Von Straßenerkennung bis Personen-detektion – Die spannende Welt der Bildverarbeitung

Sandra Ebert, Bosch Engineering GmbH, 15.06.2023

BOSCH

Females in Tech
Rules for Training

Please mute yourself (if you are not talking ☺)

Please unmute yourself and talk to me.

Chat: You can also post your question/comment in the chat.

BOSCH

Females in Tech
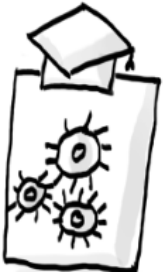## Outline

Introduction

Computer Vision

Natural Language Processing

Robotics

BOSCH

1

Introduction

BOSCH

Females in Tech - Introduction
What is AI?

Artificial Intelligence (AI)
- Build machines with
  - Intelligent behavior
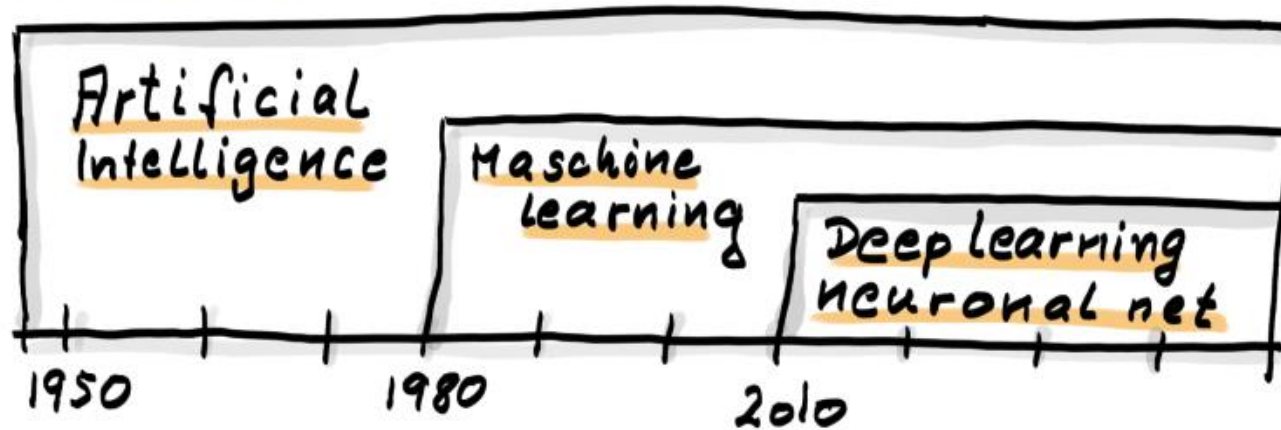 - Certain human decision-making structures

Machine Learning (ML)
- Ability to automatically learn and improve from experience
- Parse data → learn from it → prediction/decision
- Structured data

Deep Learning (DL), Neural Networks
- Closest to human learning
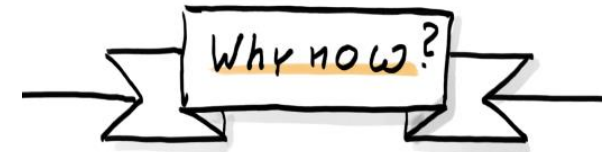- large amounts of (unstructured) data
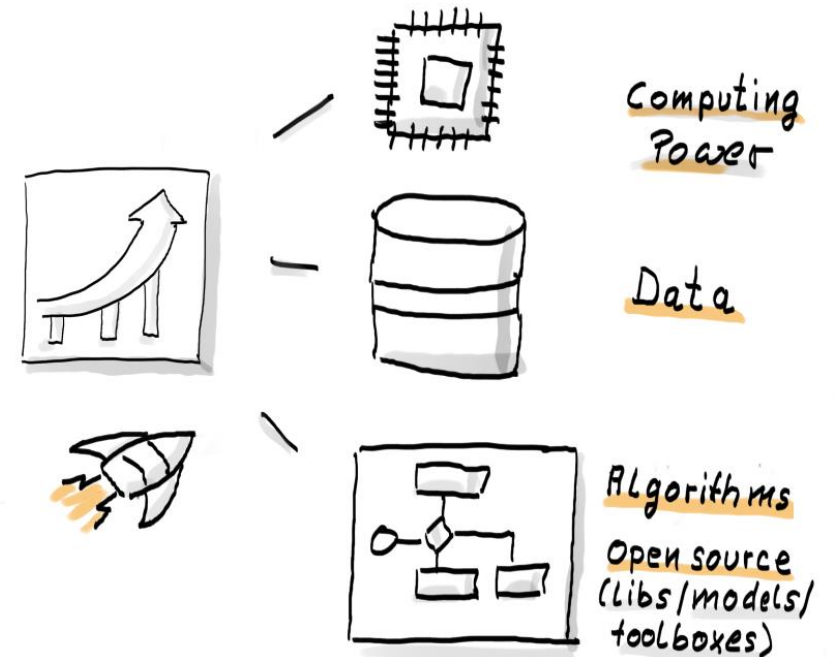
BOSCH

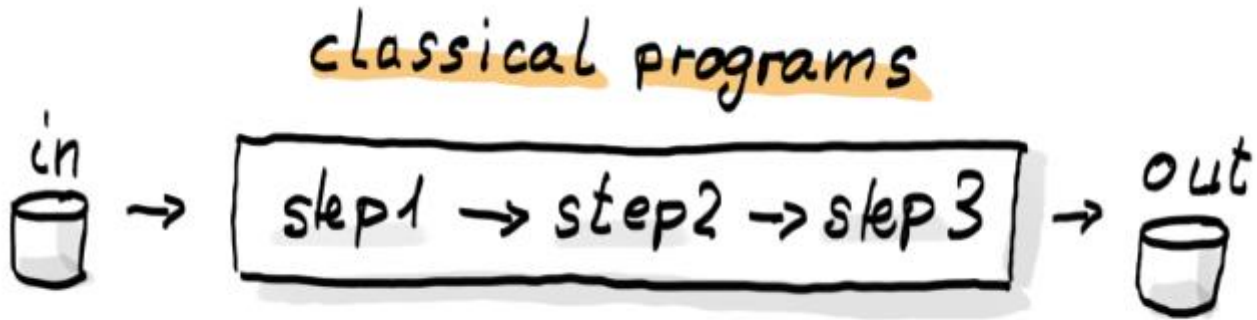# Females in Tech - Introduction
Why AI and why now?

- Why AI?
  - large potential for added value:
    - product and service innovations,
    - development of new business models

BOSCH

Example: a + b = c

BOSCH

Images licensed [CC0 1.0 Universal](#)
J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel. The German Traffic Sign Recognition
Benchmark: A multi-class classification competition. In *Proceedings of the IEEE*
International Joint Conference on Neural Networks, pages 1453–1460. 2011.

BOSCH

Training Data

Real Data

BOSCH

Machine Learning

supervised learning
unsupervised learning
reinforcement learning
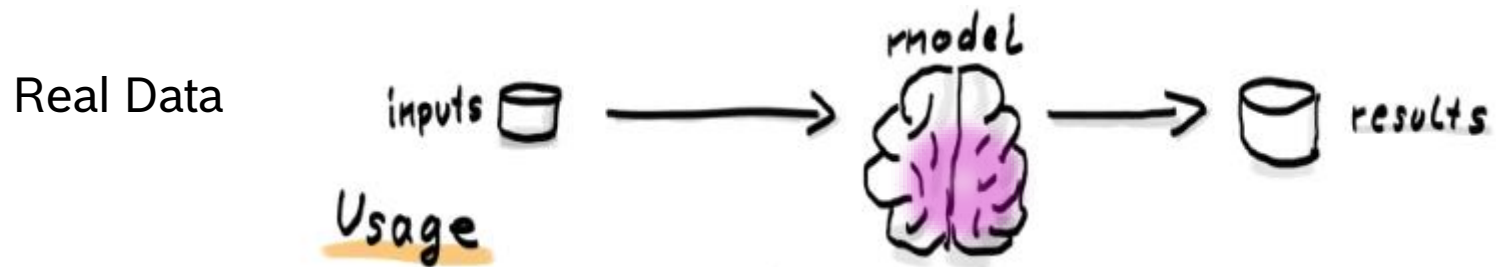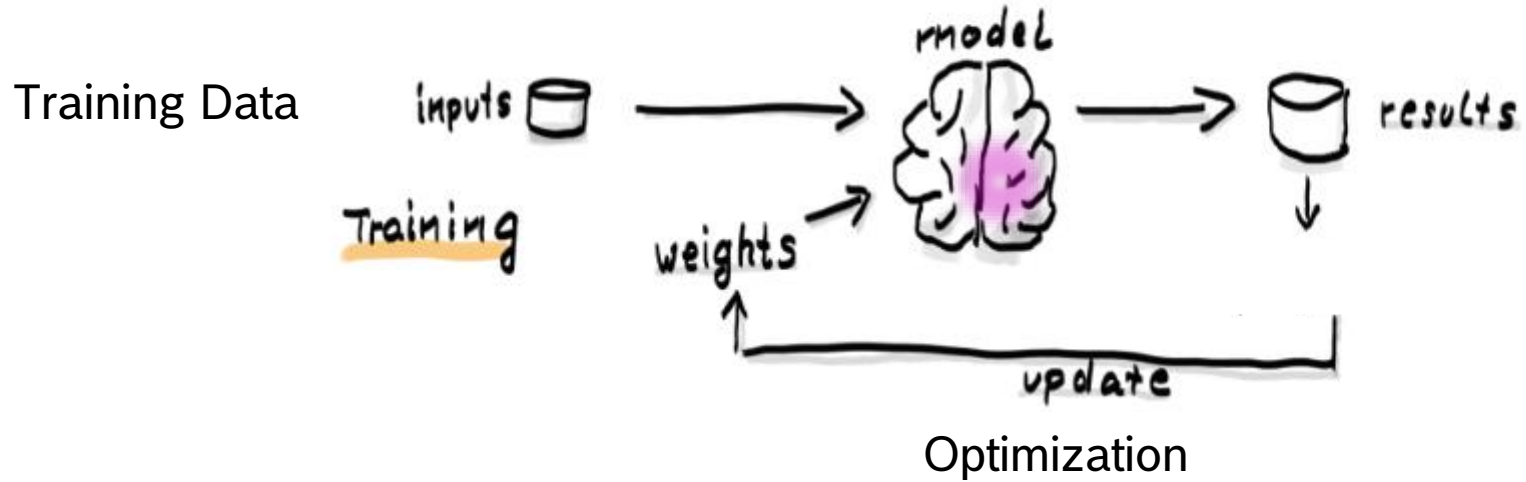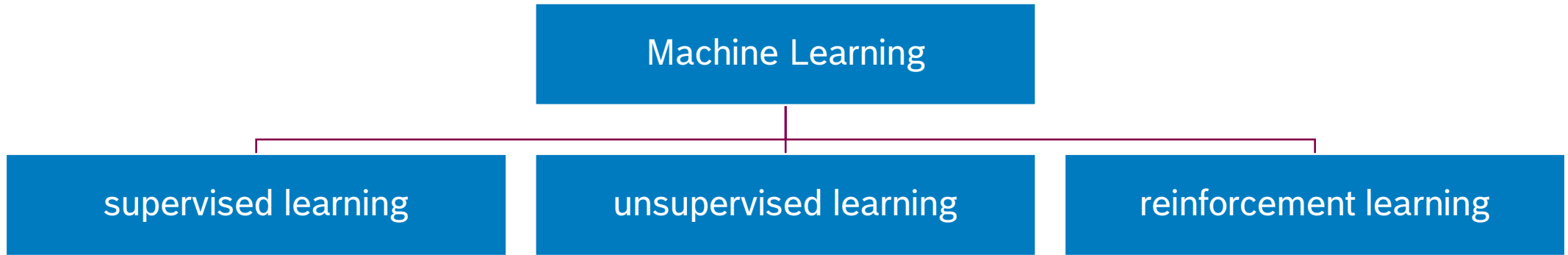
**learn from
labeled data**

 → "cat"

 → "dog"

**cluster
unlabeled data**

**learn from
reward**

BOSCH

Females in Tech - Introduction
Typical Machine Learning (Data Analytics) Process

Data Acquisition

Data Preprocessing
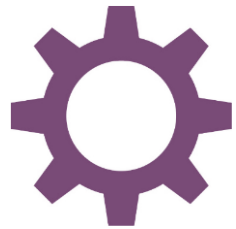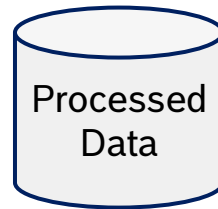
Data Processing
"Machine Learning"

Deployment

Raw Data

Raw Data

Raw Data

Processed Data

Filtering

Cleaning

Import

Extraction

Transforming
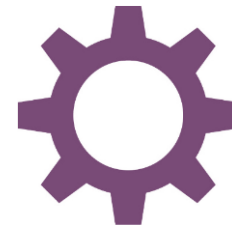
Modelling

Training

Validation

Migration

Reporting

Up to 80% of total effort

BOSCH

Females in Tech - Introduction
Training effort

large NLP model (GPT-3)
produce human-like text

355 GPU-years training time
estimated cost: 4.6 m. $
(one training run at cheapest cloud provider)

GP model for regression
e.g., predict fuel consumption

typical vision models at Bosch
e.g., traffic sign recognition

training effort

minutes   hours   days   weeks   months   years   decades   centuries
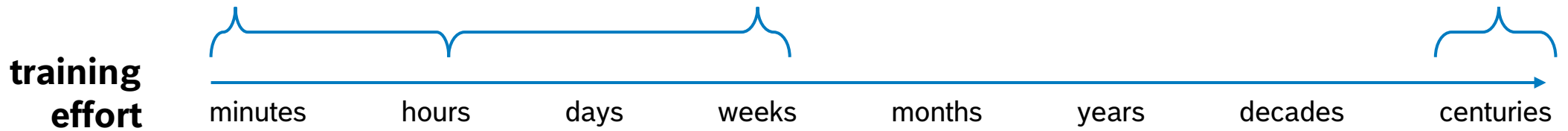
Hardware at Bosch

Notebooks & Workstations

CI GPU cluster
CI HPC cluster (CPU)

cloud services
(Azure or AWS)

GP: Gaussian Process
NLP: Natural Language Processing
CPU: Central Processing Unit

GPU: Graphics Processing Unit
GPT-3: Generative Pre-trained Transformer 3

HPC: High Performance Computing
AWS: Amazon Web Services

BOSCH

2

Computer Vision

BOSCH

Females in Tech – Computer Vision
Low-level Vision

| Images | Processing | Output |
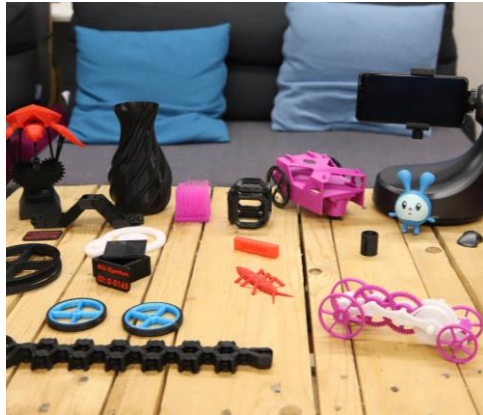
- Pixel-/patch- or line-based processing
- E.g. replace black image points of the cup with the surrounding image

BOSCH

Females in Tech − Computer Vision
Low-level Vision

```
Images  ➤  Processing  ➤  Output
```

- Equivalent to "Industrielle Bildverarbeitung"
- No information about environment, illumination, scene understanding

- **Further examples:**
- Tumor recognition in Medicine
- Optical sorting machines
- Image reconstruction

BOSCH

Females in Tech – Computer Vision
High-level Vision: Classic approach

Images → Image description → Processing → Output

- tries to understand the whole scene in the image/video

- Needs a really good representation of the image

- **Examples:**

- Object recognition

- Obstacle detection

- Face recognition e.g. background filter, People counting, …

Source: Ebert, 2012

BOSCH

Females in Tech – Computer Vision
High-level vision: Classic approach

Images ➤ Image description ➤ Processing ➤ Output

Forward Collision Warning System for trams in
The Hague #Bosch Group
May 19, 2023   239   45   13

Person detection

Traffic sign recognition

BOSCH

Females in Tech – Computer Vision
High-level vision: Challenges

Images → Image description → Processing → Output

Different illumination

Pattern recognition
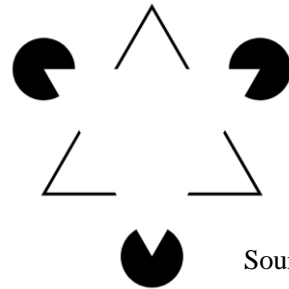
Source: Kaniza 1976

Animals are adapted to their environment

chameleon leopard mountain hare phasmatodea

flounder fish zebra pantomime

Source: Ebert 2012

BOSCH

Females in Tech – Computer Vision
High-level vision: Deep Learning

Images	Image description + Processing	Output

- Learning of both image description and task itself, e.g. object detection, scene understanding, …

- Leads to a huge number of parameter (at least number of pixel)

- But needs a huge amount of data

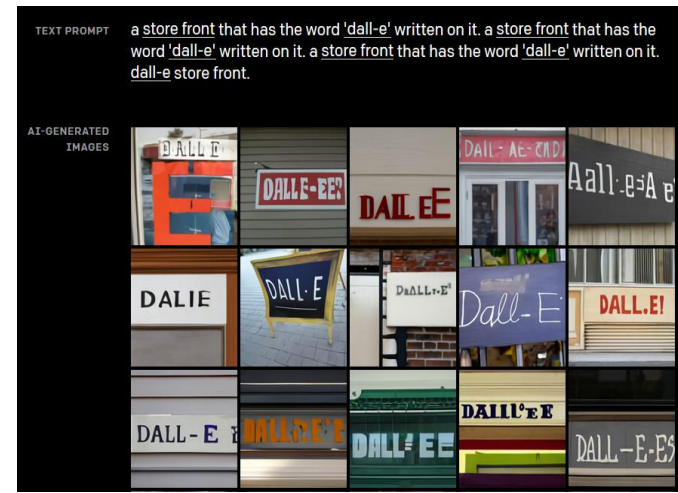**Example**: Rail track detection with semantic segmentation (Sensor4rail)

BOSCH

Females in Tech – Computer Vision
High-level vision: Deep Learning

Text → Transformer → Image

https://openai.com/dall-e-2/#demos

BOSCH

Females in Tech – Computer Vision
High-level vision: Deep Learning

Text ➤ Transformer ➤ Image

https://openai.com/dall-e-2/#demos
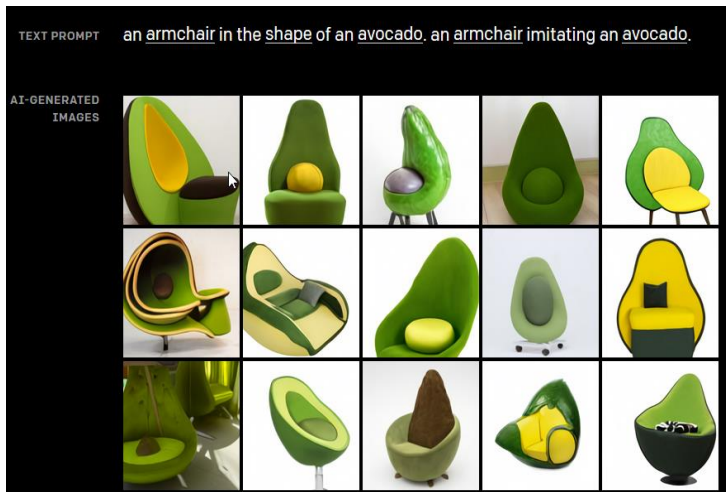
But no algorithm is perfect …

See screenshot

BOSCH

3

Natural Language Processing (NLP)

BOSCH

Natural Language Processing (NLP)
ChatGPT

- Chatbot by OpenAI launched November 2022
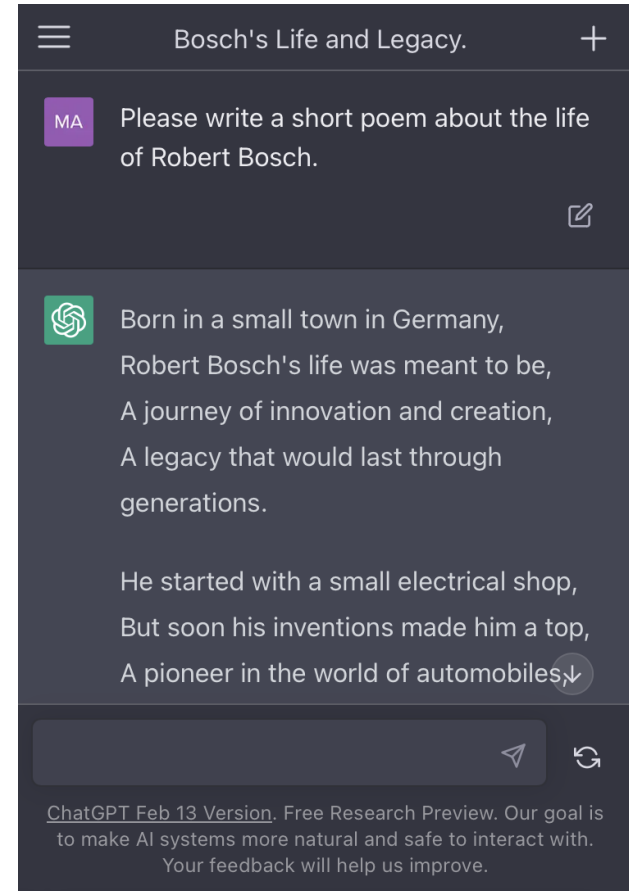- Based on GPT-3, a large language model
- Writes text based on prompts
  - answers to questions, writes letters, poems, recipes, suggestions, summaries, ...
- Can write code in various languages, comment, and debug code
- Sometimes returns plausible-sounding, but factually incorrect answers ("hallucination").

GPT: Generative Pre-trained Transformer

Females in Tech − Natural Language Processing
Voice cloning

- Question is:
  - How long should a sample be to synthesize a voice?

Speaker Adaptation for Unseen Speakers

Each column corresponds to a single speaker. The speaker name is in "Dataset SpeakerID" embedding.

These examples are sampled from the evaluation set for Table 1 and Table 2 in the paper.

VCTK p240   VCTK p260

Reference:

Synthesized:
0: Take a look at these pages for crooked creek drive.

1: There are several listings for gas station.

2: Here's the forecast for the next four days.

3: Here is some information about the Gospel of John.

  - https://youtu.be/0sR1rU3gLzQ?t=27
  - https://google.github.io/tacotron/publications/speaker_adaptation/

BOSCH

Females in Tech
Natural Language Processing (NLP)

- Sentiment Analysis

- Evaluation of e.g. product reviews
  - https://demo.allennlp.org/sentiment-analysis/MjMyNTM0Nw==
  - DIY
    - Try out the model with english sentences of your choice!
    - Where are the limits?

Sentiment Analysis

Sentiment Analysis is the task of interpreting and classifying emotions (positive or negative) in the input text.

Model

RoBERTa

This model is trained on RoBERTa large with the binary classification setting of the Stanford Sentiment Treebank. It achieves 95.11% accuracy on the test set.

Contributed by: Zhaofeng Wu

Demo   Usage

Enter text or | a very well-made, funny and entertaining picture.

Input

a very well-made, funny and entertaining picture.

Run >

Answer

The model is are quite sure the sentence is **Positive.** (100,0 %)

Answer

The model thinks the sentence is **Negative.** (66,1 %)

BOSCH

4

Robotics

BOSCH

Females in Tech
Robotics: Thinking outside the box

Leg that touches the ground less than 10% of the time

Source: [Cully et al. 2015]
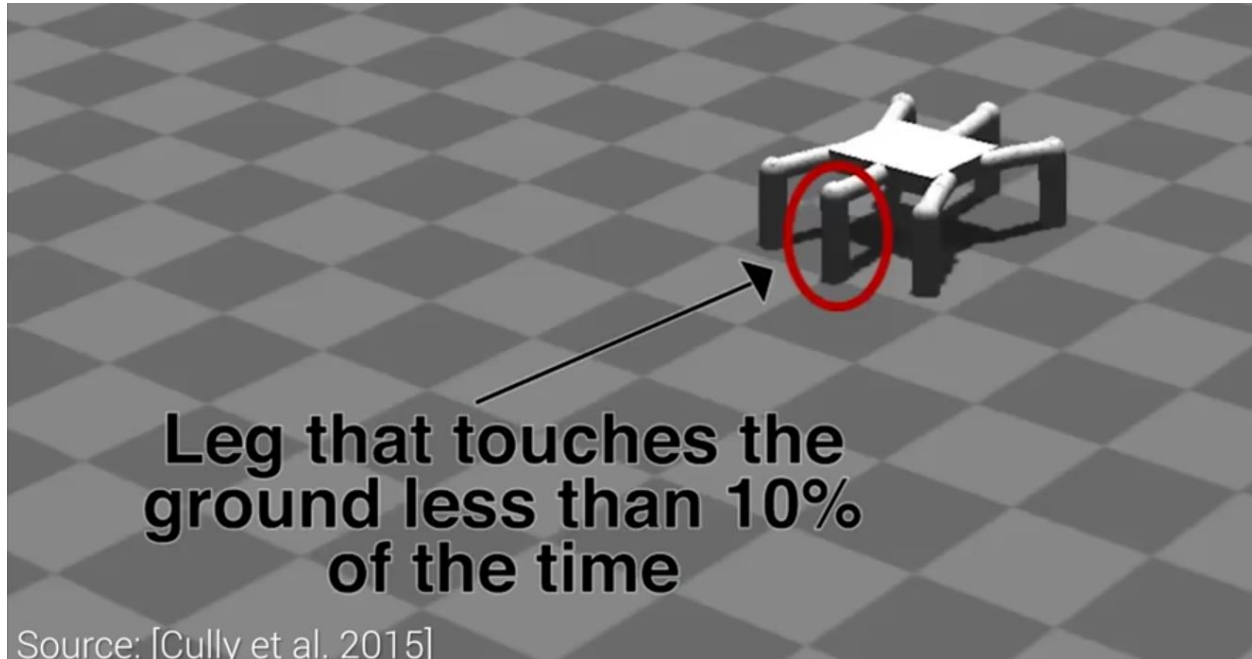
How to minimize the time, all feet touch the ground?

https://youtu.be/GdTBqBnqhaQ?t=20

BOSCH

Females in Tech
High End AI: Google's DeepMind AI

- Teaching itself how to walk

- Every motion the stick figure is making It taught itself

- The model was given virtual sensors
  which tell it information about ist orientation

- Programmers incentivised it
  to go from point A to point B

- The AI was never shown what walking looks like

https://youtu.be/gn4nRCC9TwQ

BOSCH

Females in Tech
Fragen?

BOSCH