MEMAHAMI ETIKA DAN BIAS UNTUK MEMBANGUN TRUSTWORTHY AI

Bambang Riyanto Trilaksono
Sekolah Teknik Elektro dan Informatika-ITB
Center for Artificial Intelligence-ITB
Riset.ai

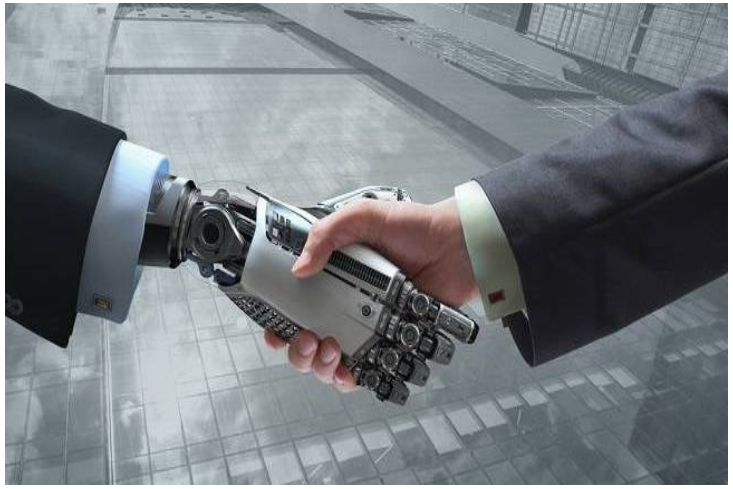
Human vs AI

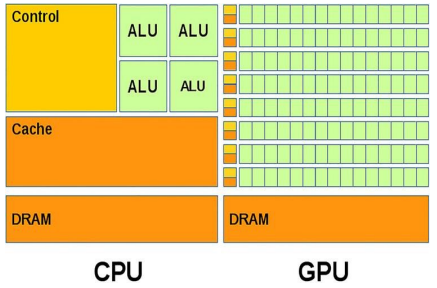
Human Intelligence

- Intuition, common sense, judgement, creativity, beliefs
- The ability to demonstrate their intelligence by communicating effectively
- Plausible reasoning, critical thinking

Artificial Intelligence

- Ability to simulate human behavior and cognitive processes
- Capture and preserve human expertise
- Fast response –the ability to comprehend large amounts of data quickly





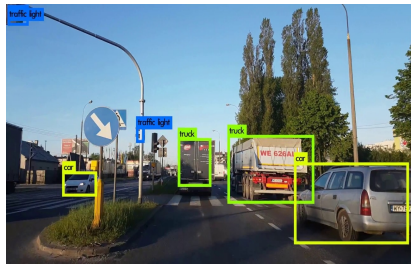
Algoritma makin baik

CPU & GPU makin kuat

Data makin berlimpah

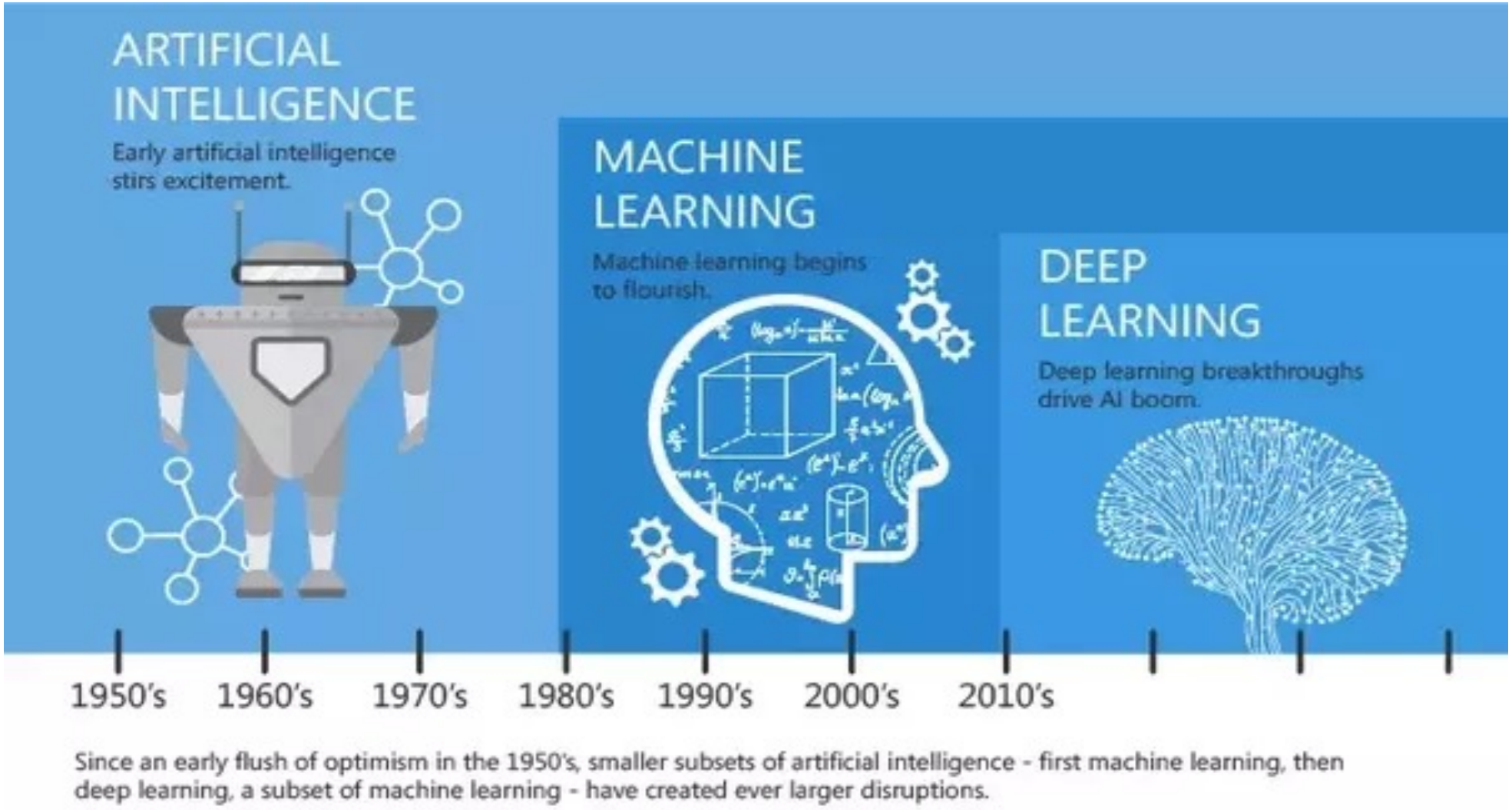
Perkembangan AI

Aplikasi makin beragam

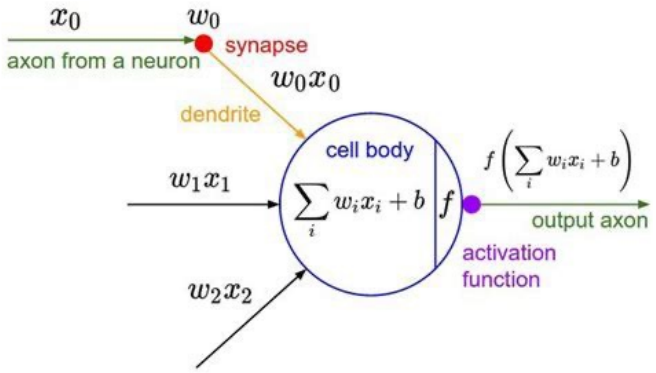


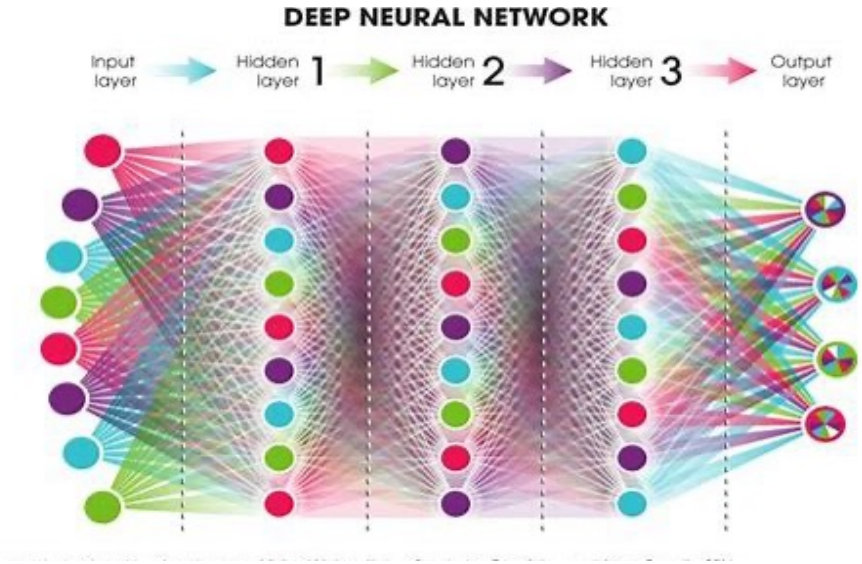
Mengapa Teknologi AI Semakin Berkembang?

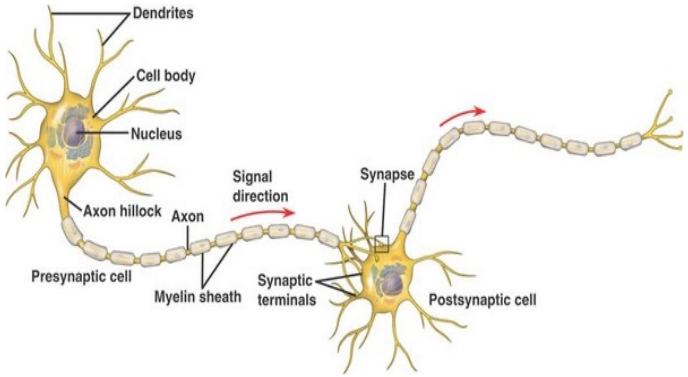
AI, Machine Learning, Deep Learning



Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.



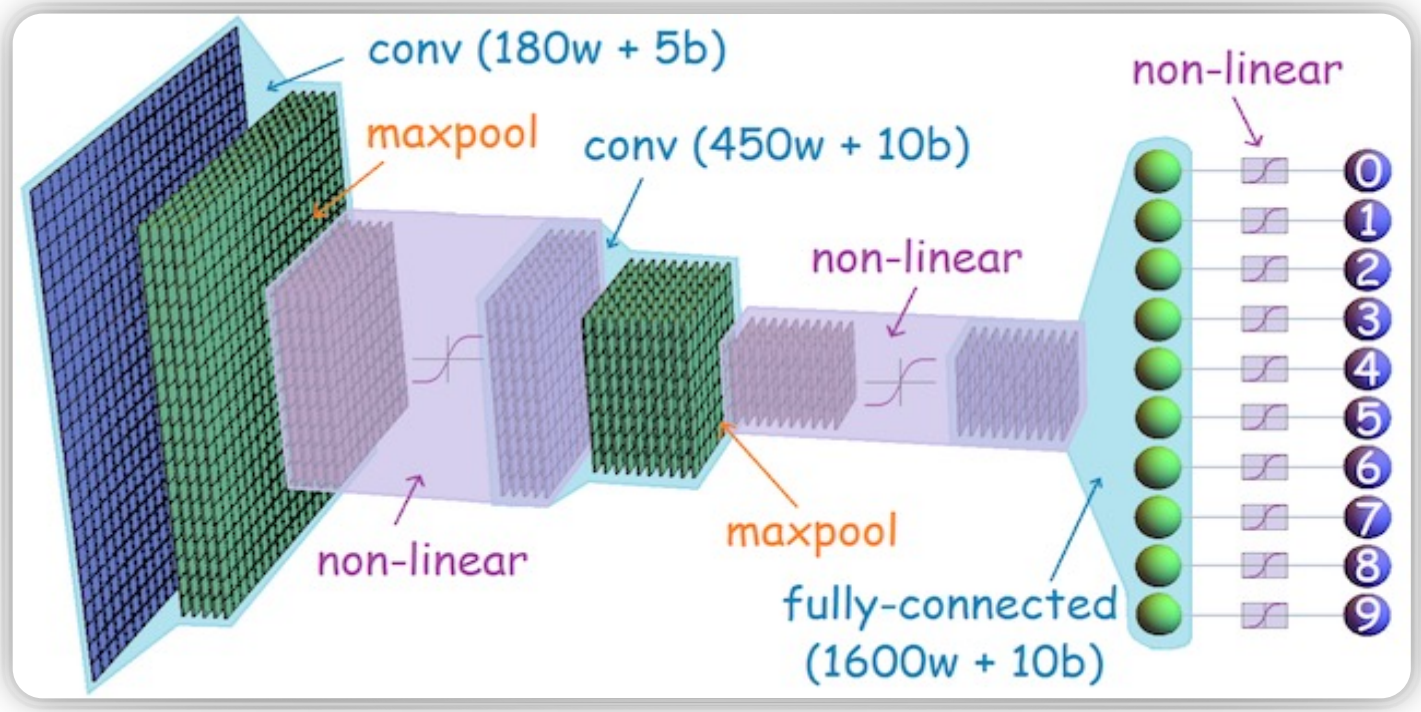




Neuron & Artificial Neural Networks Model

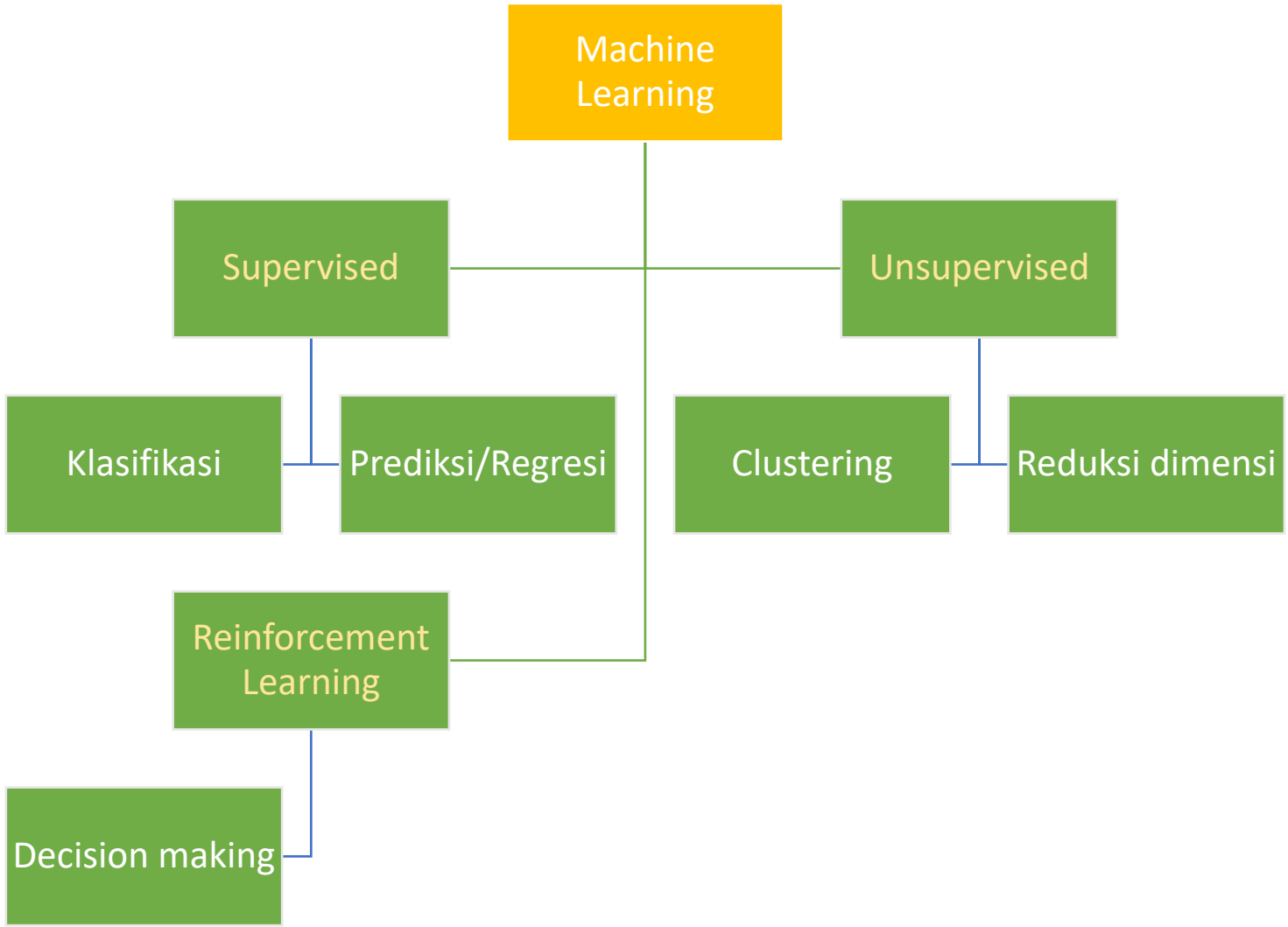
Convolutional Neural Networks

- Convolutional networks + FC networks
- Mengklasifikasi image
- Ekstraksi fitur secara otomatis
- Forward Computation & Backpropagation

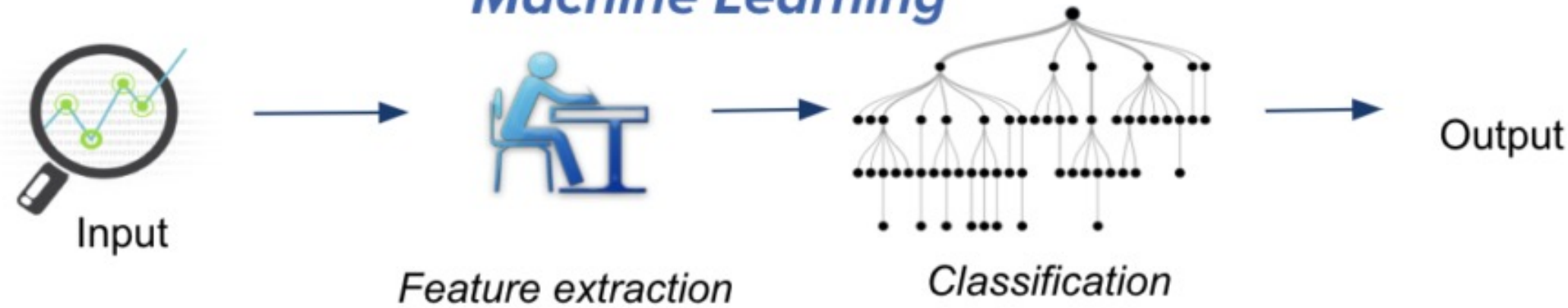


Deep Learning Architecture

Machine Learning & Deep Learning Task

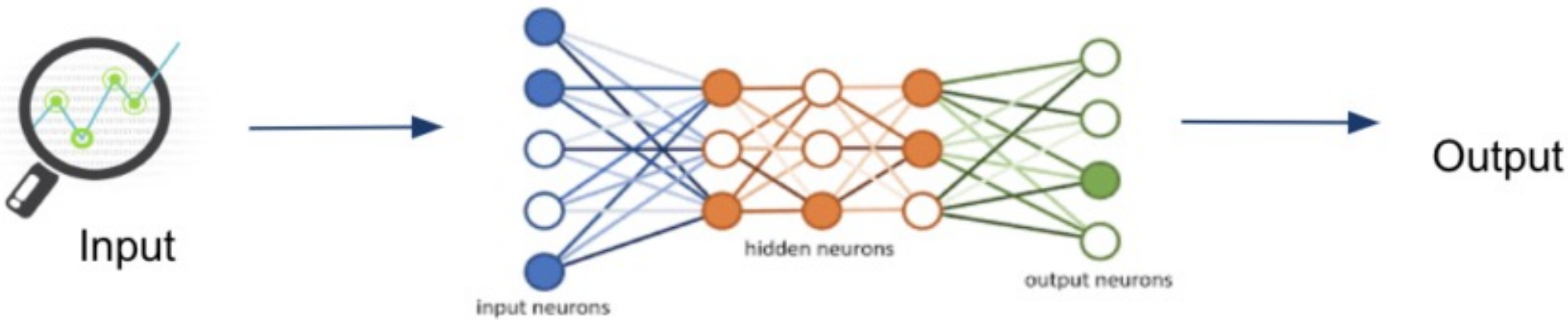


Machine Learning



Traditional machine learning uses hand-crafted features, which is tedious and costly to develop.

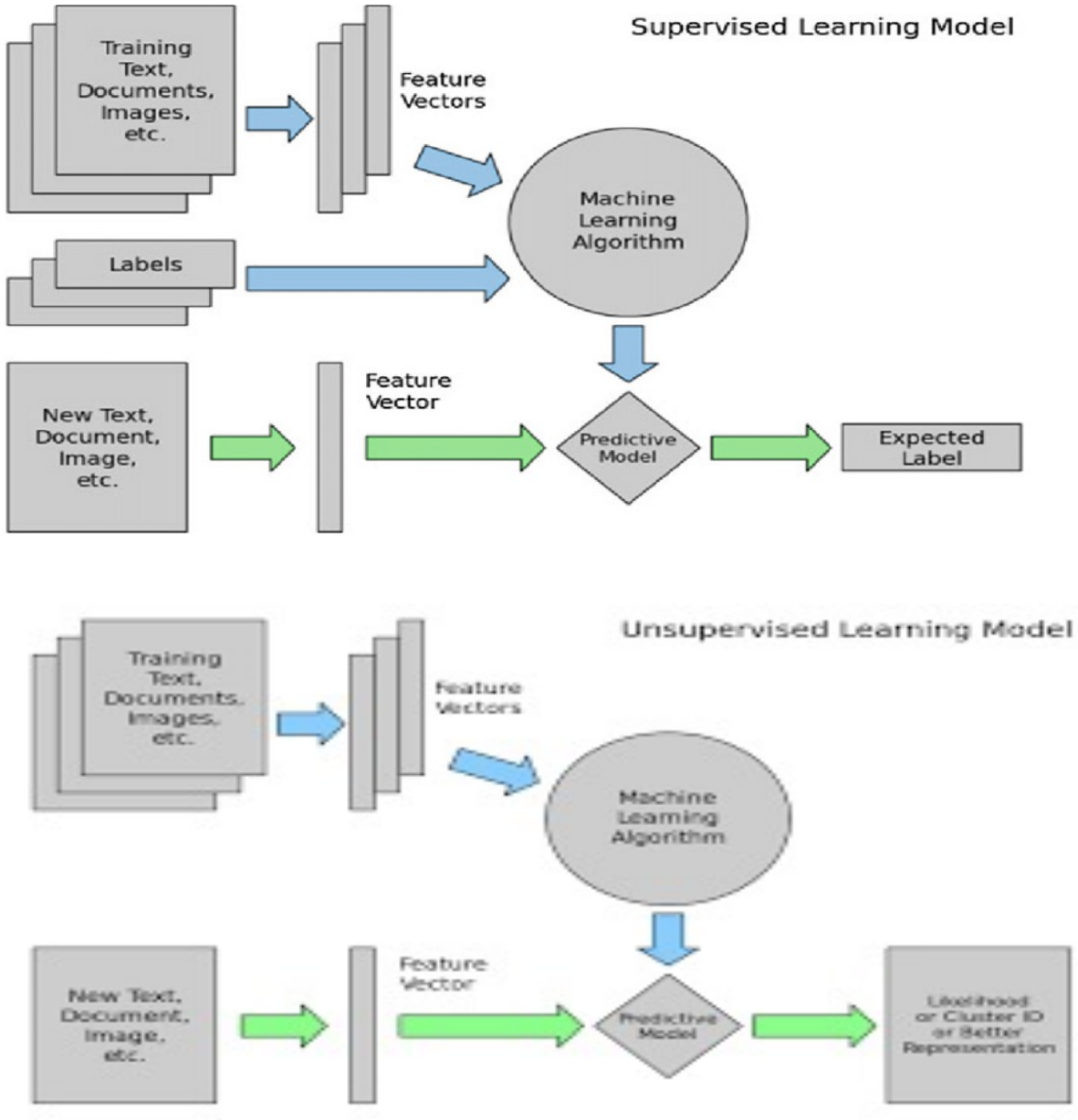
Deep Learning

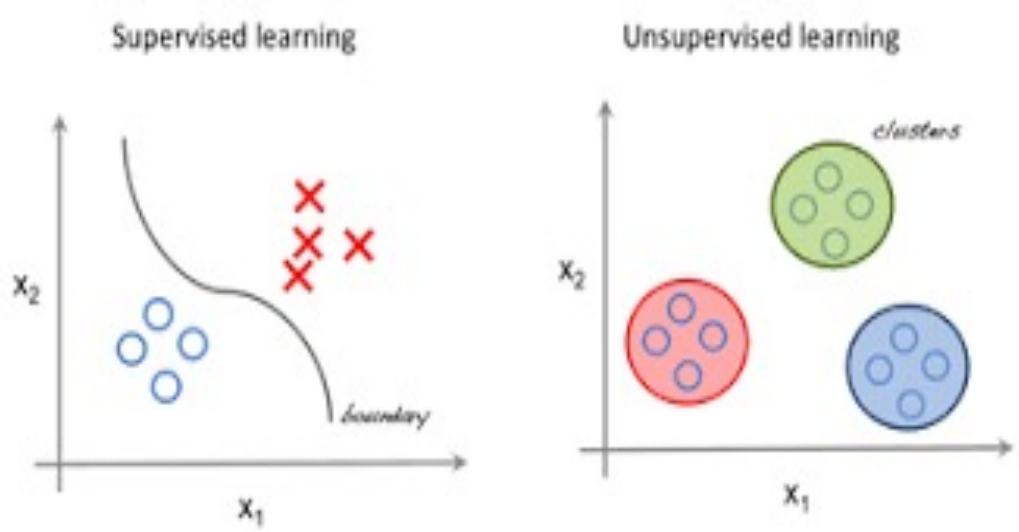


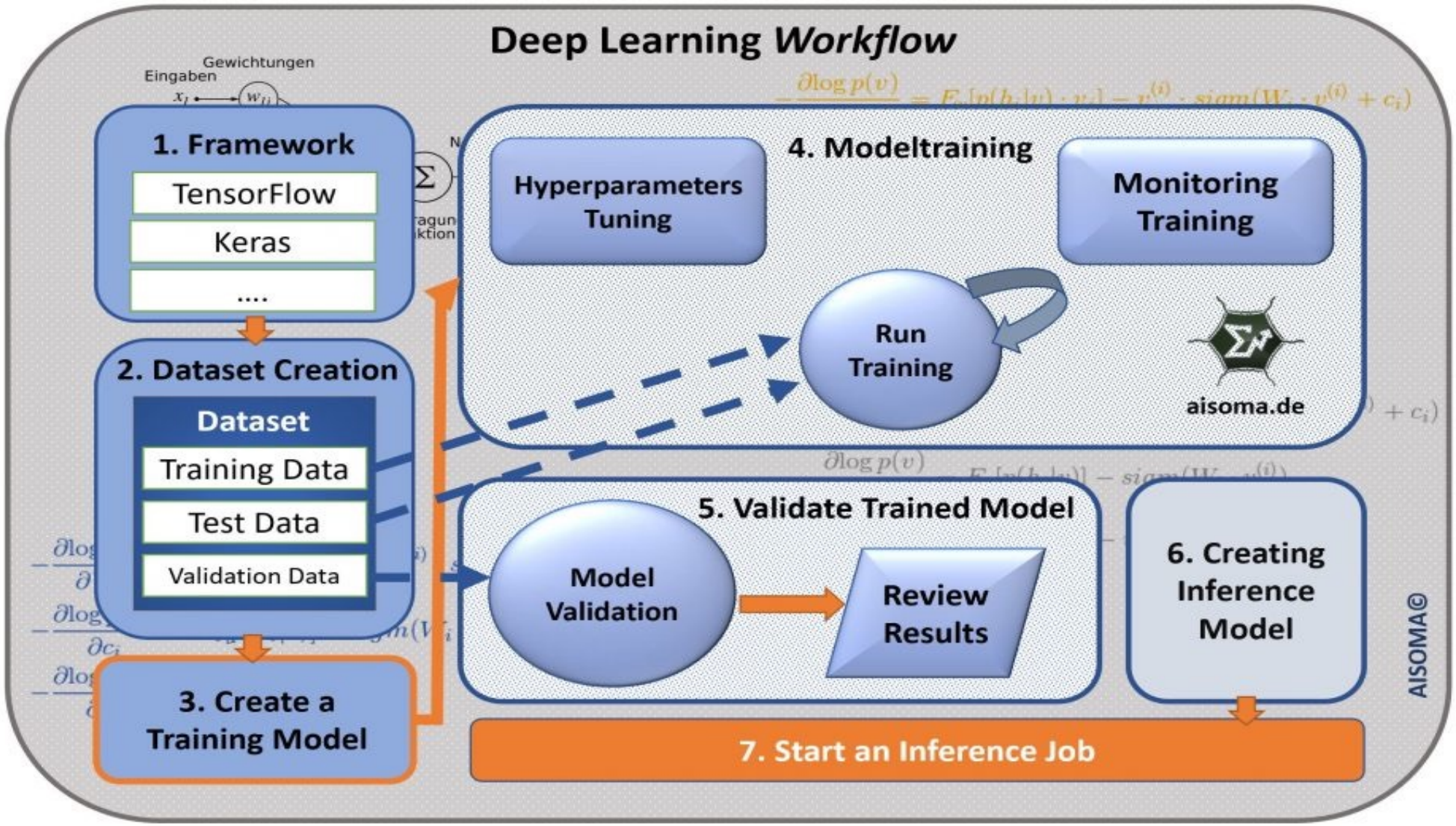
Neural Networks

Deep learning learns hierarchical representation from the data itself, and scales with more data.

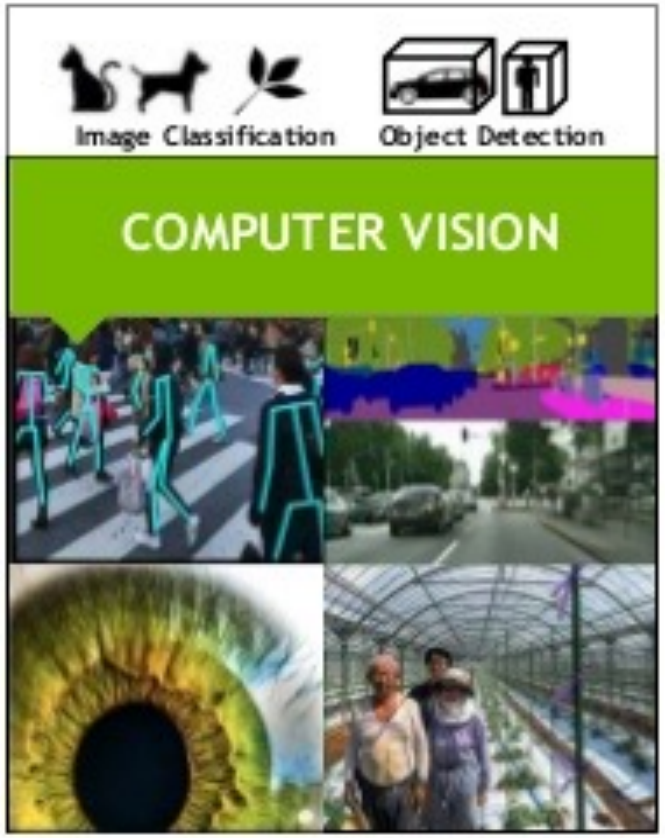
Machine Learning Workflow

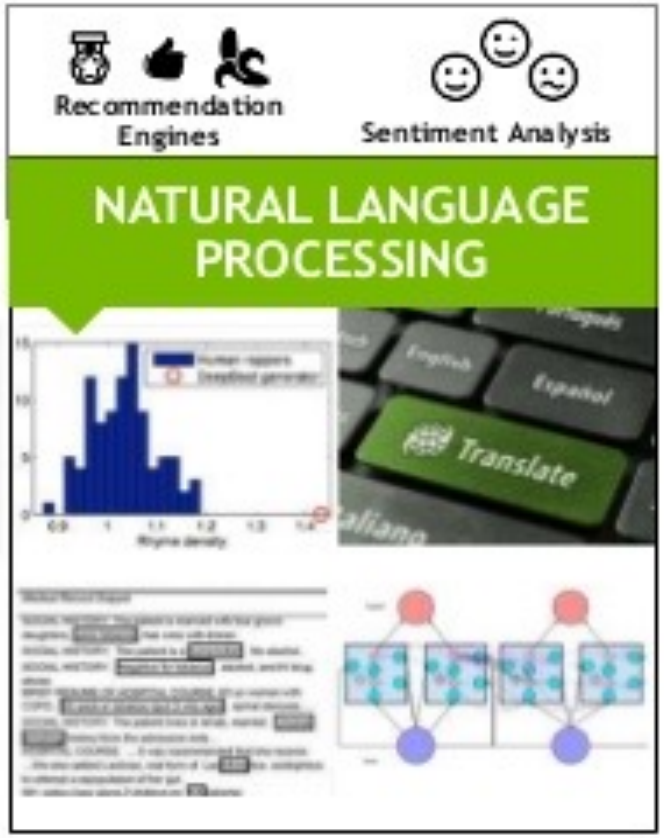





Deep Learning *Workflow*
Gewichtungen
Eingaben
1. Framework
TensorFlow
Keras
....
2. Dataset Creation
Dataset
Training Data
Test Data
Validation Data
3. Create a Training Model
4. Modeltraining
Hyperparameters Tuning
Monitoring Training
Run Training
aisoma.de
5. Validate Trained Model
Model Validation
Review Results
6. Creating Inference Model
7. Start an Inference Job
AISOMA©

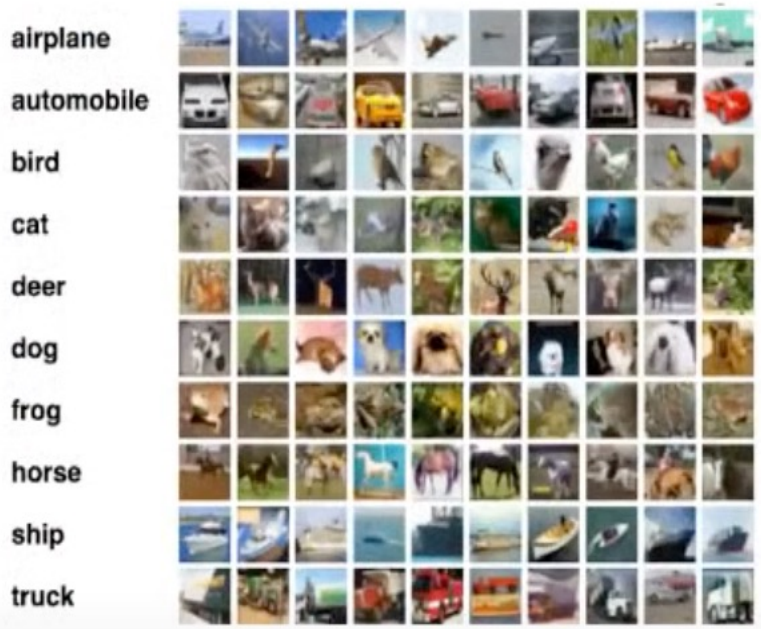
AI APPLICATIONS

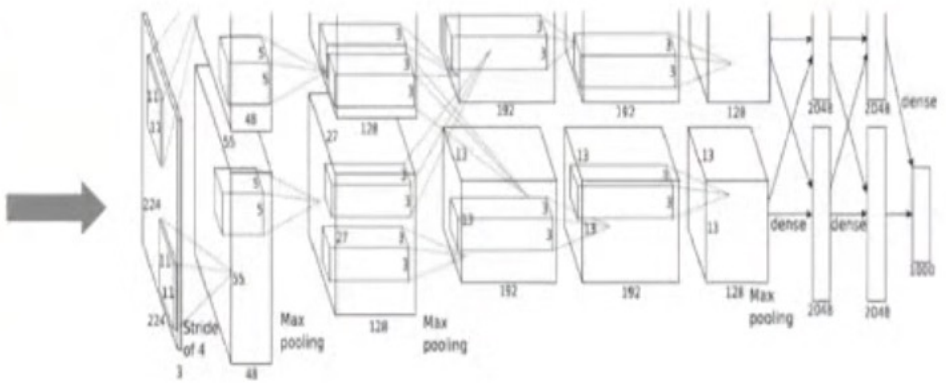


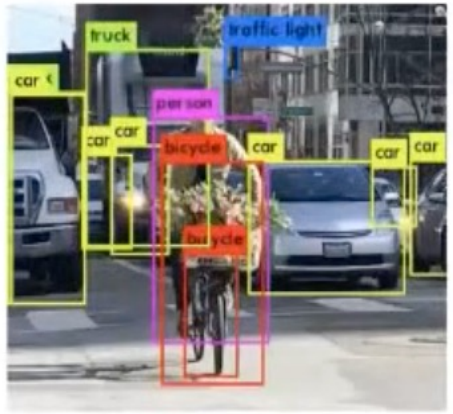


Big Data Analytics

Supervised (Deep) Learning

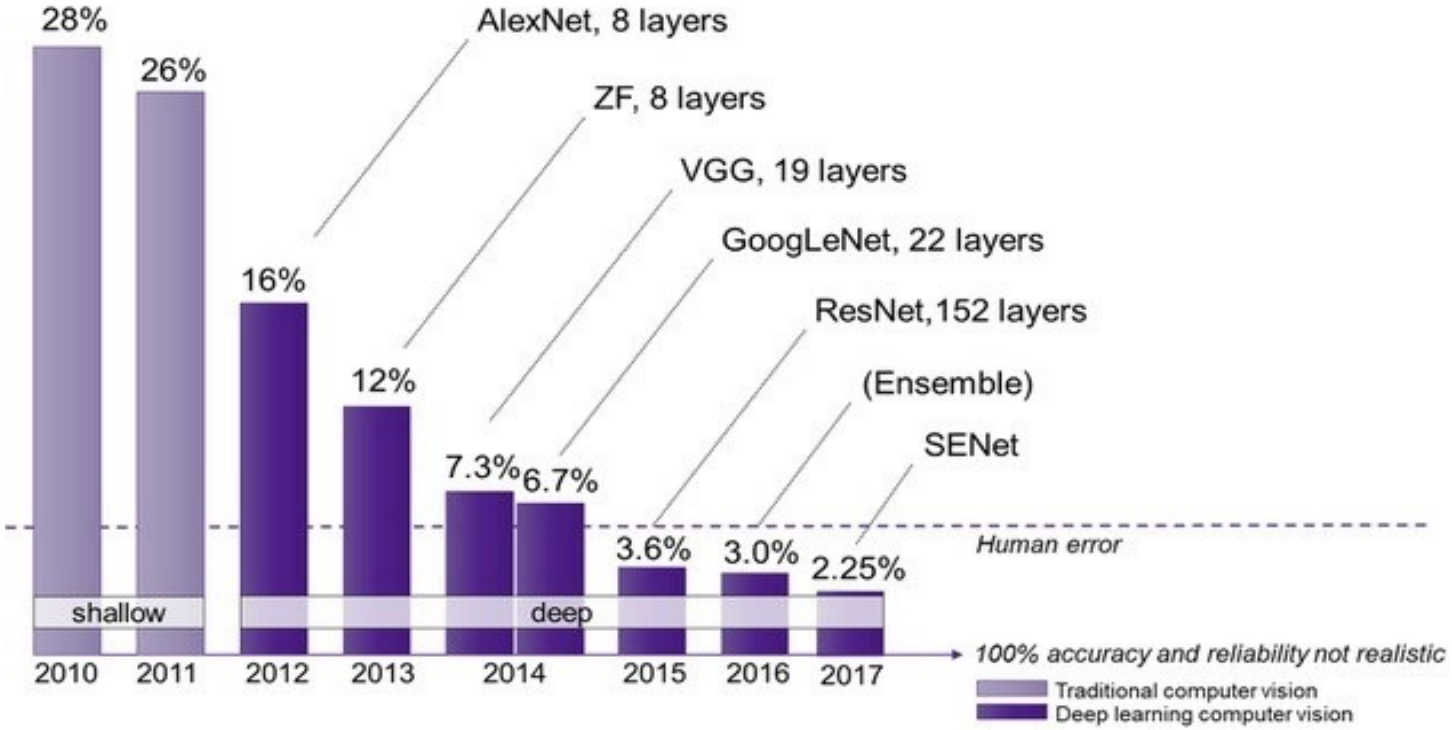




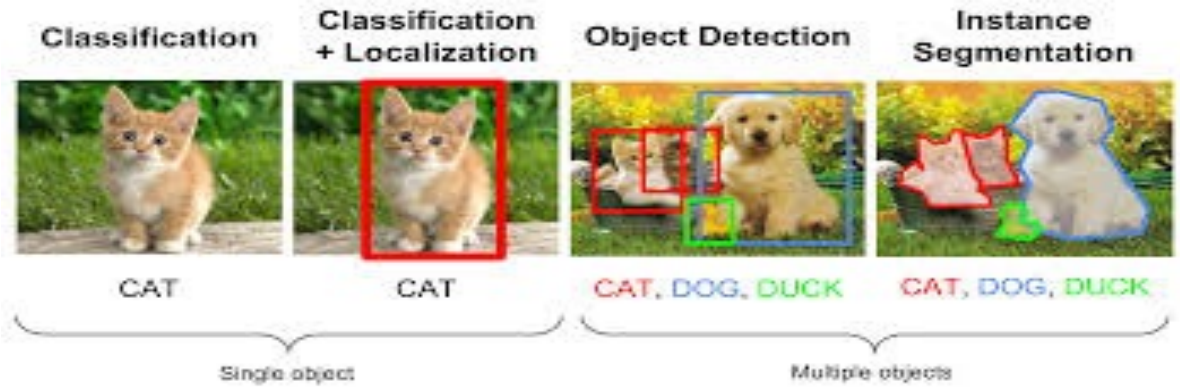


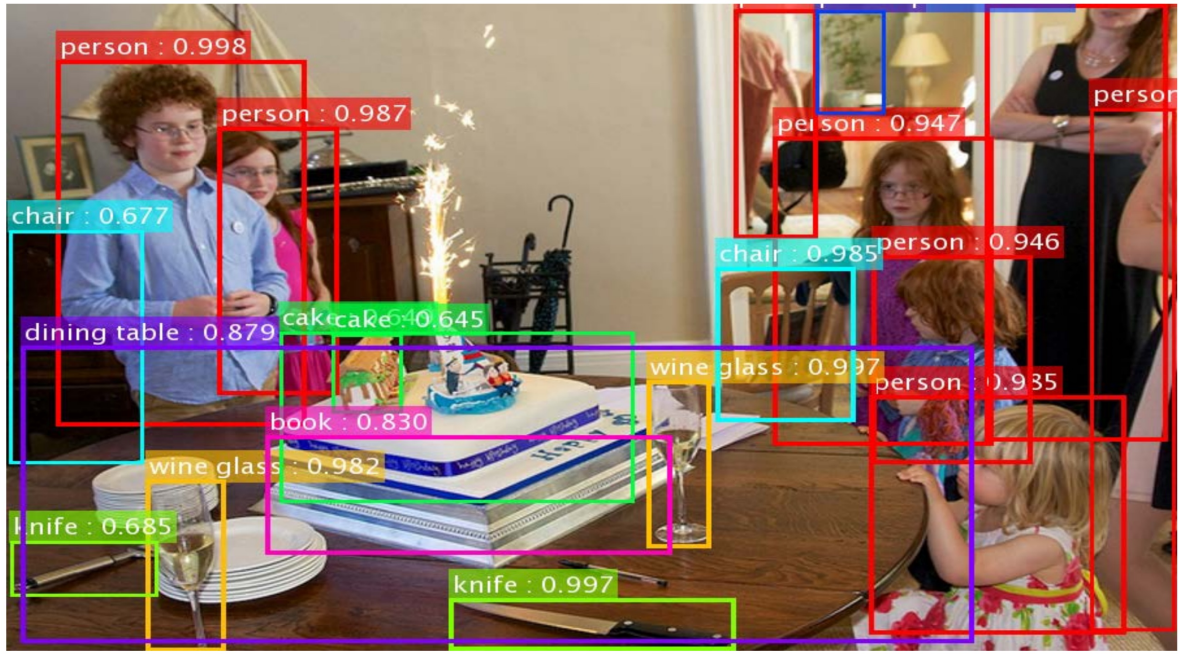
- Predefine the set of visual concepts to be learned
- Collect diverse and large number of examples for each of them
- Train a deep model for several GPU hours / days

Kompetisi ImageNet



Aplikasi AI untuk Computer Vision





AI & Etika

Bias	• Is AI fair?
Liability	• Who is responsible for AI?
Human interaction	• Will we stop talking to one another?
Employment	• Is AI getting rid of jobs?
Wealth	• Who benefits from AI?
Power & control	• Who decides how to deploy AI?
Robot rights	• Can AI suffer?

Contoh Bias pada AI

Amazon recruiting tool

- Aplikasi rekrutmen yg mereview resume applicant dan memberikan rating
- Bias thd gender (wanita)

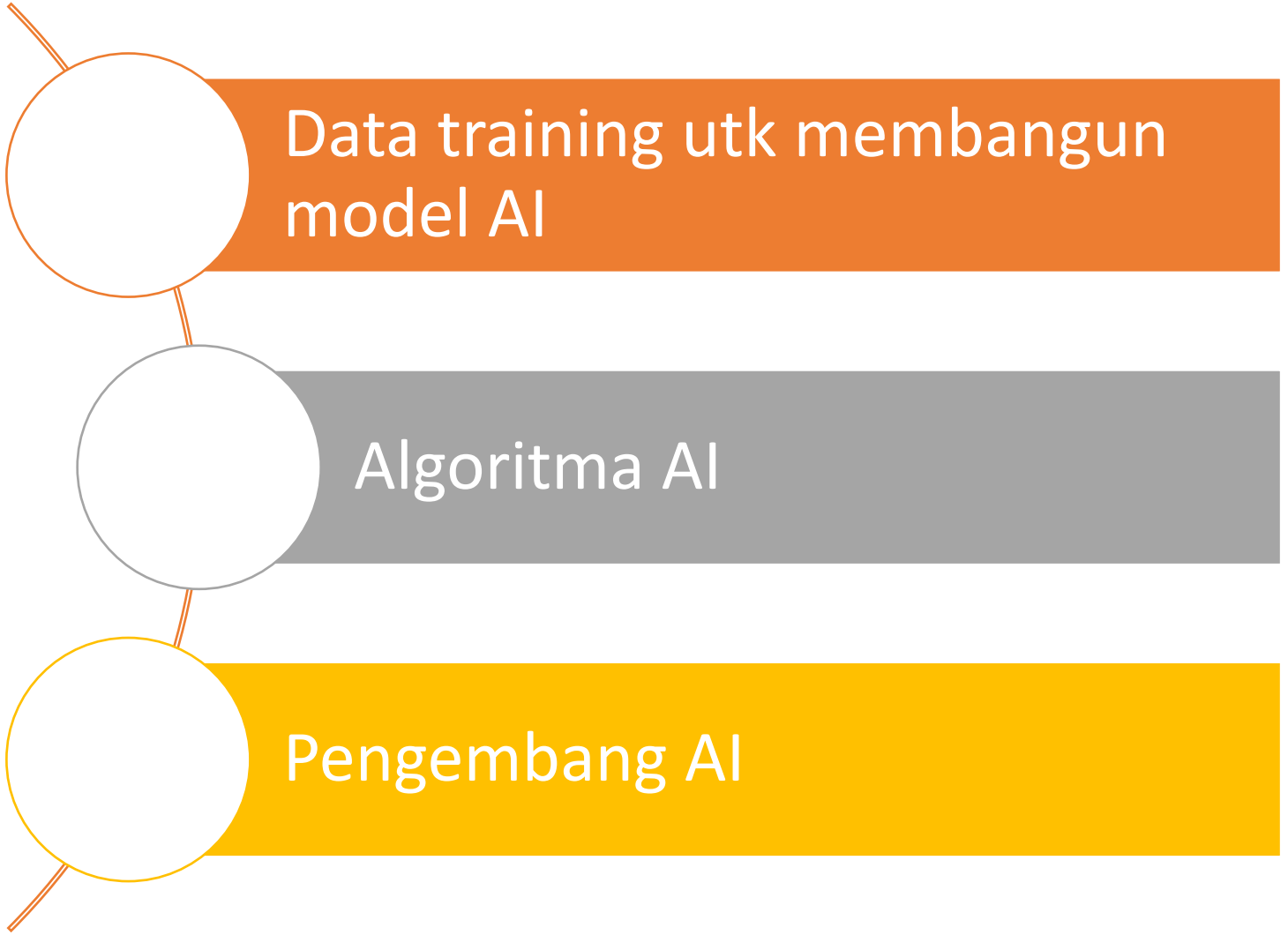
Healthcare risk prediction

- Memprediksi pasien yg memerlukan layanan kesehatan ekstra
- Memprioritaskan pasien kulit putih dibandingkan kulit hitam

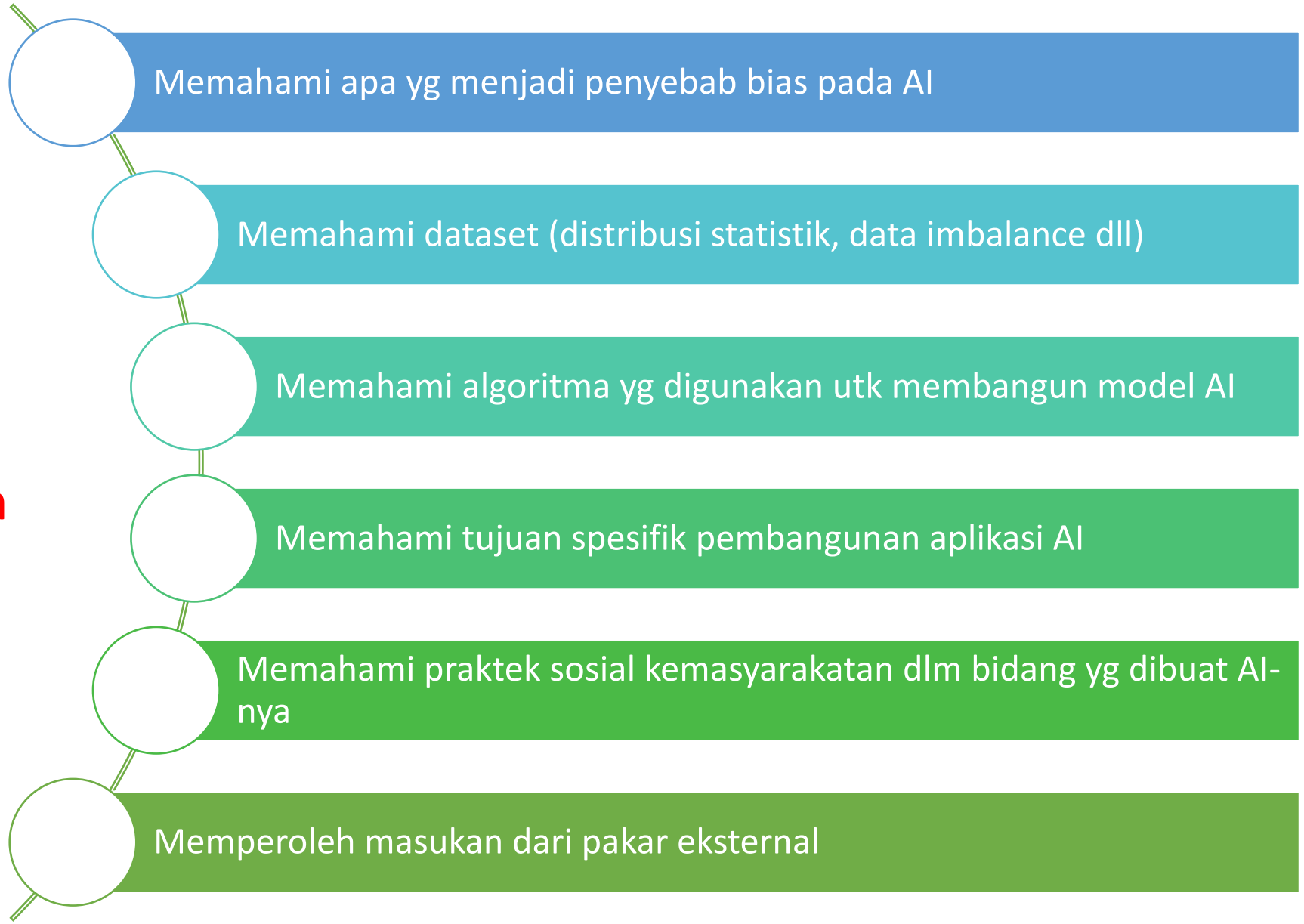
Face recognition

- Temuan ACLU & MIT : akurat utk pria kulit putih, tdk akurat pada wanita kulit hitam

Apa yg menjadi penyebab terjadinya bias pada AI?



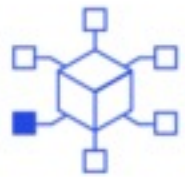
Bagaimana cara mengatasi atau memecahkan masalah bias pada AI?



Minimizing bias will be critical if artificial intelligence is to reach its potential and increase people's trust in the systems.

Six potential ways forward for artificial-intelligence (AI) practitioners and business and policy leaders to consider

1



Be aware of contexts in which AI can help correct for bias and those in which there is high risk for AI to exacerbate bias

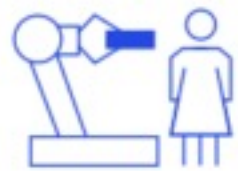
2

Establish processes and practices to test for and mitigate bias in AI systems

3

Engage in fact-based conversations about potential biases in human decisions

4



Fully explore how humans and machines can best work together

5

Invest more in bias research, make more data available for research (while respecting privacy), and adopt a multidisciplinary approach

6

Invest more in diversifying the AI field itself

McKinsey & Company

Tools utk Mengurangi Bias

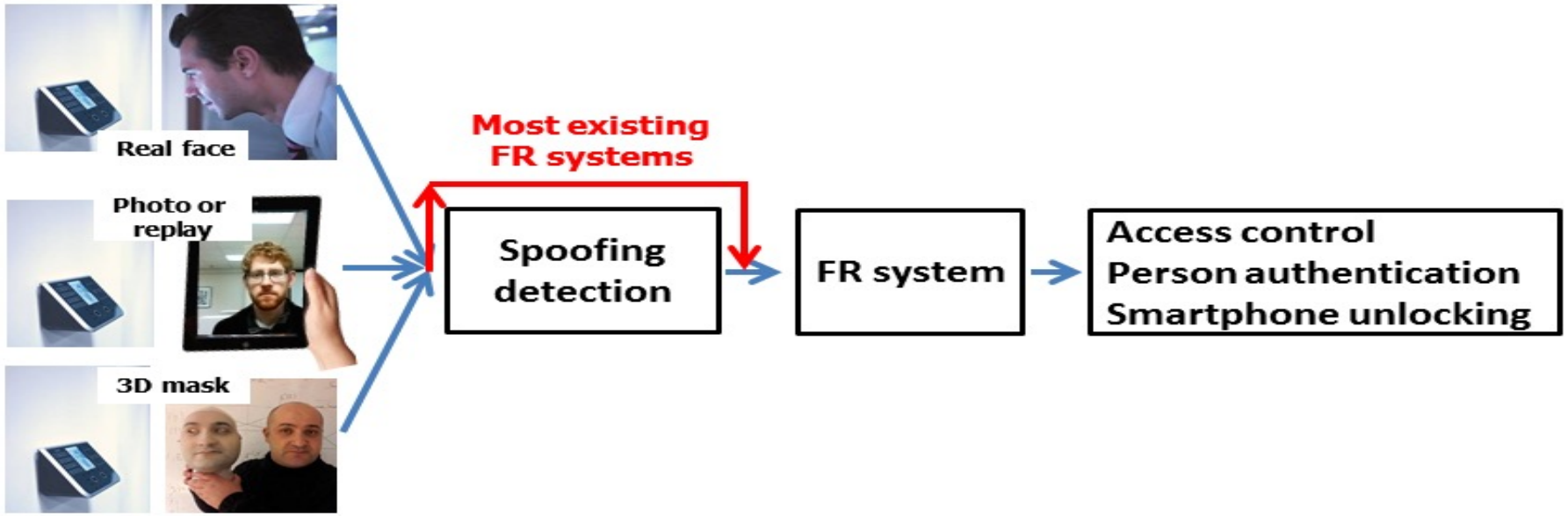
AI Fairness 360 (IBM)	• Menguji bias pada data dan model • Terbatas utk klasifikasi biner
IBM Watson OpenScale	• Memeriksa bias secara real-time ketika AI membuat prediksi
Google What-If Tool	• Menguji performance • Menganalisis atribut data yg berbeda • Visualisasi

Trustworthy AI

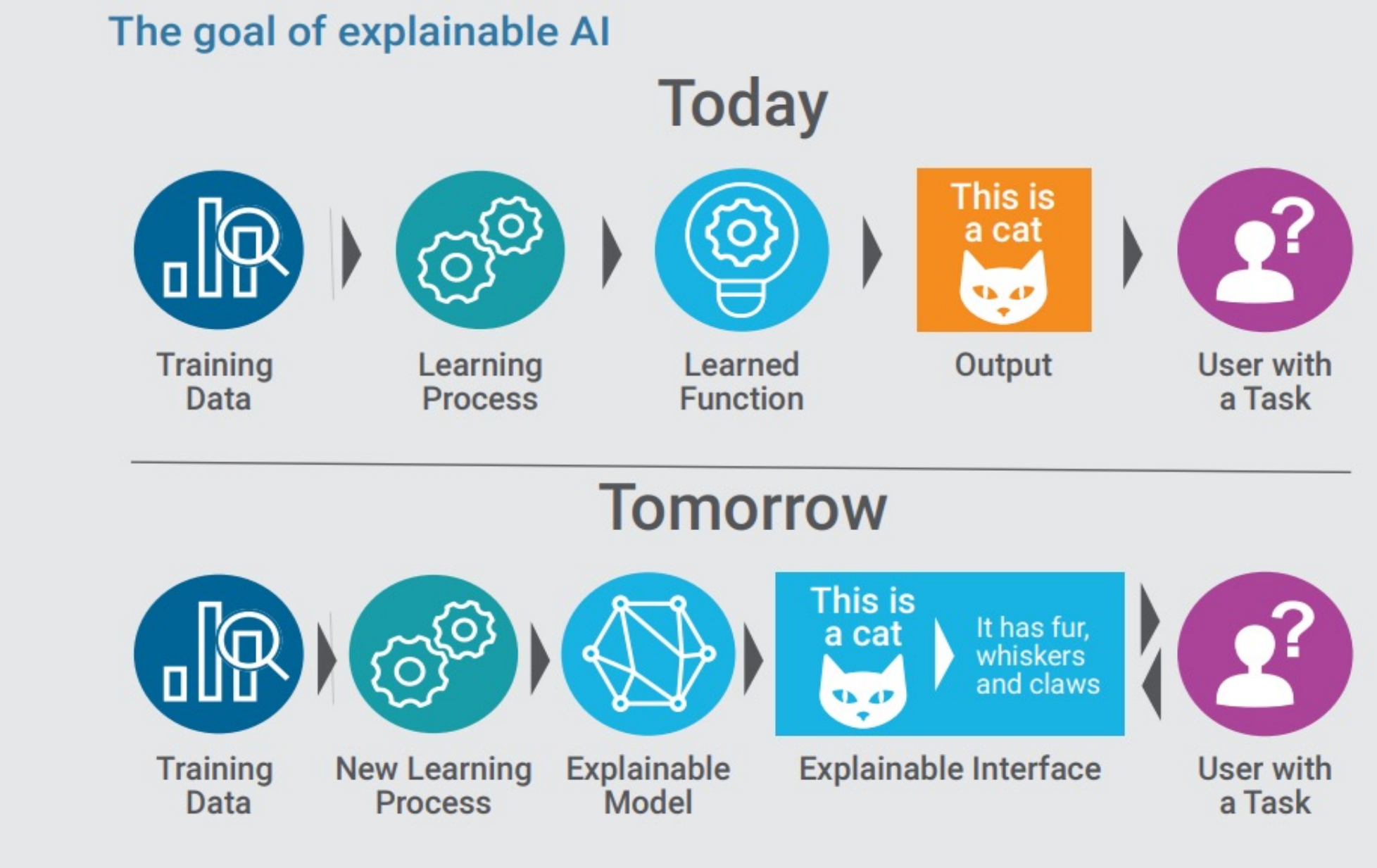
- **Robust** : tahan thd serangan (attack)
- **Fair** : berkeadilan (tidak bias)
- **Responsible** : Akuntabel
- **Transparent** : dapat menjelaskan
- **Secure** : aman
- **Privacy : menjaga privasi**

Spoofing--serangan pada pengenalan wajah

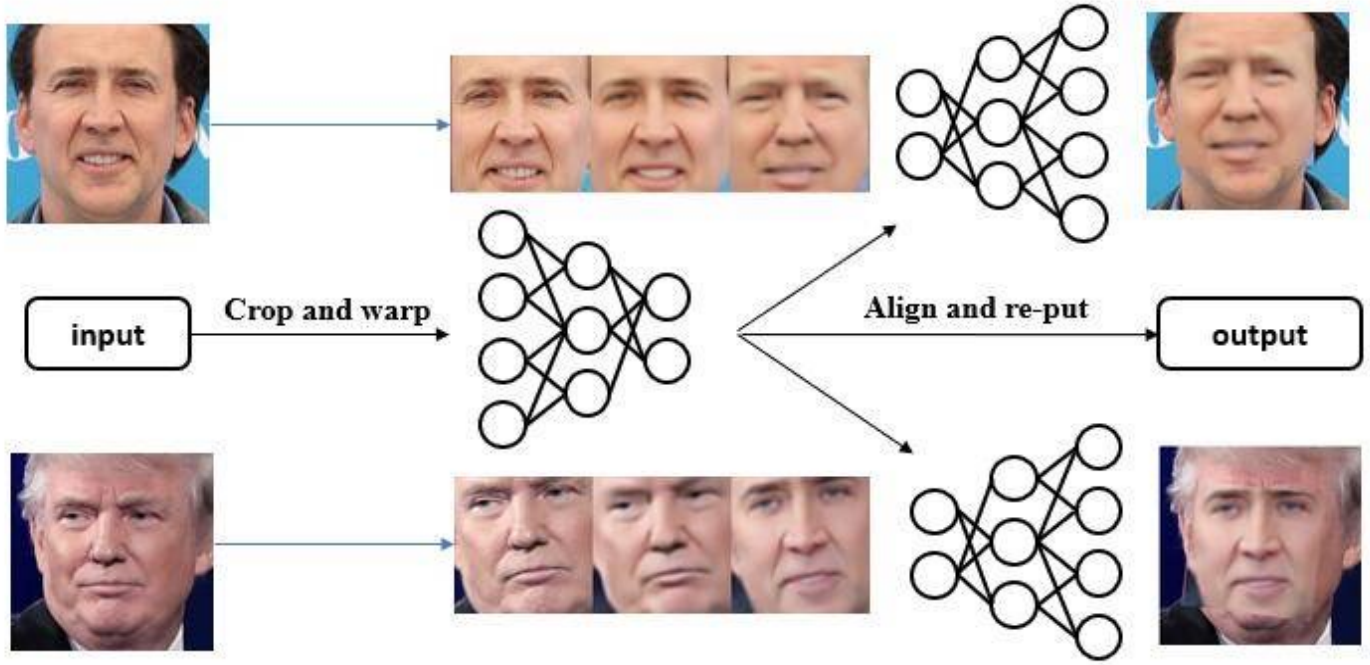
- Deteksi Liveness
- Aktif vs pasif



Explain-able AI

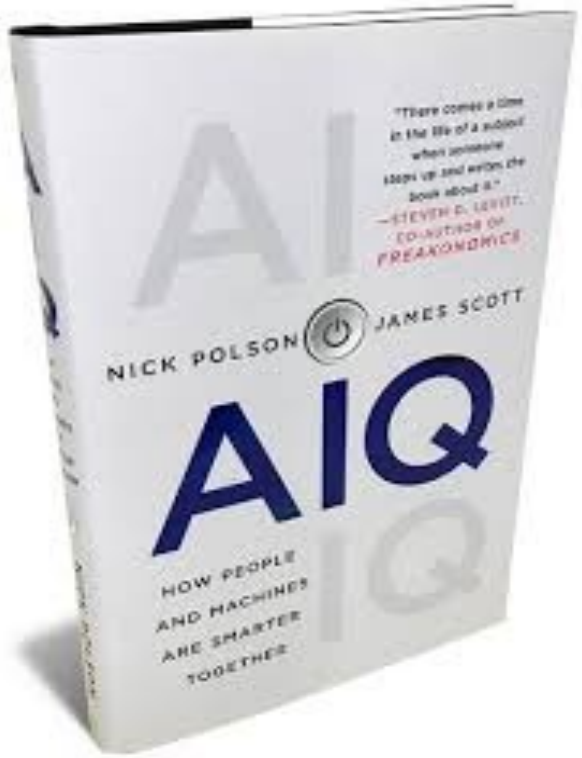


Deepfake



- Memiliki kemampuan utk membuat foto dan video (digabungkan dg suara) palsu yg seolah-olah benar/valid
- AI yg bertanggung-jawab?
- Microsoft Deepfake Detection Tool

Masih diperlukan riset yg mendalam, komprehensif dan multidisiplin untuk memahami etika AI, mengatasi bias, dan membangun Trustworthy AI



A revolution of intelligent machines, from self driving cars to smart digital assistants, is now remaking our world, just as the Industrial Revolution remade the world of the nineteenth century.

AI is not some science fiction droid from the future. It's right here, right now, and it's changing our lives at lightning-fast speed.

Artificial Intelligence Quotient : How People and Machines Are Smarter Together (Nick Polson & James Scott, St. Martins Press, 2018)

TERIMAKASIH

riset.ai