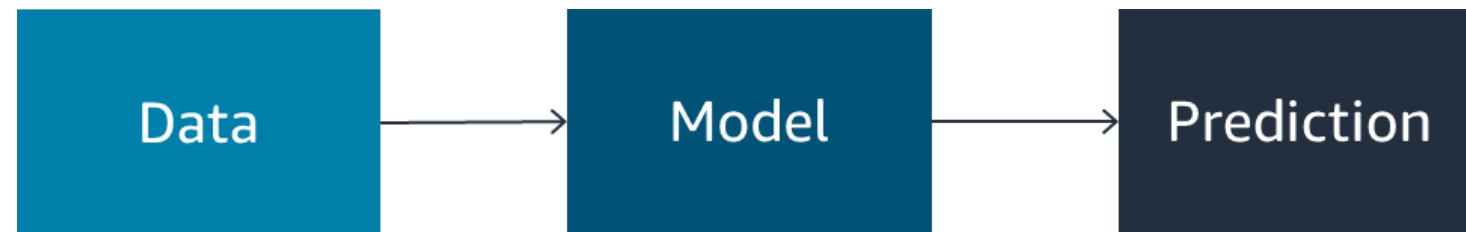
Machine Learning: Definition

- Artificial Intelligence (AI) is the broad field of **building machines to perform human tasks**. Machine learning is a sub-domain of AI.
- Machine Learning is the scientific study of **algorithms and statistical models** to perform a task using **inference** instead of instructions.
- Machine Learning focuses on *using data* to *train ML models* so these models can *make predictions*.

Source: Amazon Web Services

Machine Learning Flow: 3 Steps

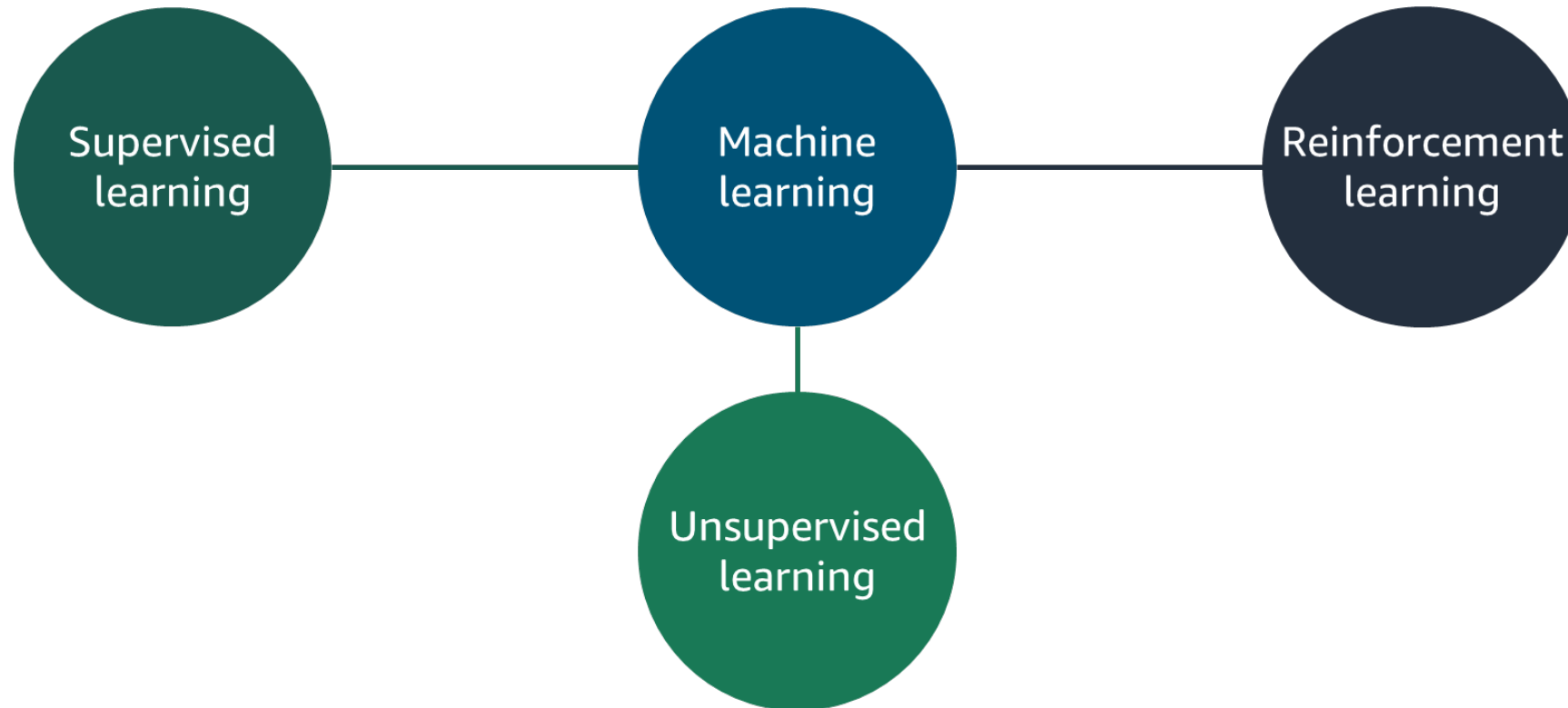
Machine Learning focuses on *using data* to *train ML models* so these models can *make predictions*.



Machine learning flow

Source: Amazon Web Services

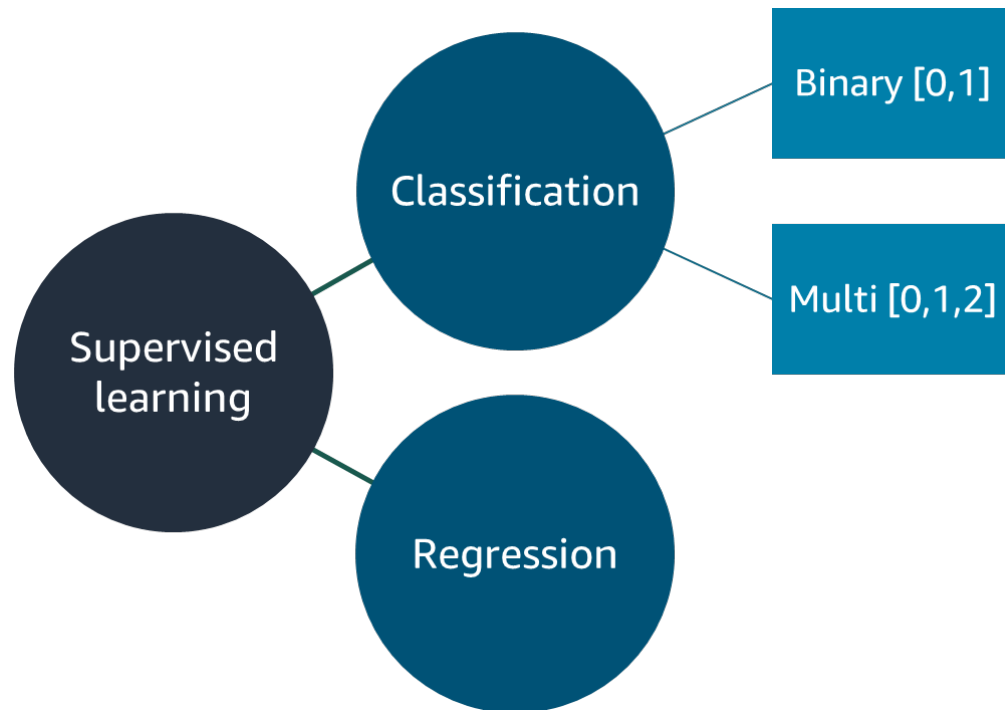
Types of Machine Learning



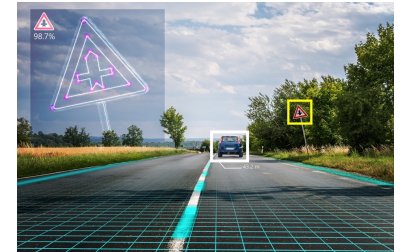
Source: Amazon Web Services

Supervised learning

Learn by identifying patterns in data that is **already labeled**.

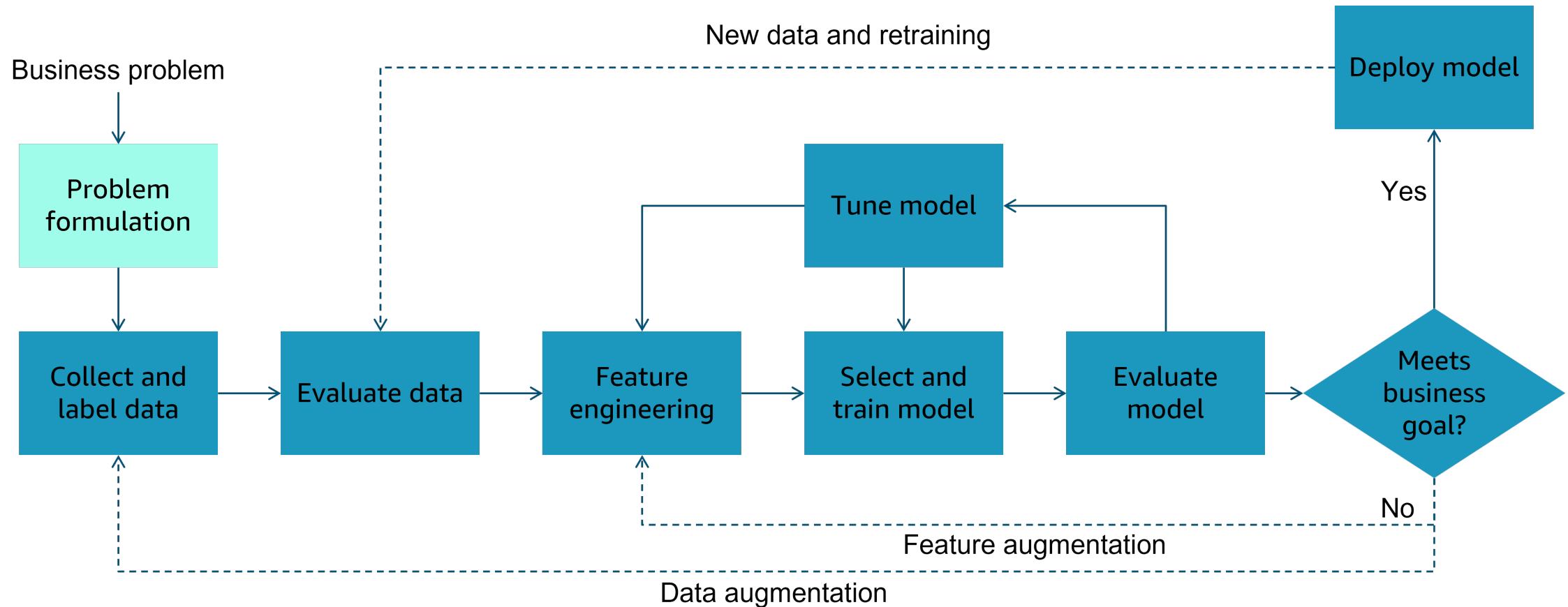


- Fraud detection
- Image recognition
- Personalized advertising
- Medical diagnostics
- Product sales prediction
- Weather forecasting
- Population growth prediction



Source: Amazon Web Services

Machine Learning Workflow



Source: Amazon Web Services

The Machine Learning Pipeline

Workflow / Pipeline Distinction

- **Workflow** describes the entire ML learning model development and deployment process including activities that need to be done by humans.
- **Pipeline** describes the data processing manifestation of the workflow process.
- In this course, our pipelines are going to be implemented using Python-based data science packages in Jupyter Notebooks.
- Our Jupyter Notebooks will include both implementing code and documentation of human decisions that impact the ML project.

Source: Amazon Web Services

The IBM Perspective on ML Pipelines

- The article
- Kevin's mind map

Source: Amazon Web Services

ML Jupyter Notebook Styles Evolve

- Exploratory Style:
 - More than one alternative explored.
 - Many data displays.
 - Most code not organized into functions.
- Transitional Style:
 - One approach emerges.
 - Fewer data displays.
 - Code migrates to functions placed in code cells.
- Production Style
 - Functions migrate to separate .py files.
 - Convenience features like results log files added.

Source: Amazon Web Services

Questions or Comments?