

# Giới thiệu về Trí tuệ nhân tạo (AI)

AI là ngành khoa học tạo ra các hệ thống mô phỏng trí thông minh con người: học, lý luận, nhận thức, giải quyết vấn đề, giao tiếp và ra quyết định.



# Các Lĩnh vực Chính của AI



**Machine Learning (ML)**

Học máy



**Deep Learning**

Học sâu (mạng nơ-ron)



**Computer Vision**

Thị giác máy tính



**NLP**

Xử lý ngôn ngữ tự nhiên



**Robotics**

Robot & tự động hóa

# Ứng dụng AI trong Thực tế



## Y tế

Chẩn đoán bệnh từ ảnh X-quang/MRI.



## Giao thông

Xe tự lái, tối ưu hóa tuyến đường.



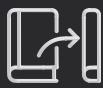
## Giáo dục

Học tập cá nhân hóa.



## Kinh doanh

Phân tích dữ liệu khách hàng, dự đoán doanh thu.



## Giải trí

Gợi ý phim, nhạc (Netflix, Spotify).



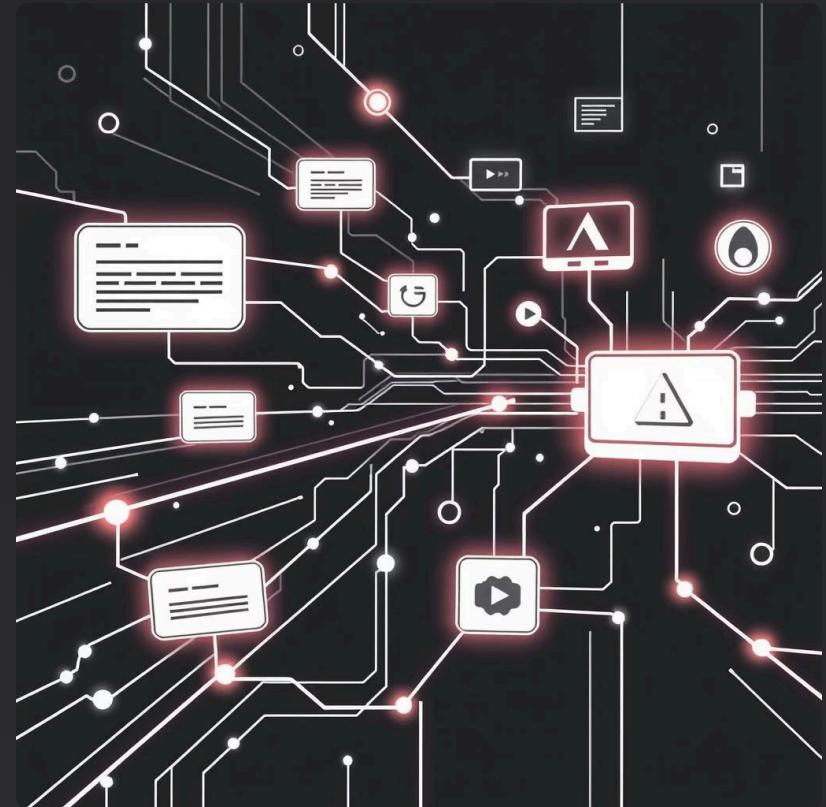
## Chatbot/Trợ lý ảo

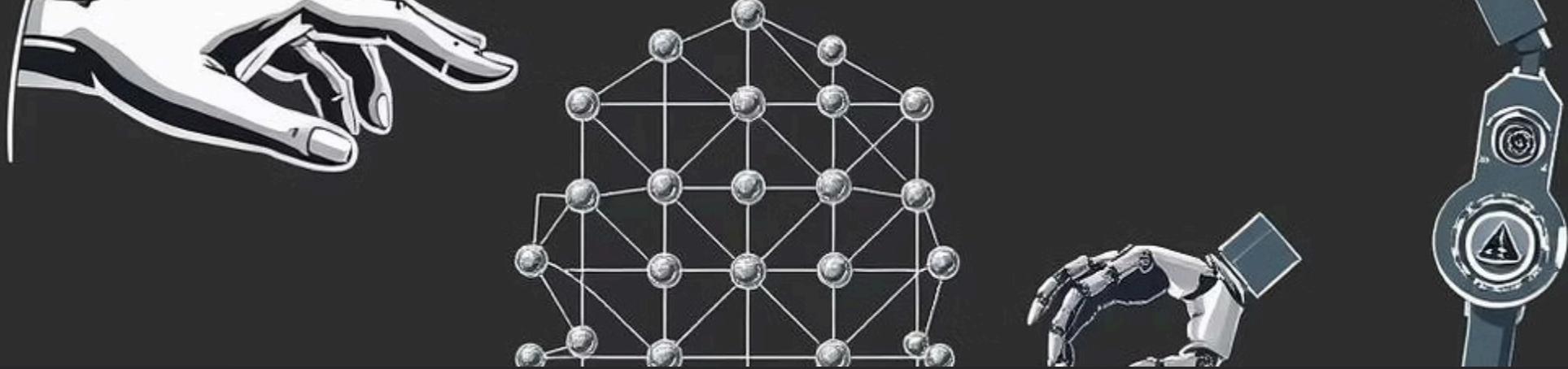
Siri, Google Assistant, ChatGPT.

# Tầm quan trọng của Dữ liệu trong AI

Dữ liệu là tập hợp thông tin đa dạng: chữ, số, hình ảnh, âm thanh. AI, đặc biệt là Machine Learning, học từ dữ liệu giống như con người học từ trải nghiệm. Không có dữ liệu, AI không thể học và mô hình trở nên vô nghĩa.

- Dữ liệu là "nhiên liệu" cho AI. Chất lượng dữ liệu quyết định chất lượng AI.





# Ba Loại Bài Toán Chính trong Machine Learning

1

## Học có giám sát (Supervised Learning)

Dữ liệu có nhãn. Mục tiêu: dự đoán giá trị/nhãn.

2

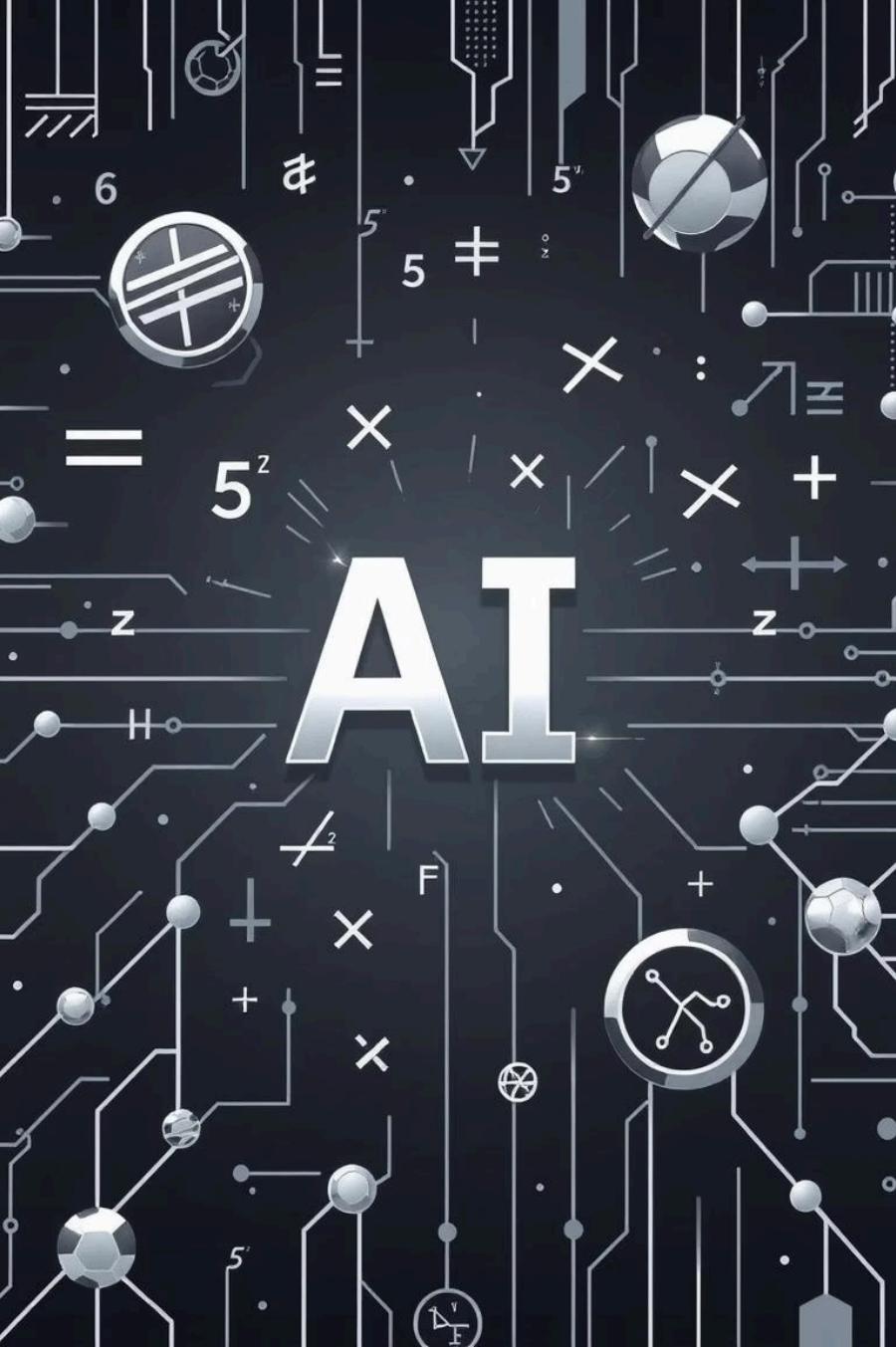
## Học không giám sát (Unsupervised Learning)

Dữ liệu không có nhãn. Mục tiêu: tìm cấu trúc ẩn, phân nhóm.

3

## Học tăng cường (Reinforcement Learning)

Agent tương tác môi trường, nhận thưởng/phạt, tối ưu hành vi.



# Thuật ngữ và Ký hiệu Cơ bản trong AI

## Thuật ngữ cơ bản

- Dataset: Bộ dữ liệu
- Feature: Thuộc tính
- Label: Nhãn
- Model: Mô hình
- Training: Huấn luyện
- Inference: Dự đoán
- Overfitting/Underfitting
- Loss function: Hàm mất mát
- Neural Network: Mạng nơ-ron

## Ký hiệu thường dùng

- $x$ : Input
- $y$ : Output
- $\hat{y}$ : Dự đoán
- $w$ : Trọng số
- $b$ : Độ chêch
- $f(x)$ : Hàm mô hình
- $L(y, \hat{y})$ : Hàm loss
- $\theta$ : Tập tham số
- $\nabla$ : Gradient



# Các Thư viện Python Phổ biến cho AI

## NumPy

Mảng n-dimensional, đại số tuyến tính tốc độ cao. Nền tảng cho nhiều thư viện khác.

## Pandas

Xử lý dữ liệu dạng bảng (DataFrame). Thao tác, làm sạch, phân tích dữ liệu mạnh mẽ.

## Matplotlib

Thư viện trực quan hóa dữ liệu, tạo biểu đồ 2D/3D.

## PyTorch

Thư viện ML/DL mạnh mẽ, xây dựng mô hình mạng nơ-ron. Ưu tiên nghiên cứu.

# Môi trường Phát triển và Nguồn Dữ liệu

## Môi trường Notebook

- **Jupyter Notebook:** Môi trường lập trình trong trình duyệt, kết hợp code và giải thích.
- **Google Colab:** Jupyter Notebook trực tuyến, miễn phí GPU/TPU, dễ chia sẻ.
- **Kaggle Notebook:** Môi trường online của Kaggle, tích hợp sẵn dữ liệu và thư viện ML.

## Nguồn Dữ liệu

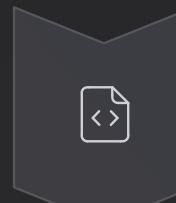
- **Kaggle:** Nền tảng lớn nhất chia sẻ dataset, thi đấu ML/AI, học máy.
- **Hugging Face:** Cộng đồng AI mã nguồn mở, tập trung NLP, mô hình ngôn ngữ, dataset.





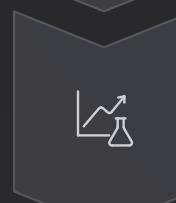
# Phân tích Dữ liệu Sơ bộ (EDA)

EDA là quá trình khám phá, trực quan hóa và tóm tắt dữ liệu để hiểu rõ, tìm quy luật, phát hiện vấn đề và đưa ra giả thuyết trước khi xây dựng mô hình.



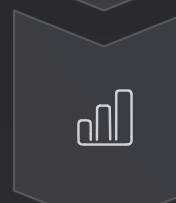
## Hiểu dữ liệu

Cấu trúc, giá trị thiểu, bất thường, ngữ cảnh.



## Phân tích dữ liệu

Thống kê mô tả, mối quan hệ biến, phân phối.



## Trực quan hóa

Biểu đồ đơn/hai/đa biến, phát hiện xu hướng.



# Trực quan hóa Dữ liệu trong EDA

## Các loại biểu đồ

- Đơn biến:** Histogram, Boxplot, Countplot.
- Hai biến:** Scatter plot, Bar chart, Line chart.
- Nhiều biến:** Heatmap, Pairplot, Bubble chart.

## Ý nghĩa

- Phát hiện xu hướng và quy luật ẩn.
- Tìm kiếm bất thường trong dữ liệu.
- Giải thích kết quả dễ hiểu hơn.
- Hỗ trợ ra quyết định hiệu quả.