

Artificial Intelligence

Fundamental with Python



Kholed Langsari

AI/ML Engineer, Software Architect,
Instructor at Fatoni University

langsari@ftu.ac.th

<< Setup works environment

<< Python for AI/ML

The Machine Learning Landscape

End-to-End Machine Learning Project >>

Classification >>

What is Machine Learning?

What is ML?

- A field of study that **gives computers the ability to learn** without being explicitly programmed. — Arthur Samuel, 1959
- A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E . — Tom Mitchell, 1997
- ML is a prominent sub-field in AI

- The science (and art) of programming computers so they can ***learn from data***.

Why Use Machine Learning?

- A field of study that gives computers the ability to learn without being explicitly programmed.

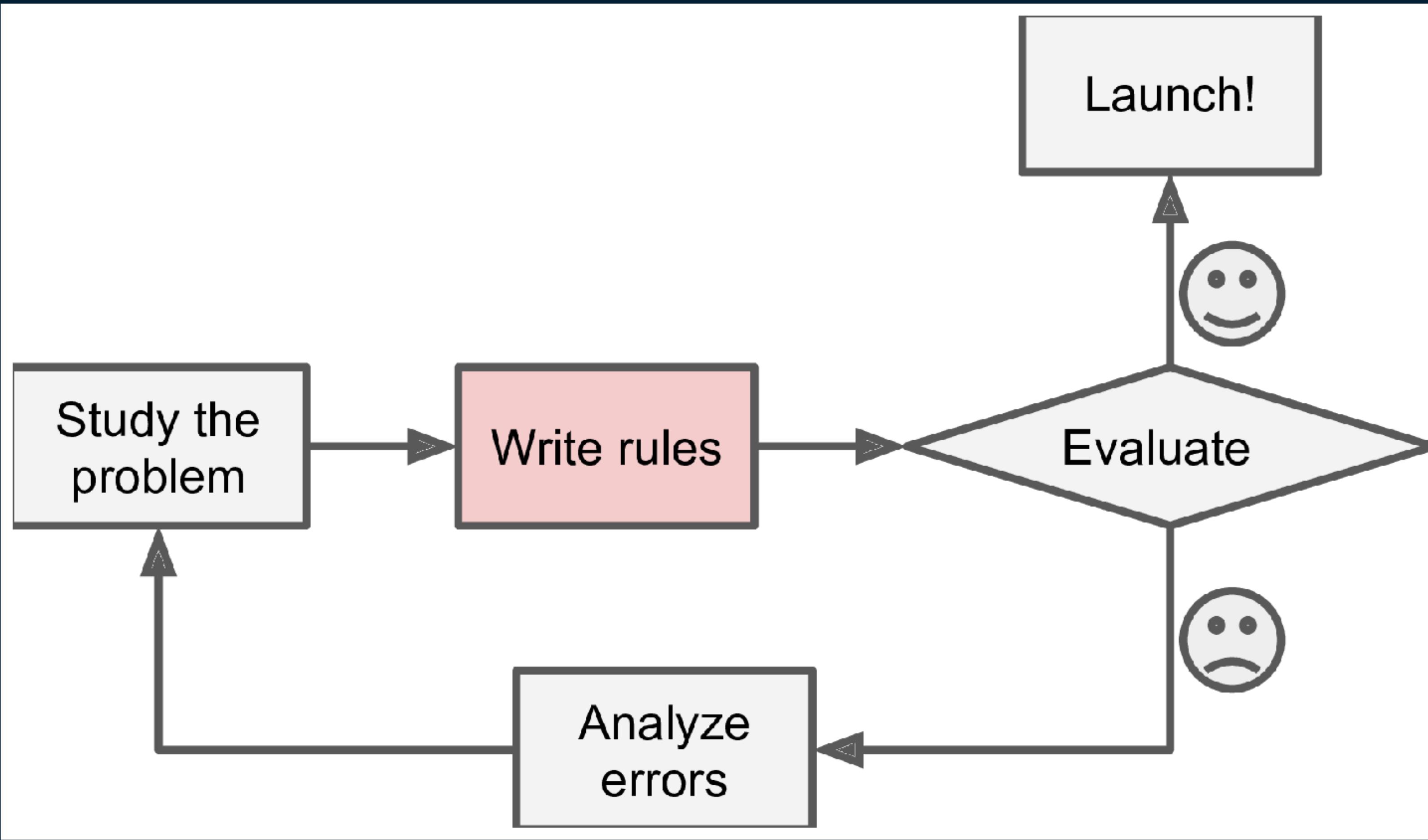
From Programming to Learning



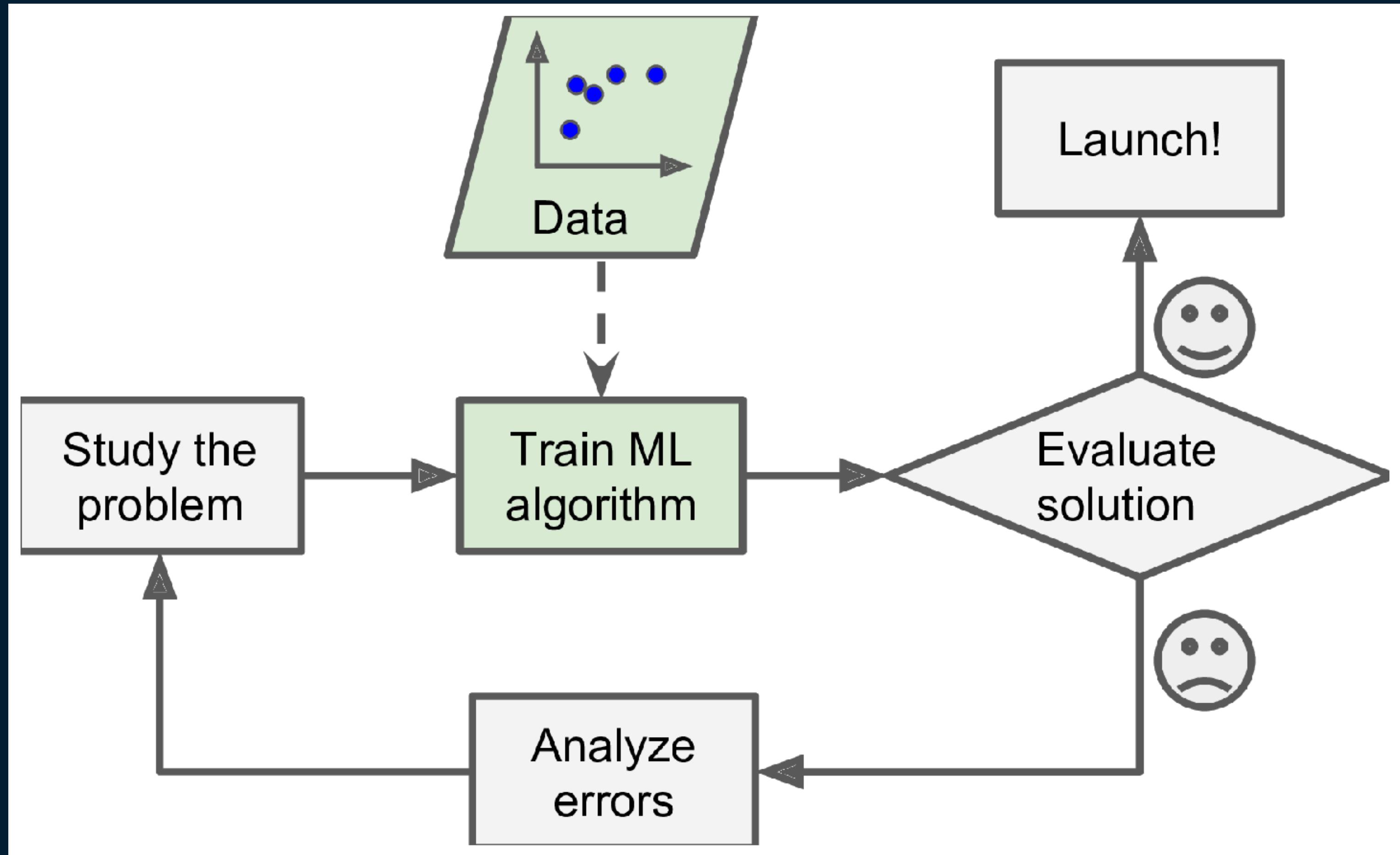
The traditional programming flow



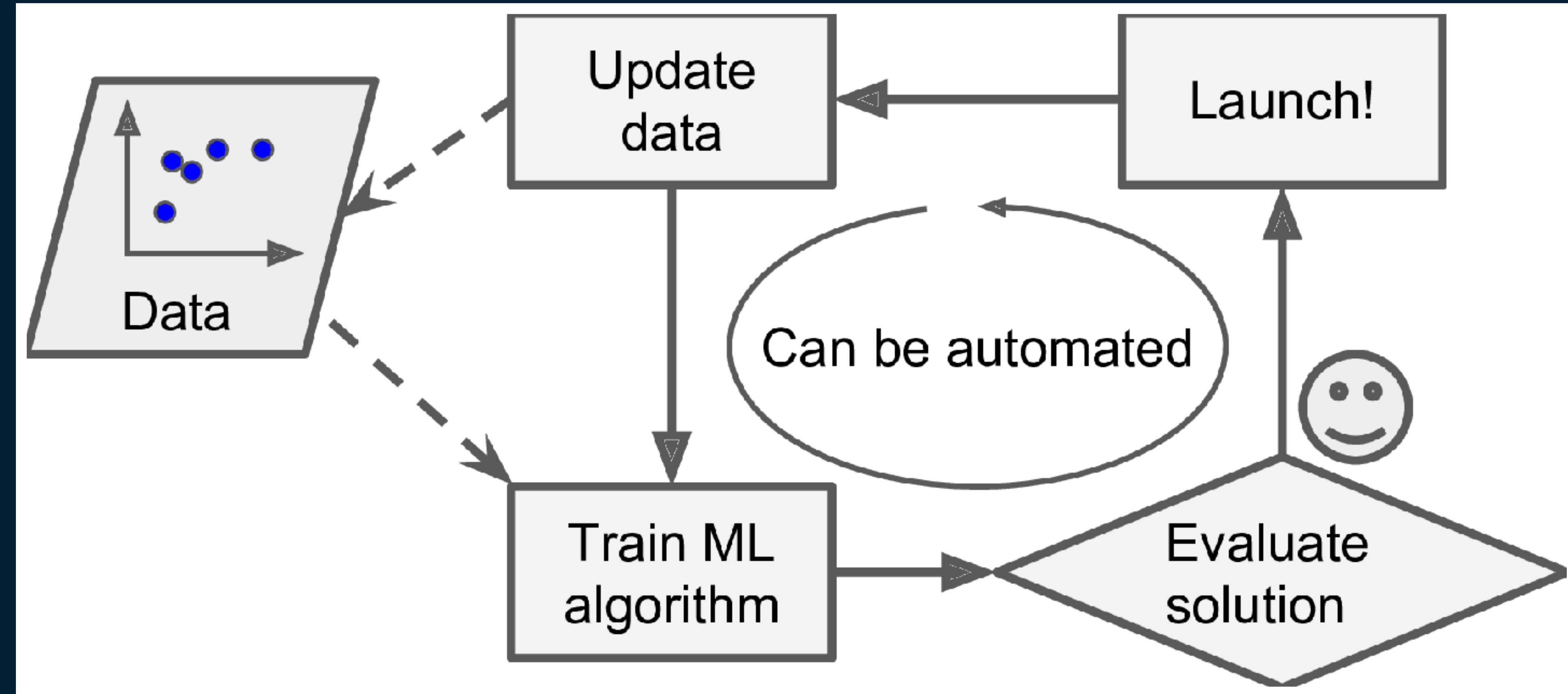
Changing the axes to get machine learning



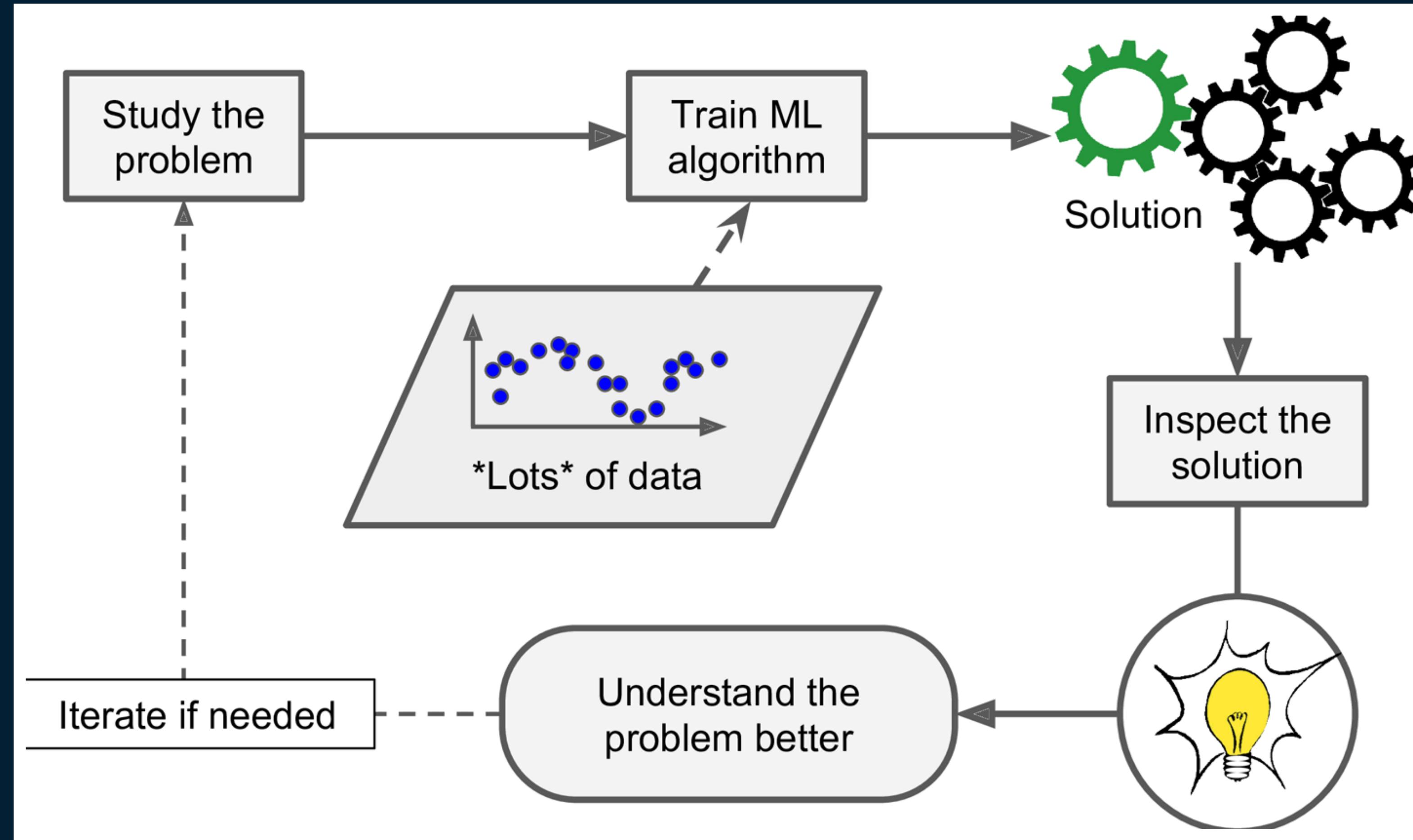
The traditional approach



The Machine Learning approach



Automatically adapting to change



Machine Learning can help humans learn

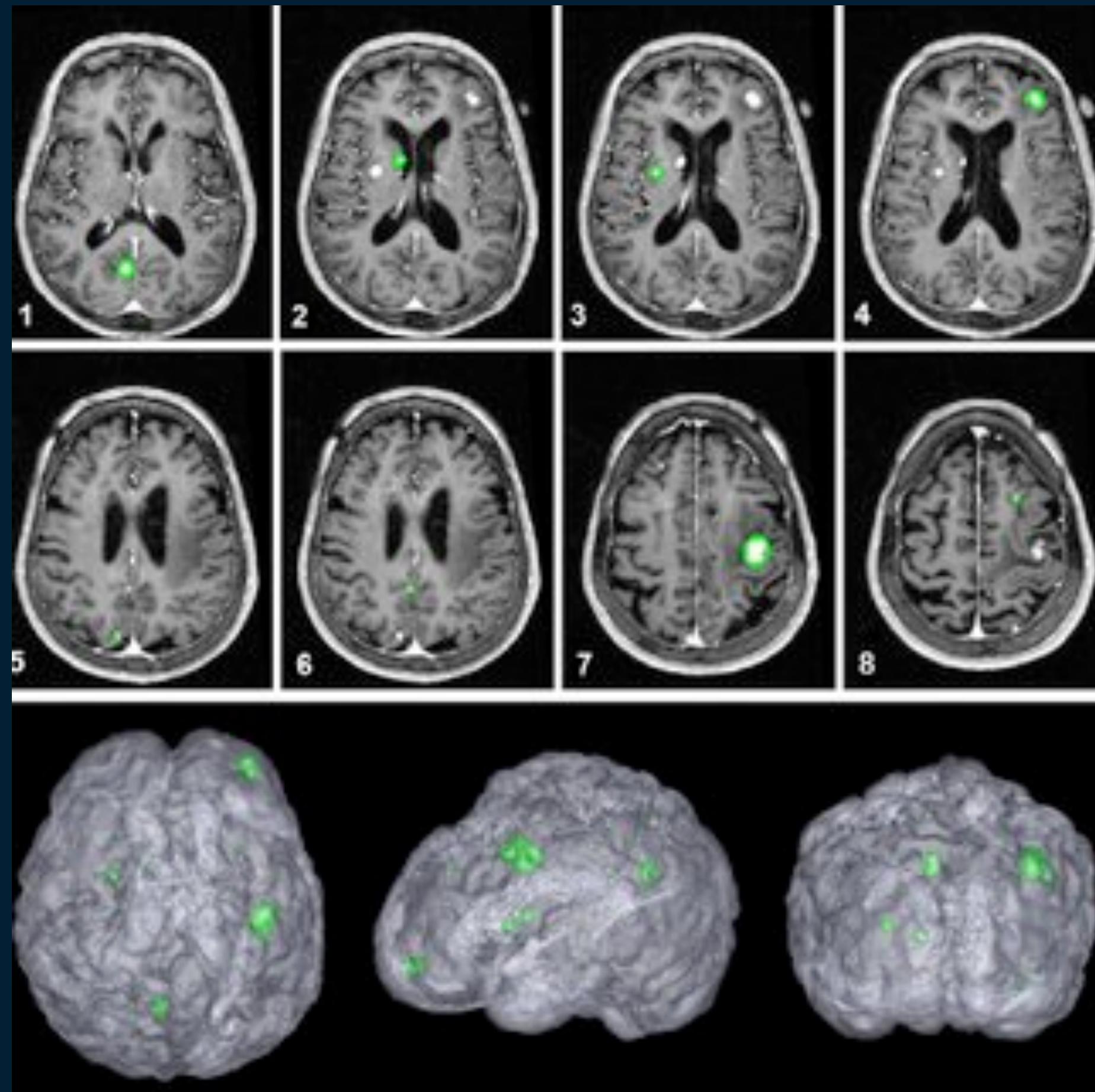
- A machine-learning system is **trained** rather than explicitly programmed.

Examples of ML tasks



Analyzing images of products on a production line to automatically classify them

This is image classification, typically performed using convolutional neural networks (CNN)



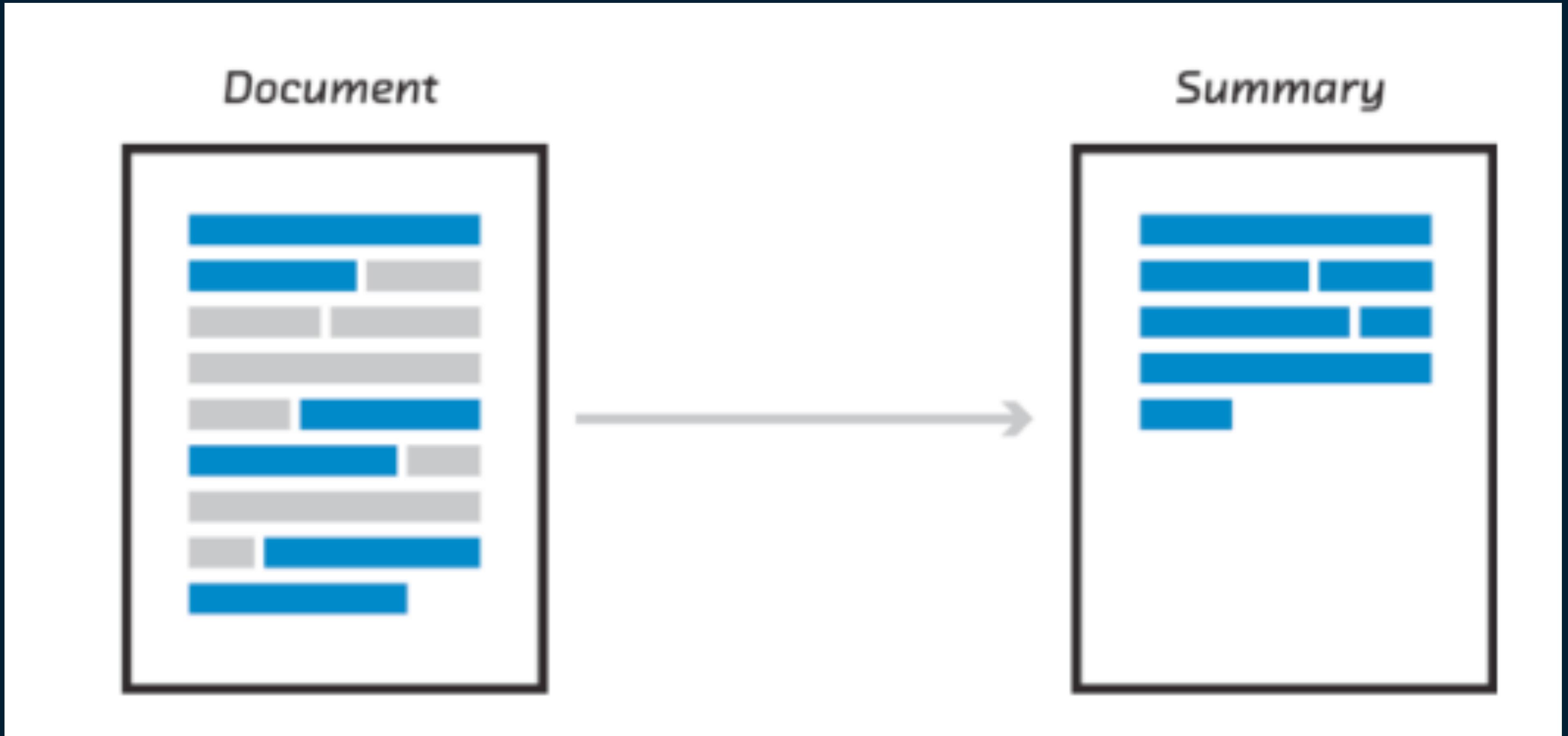
Detecting tumors in brain scans

This is semantic segmentation, where each pixel in the image is classified typically using CNNs as well



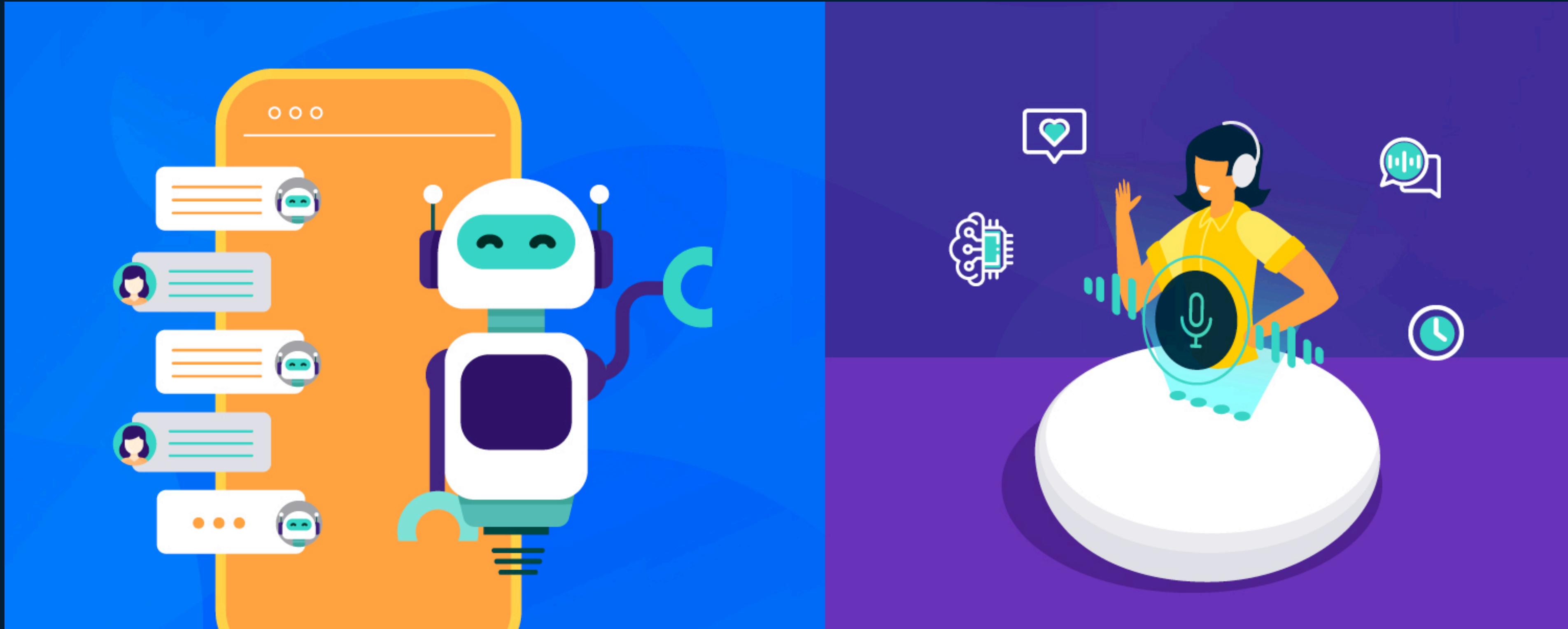
Automatically classifying news articles

This is natural language processing (NLP), and more specifically
text classification



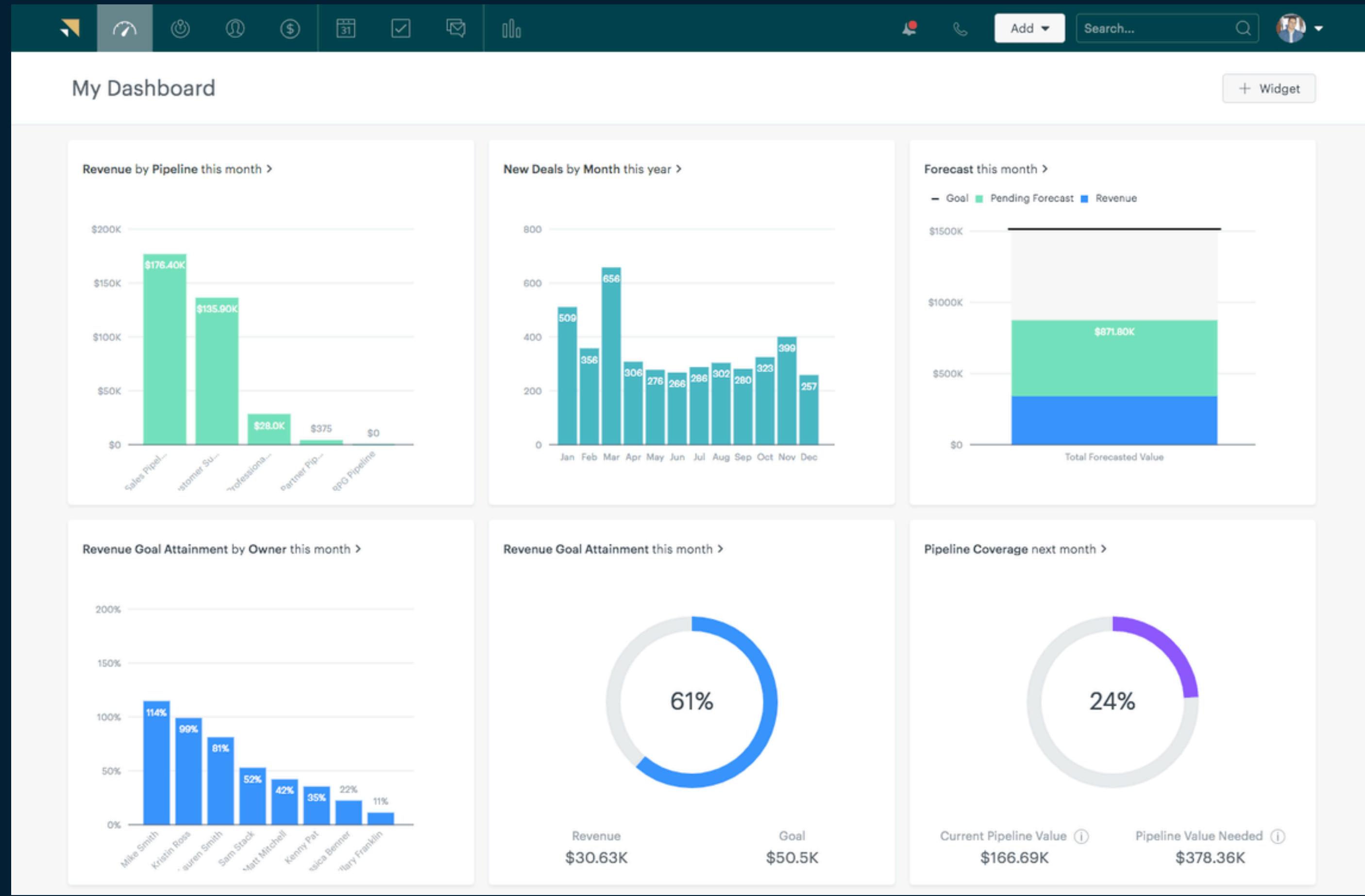
Summarizing long documents automatically

This is a branch of NLP called text summarization



Creating a chatbot or a personal assistant

This involves many NLP components, including natural language understanding (NLU) and question-answering modules.



*Forecasting your company's revenue next year,
based on many performance metrics*

This is a regression task (i.e., predicting values), Linear Regression or Polynomial Regression model



“Hey Siri, OK Google”

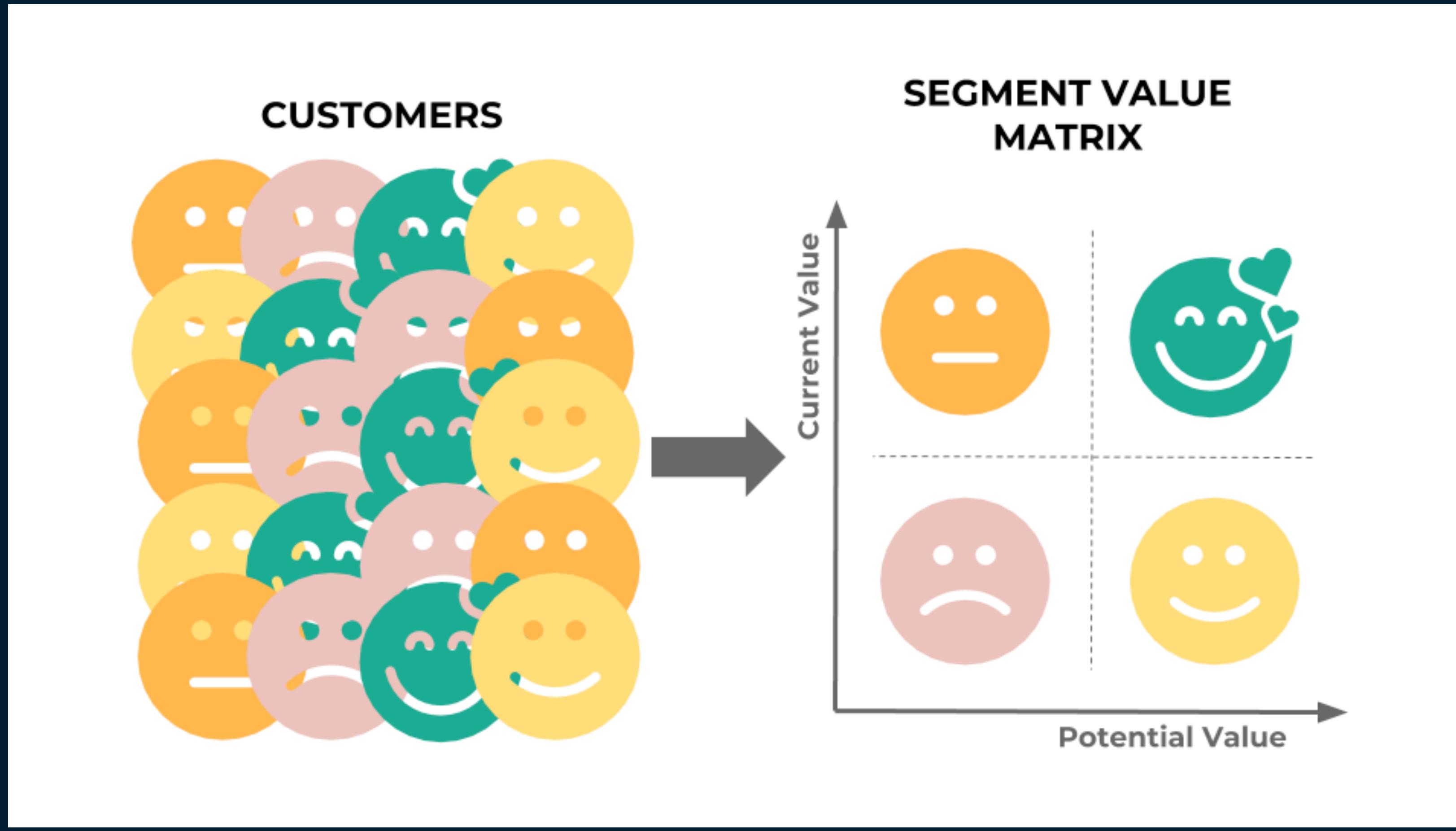
Making your app react to voice commands

This is speech recognition, typically processed using RNNs,
CNNs, or Transformers



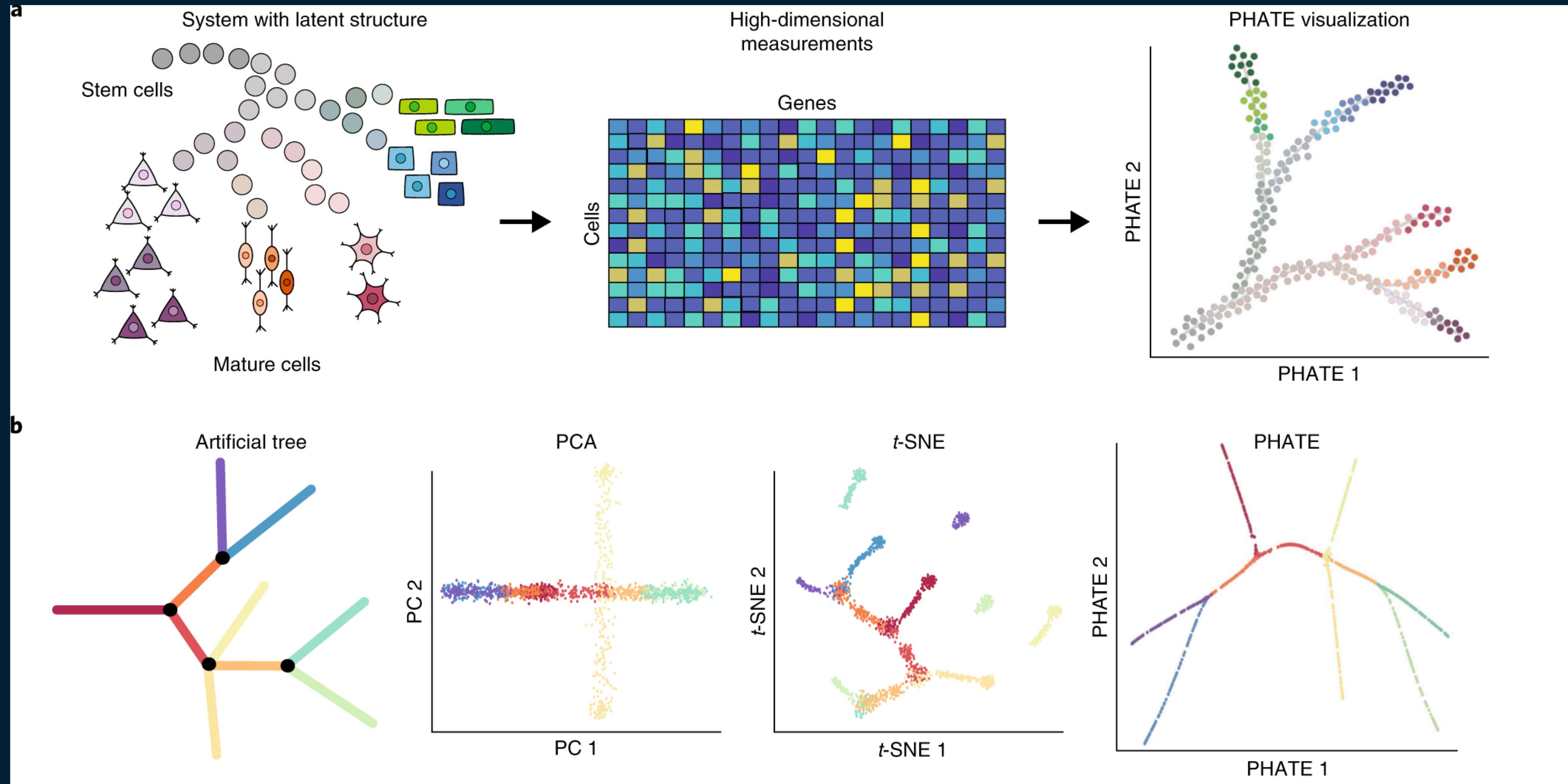
Detecting credit card fraud

This is anomaly detection



Segmenting clients based on their purchases so that you can design a different marketing strategy for each segment

This is clustering



Representing a complex, high-dimensional dataset in a clear and insightful diagram

This is data visualization, often involving dimensionality reduction techniques

Your recently viewed items and featured recommendations

Recommendations & Popular Items

Page 1 of 2

The screenshot displays a grid of recommended products. The top row includes SanDisk 32GB Ultra Class 10 SDHC UHS-I Memory Card, SanDisk Ultra 32GB microSDHC UHS-I Card with Adapter, SanDisk 64GB Ultra microSDXC UHS-I Memory Card with Adapter, Samsung 32GB 95MB/s (U1) MicroSD EVO Select Memory Card with Adapter, NETGEAR N300 WiFi Range Extender, and AmazonBasics Mini DisplayPort (Thunderbolt) to HDMI Adapter. The bottom row includes Kindle Unlimited books: The Magnolia Story, 1984, I Am Watching You, Dark Sacred Night, Girl, Wash Your Face, and Bleak Harbor.

Product	Description	Rating	Price
SanDisk 32GB Ultra Class 10 SDHC UHS-I Memory Card	Up to 80MB, Grey/Black (SDSDUNC...)	★★★★★ 7,448	\$8.99
SanDisk Ultra 32GB microSDHC UHS-I Card with Adapter	Grey/Red, Standard Packaging...	★★★★★ 31,062	\$8.99
SanDisk 64GB Ultra microSDXC UHS-I Memory Card with Adapter	100MB/s, C10, U1, Full...	★★★★★ 9,358	\$11.49
Samsung 32GB 95MB/s (U1) MicroSD EVO Select Memory Card with Adapter	(MB-ME32GA/AM)	★★★★★ 10,983	\$5.99
NETGEAR N300 WiFi Range Extender (EX2700)		★★★★★ 30,748	\$29.95
AmazonBasics Mini DisplayPort (Thunderbolt) to HDMI Adapter		★★★★★ 5,363	\$9.99
The Magnolia Story (with Bonus Content)	by Chip Gaines	★★★★★ 5,342	\$2.99
1984	by George Orwell	★★★★★ 6,594	\$2.99
I Am Watching You	by Teresa Driscoll	★★★★★ 7,459	\$1.99
Dark Sacred Night (A Ballard and Bosch Novel...)	by Michael Connelly	★★★★★ 626	\$14.99
Girl, Wash Your Face: Stop Believing the Lies About...	by Rachel Hollis	★★★★★ 7,349	\$6.99
Bleak Harbor: A Novel	by Bryan Gruley	★★★★★ 171	\$4.99

Page 2 of 8 | Start over

Recommending a product that a client may be interested in, based on past purchases
This is a recommender system



Building an intelligent bot for a game

This is often tackled using Reinforcement Learning (RL)

Types of Machine Learning Systems?

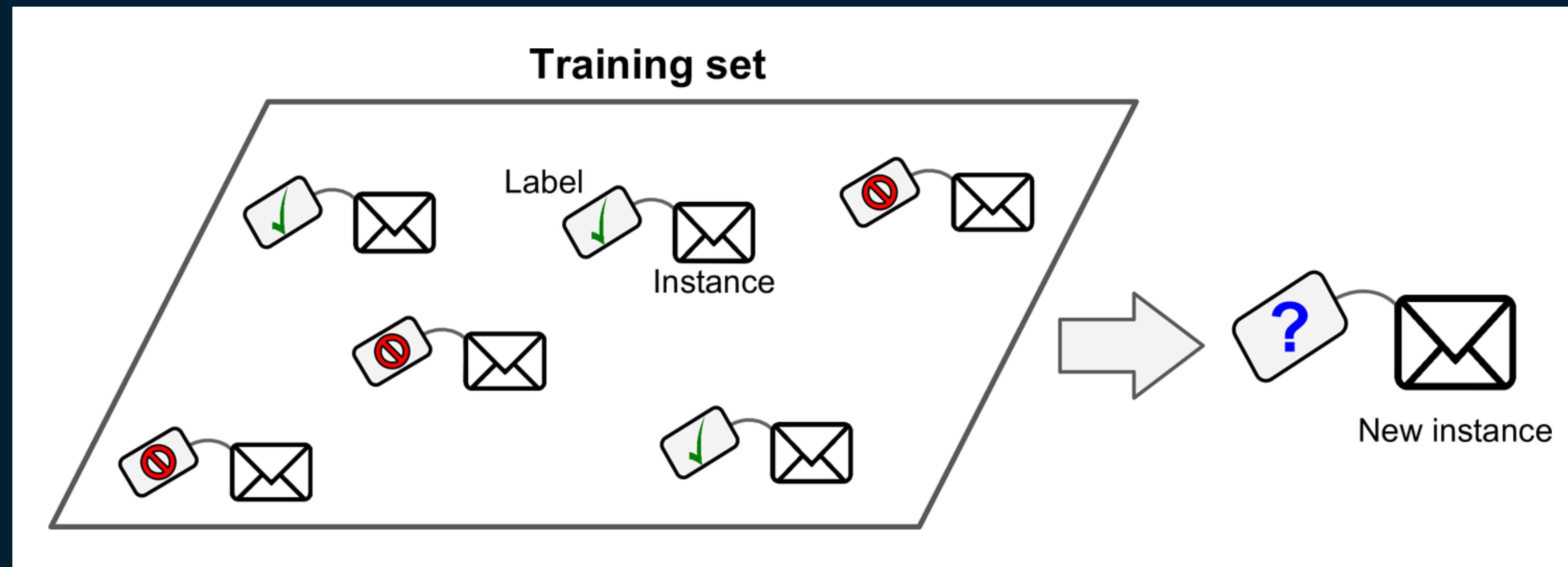
Types of Machine Learning Systems?

- Whether or not they are trained with human supervision (*supervised*, *unsupervised*, *semisupervised*, and *Reinforcement Learning*)
- Whether or not they can learn incrementally on the fly (*online* versus *batch learning*)
- Whether they work by simply comparing new data points to known data points, or instead by detecting patterns in the training data and building a predictive model, much like scientists do (*instance-based* versus *model-based learning*)

Supervised and Unsupervised Learning

Types of Machine Learning Systems?

- **Supervised Learning** - Training Data contains desired solutions, or *labels*
- Email spam classification

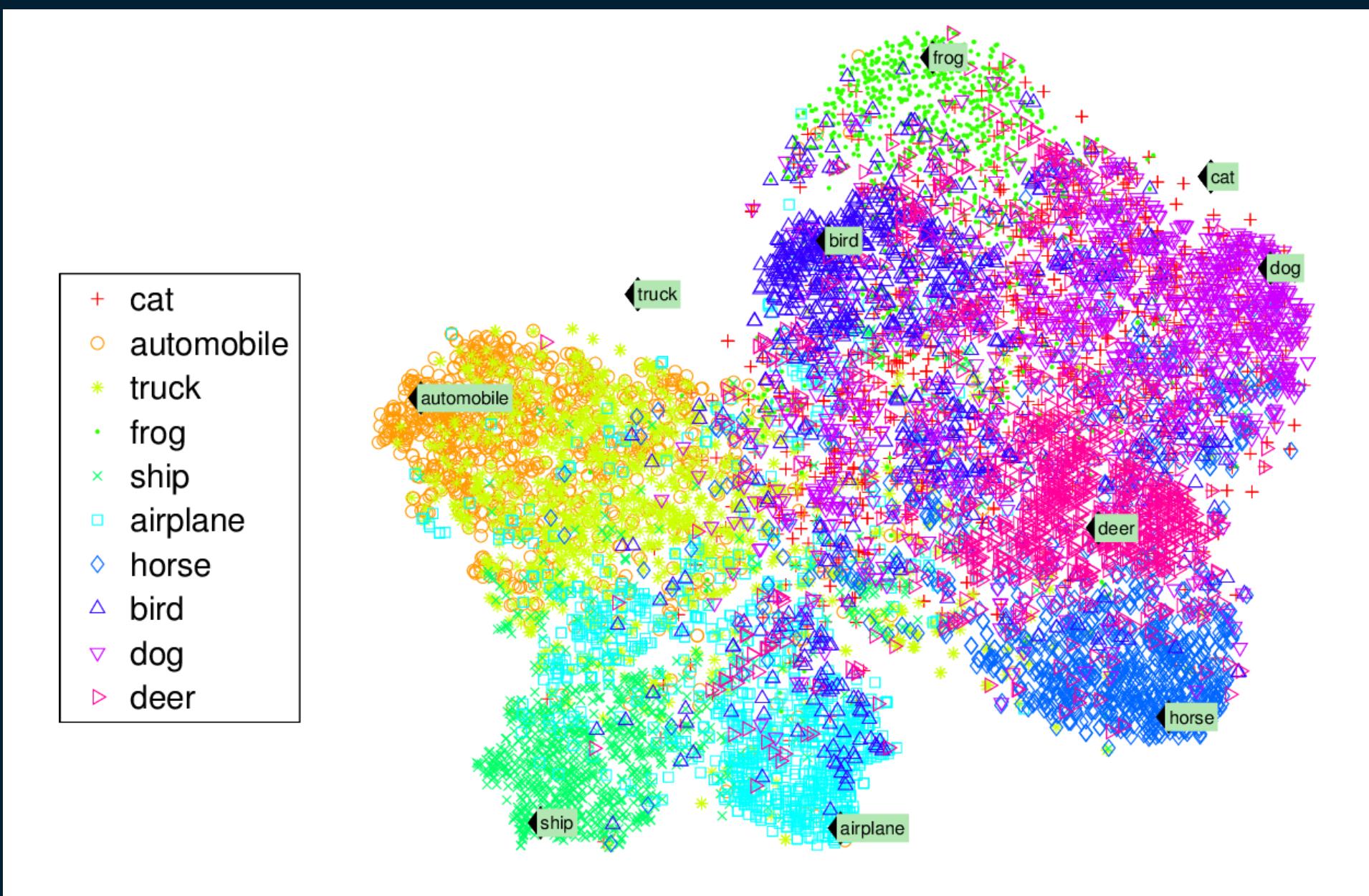


Supervised learning algorithms

- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees and Random Forests
- Neural networks

Types of Machine Learning Systems?

- **Unsupervised learning** is the training data is *unlabeled*. The system tries to learn without a teacher.
- t-SNE visualization highlighting semantic clusters of how animals are rather well separated from vehicles

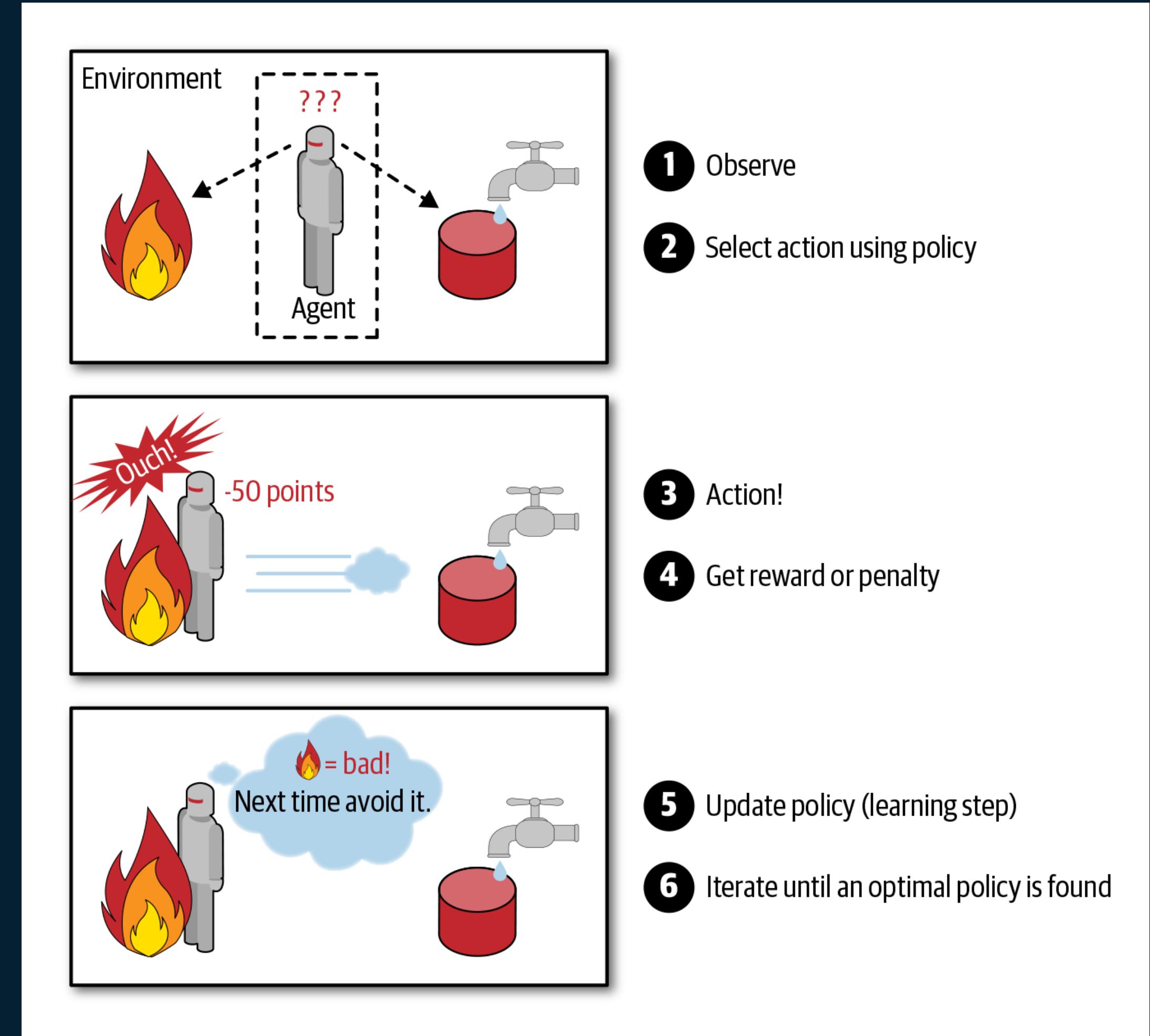


Unsupervised learning algorithms

- **Clustering**
 - K-Means
 - DBSCAN
 - Hierarchical Cluster Analysis (HCA)
- **Anomaly detection and novelty detection**
 - One-class SVM
 - Isolation Forest
- **Visualization and dimensionality reduction**
 - Principal Component Analysis (PCA)
 - Kernel PCA
 - Locally Linear Embedding (LLE)
 - t-Distributed Stochastic Neighbor Embedding (t-SNE)
- **Association rule learning**
 - Apriori
 - Eclat

Types of Machine Learning Systems?

- **Reinforcement Learning** - Training Data does not contain target output, but instead contains some possible output together with a measure of how good that output is.
- DeepMind's AlphaGo

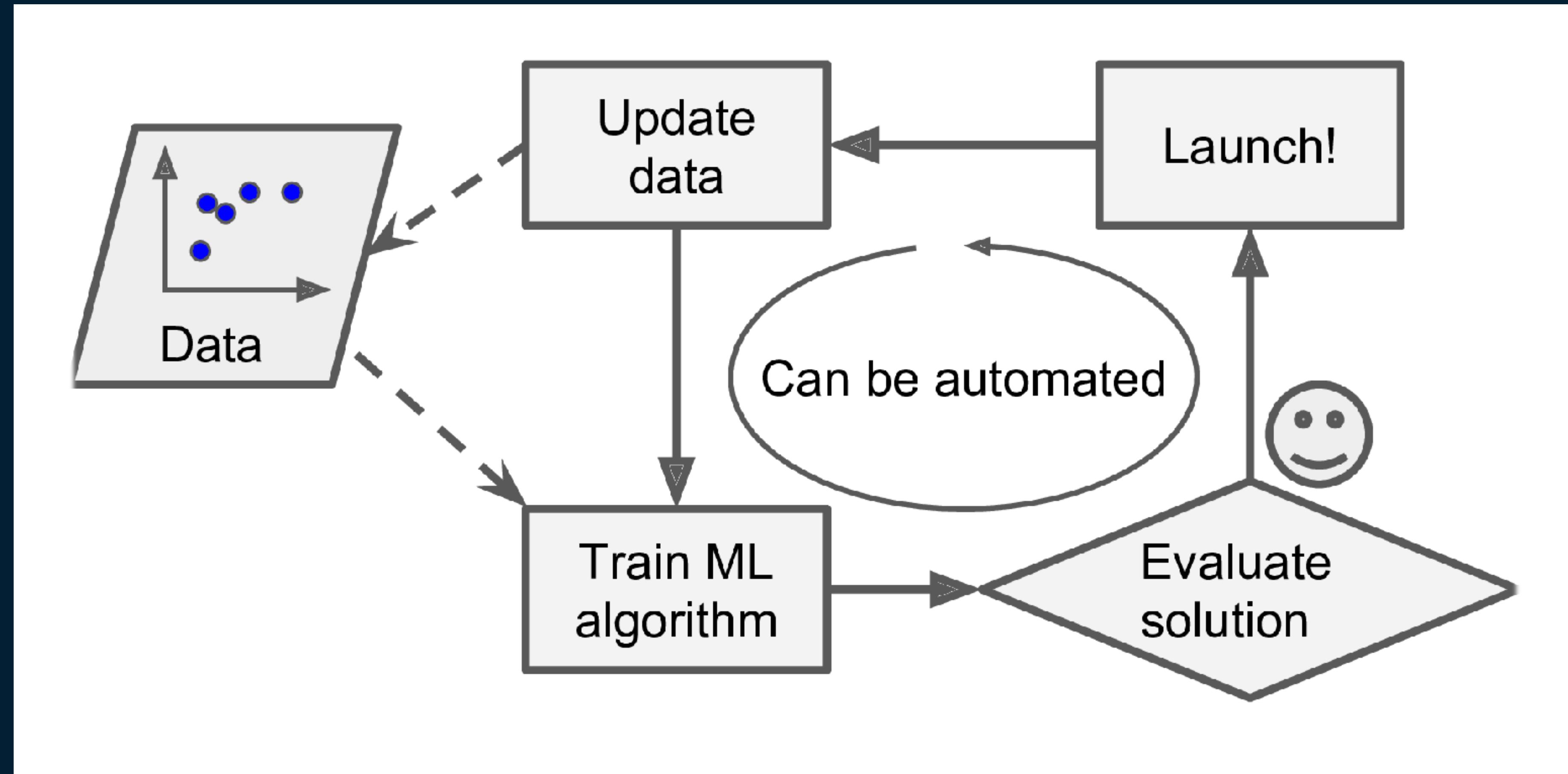


Batch and Online Learning

Batch Learning

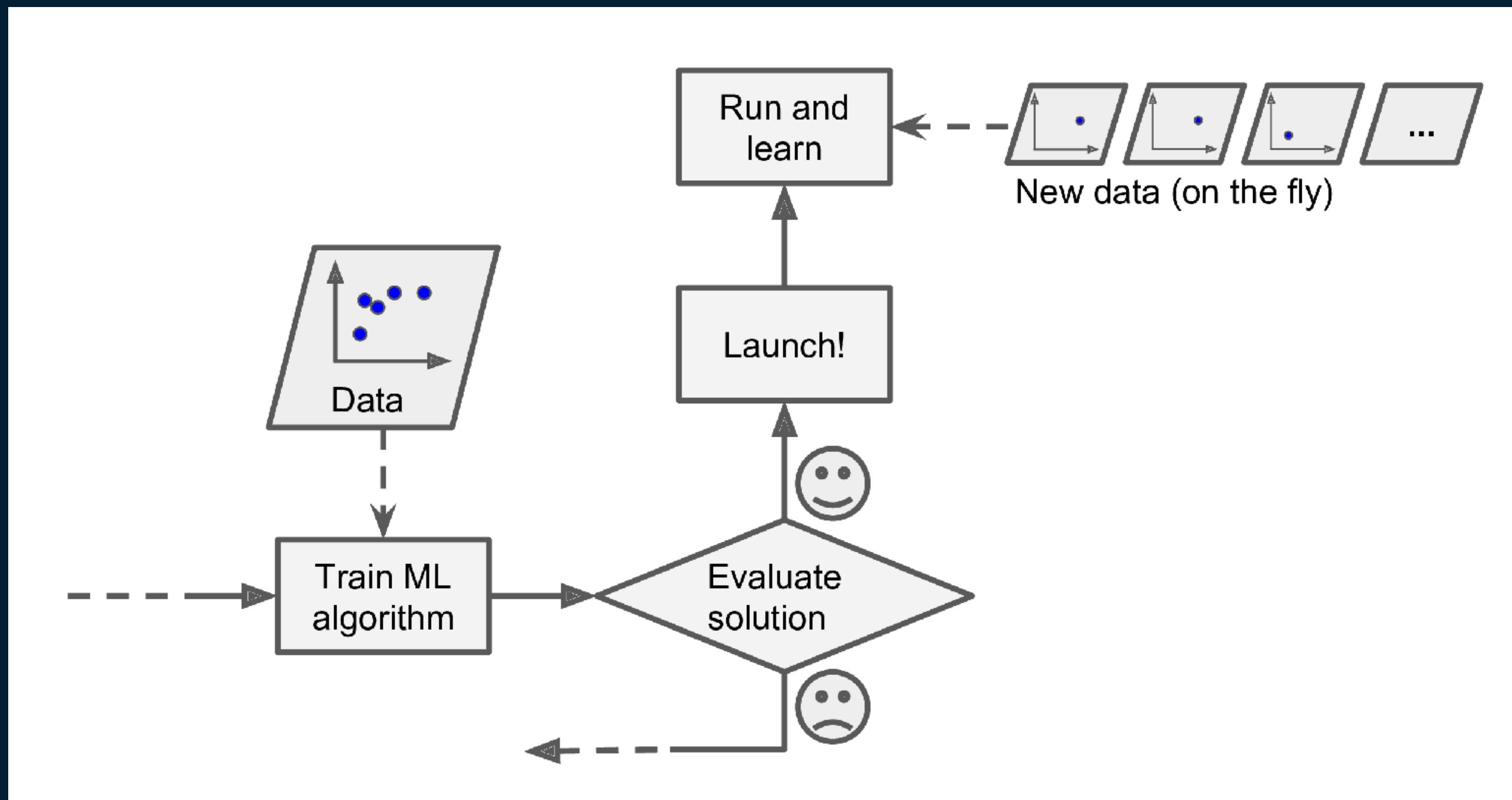
- The system is incapable of learning incrementally: it must be trained using all the available data.
- This will generally take a lot of time and computing resources, so it is typically done offline.
- First the system is trained, and then it is launched into production and runs without learning anymore; it just applies what it has learned. This is called **offline learning**.
- If you want a batch learning system to know about new data, you need to train a new version of the system from scratch on the full dataset (not just the new data, but also the old data), then stop the old system and replace it with the new one.

Batch Learning



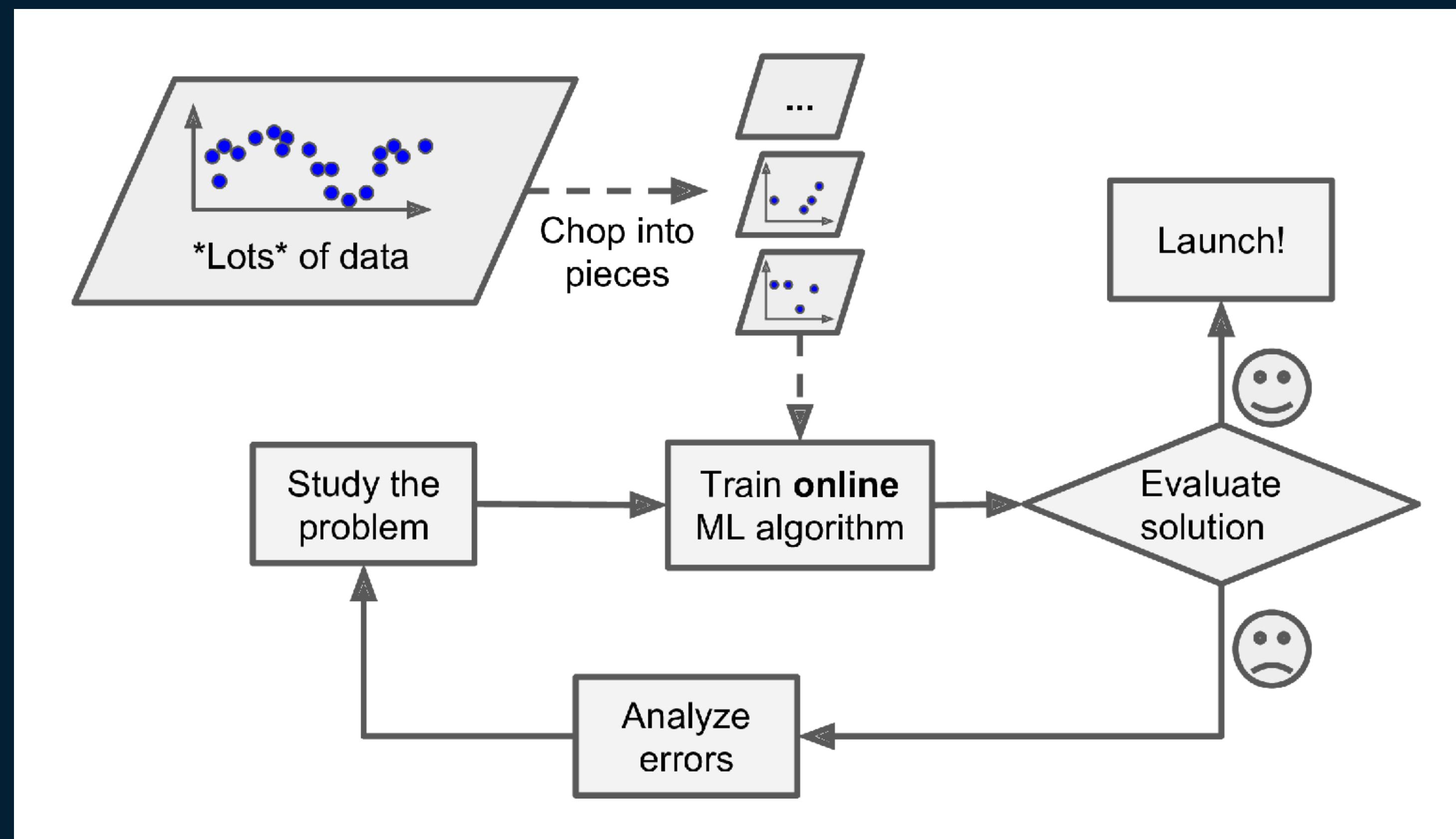
Online Learning

- You **train the system incrementally** by feeding it data instances sequentially, either individually or in small groups called mini-batches. Each learning step is fast and cheap, so the system can learn about new data on the fly, as it arrives



Online Learning

- Using online learning to handle huge datasets



Instance-Based Versus Model-Based Learning

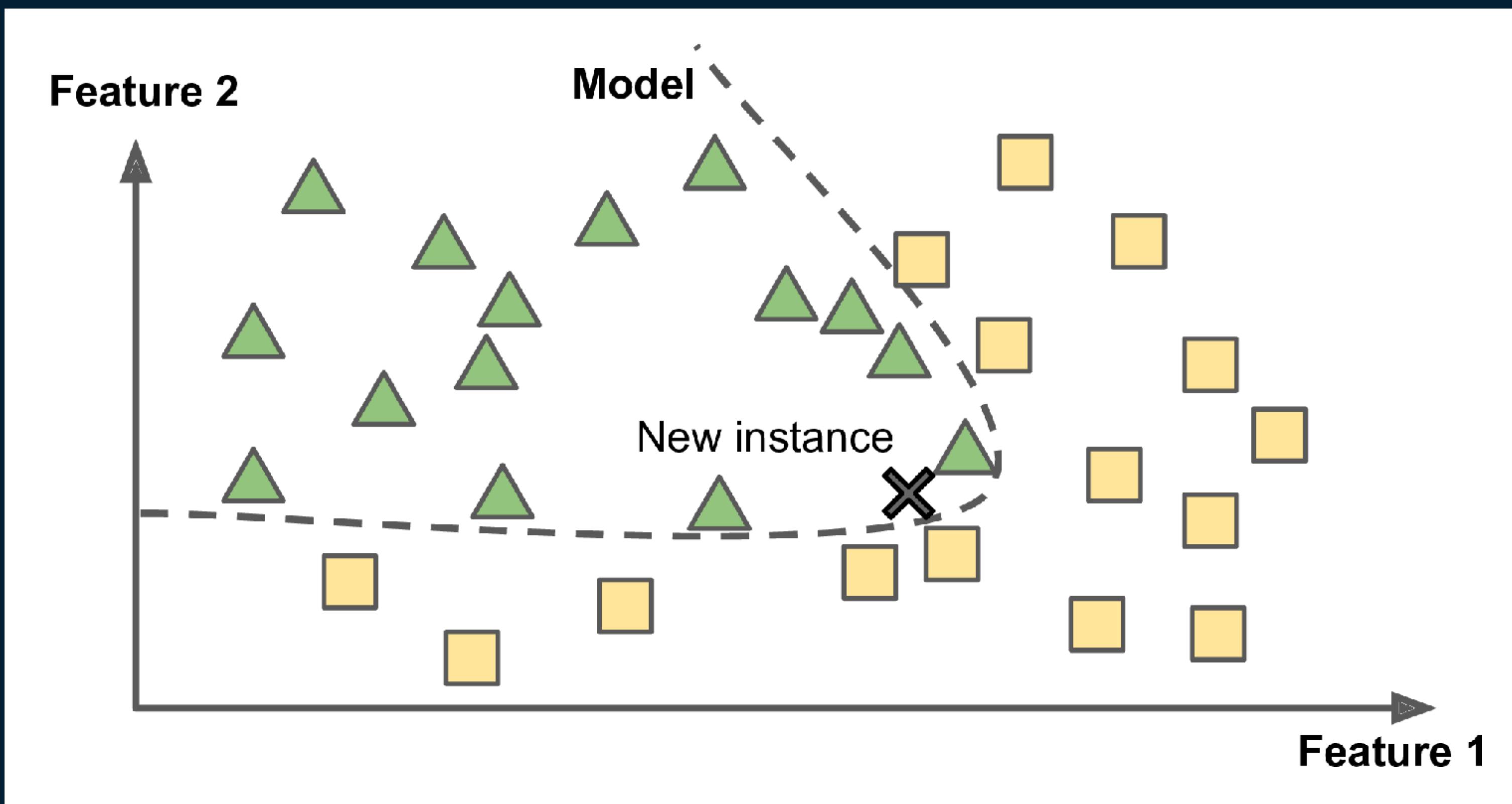
Instance-Based Learning

- The system learns the examples by **heart**, then generalizes to new cases by using a **similarity measure to compare** them to the learned examples (or a subset of them)



Model-Based Learning

- To generalize from a set of examples is to **build a model of these examples** and then use that **model to make predictions**.



Main Challenges of Machine Learning

Main Challenges of Machine Learning

- Insufficient Quantity of Training Data
- Nonrepresentative Training Data
- Poor-Quality Data
- Irrelevant Features
- Overfitting the Training Data
- Underfitting the Training Data

Testing and Validating

Testing and Validating

- **Training set** and the **test set**
- **Hyperparameter Tuning** and **Model Selection**
- **Data Mismatch**

Summary

- Machine Learning is about making machines get better at some task by learning from data, instead of having to explicitly code rules.
- Supervised or not, batch or online, instance-based or model-based.
- The system will not perform well if your training set is too small, or if the data is not representative, is noisy, or is polluted with irrelevant features (garbage in, garbage out).
- The model needs to be neither too simple (in which case it will underfit) nor too complex (in which case it will overfit).
- Testing and validating the model by splitting data to training set and the test set.
- And tuning hyperparameter and select the suitable Model for the system

Weekly Open Questions

Reply through [@thefutureisdata](#)

How would you define Machine Learning?

คุณจะนิยามแมชชีนเลิร์นนิ่งว่าอย่างไร?

Artificial Intelligence

Fundamental with Python



Kholed Langsari

AI/ML Engineer, Software Architect,
Instructor at Fatoni University

langsari@ftu.ac.th