Beyond the Hype
Fundamentals and Limits of Artificial Intelligence

Christian Heitzmann, SimplexaCode AG

SimplexaCode

SimplexaCode AG | Grimselweg 11/501 | 6005 Lucerne
043 810 06 03 | info@simplexacode.ch | www.simplexacode.ch

DINAcon 2019 | Welle7 Workspace | Bern
October 18, 2019

Shortlink

Shortlink to

- these slides
- source code examples

$\Longrightarrow$ *link.simplexacode.ch/8yn4*

Terms and Definitions

- algorithm
- artificial intelligence (AI)
- machine learning
- artificial neural networks (ANNs)
- deep learning
- convolutional neural networks (CNNs)

algorithm:

- sequence of instructions for processing data
- examples: sorting algorithms, JPEG compression, AI methods

Terms and Definitions

- algorithm
- artificial intelligence (AI)
- machine learning
- artificial neural networks (ANNs)
- deep learning
- convolutional neural networks (CNNs)

artificial intelligence (AI):

- John McCarthy: “the science and engineering of making intelligent machines”
- Merriam-Webster: “a branch of computer science dealing with the simulation of intelligent behavior in computers”

Terms and Definitions

- algorithm
- artificial intelligence (AI)
- machine learning
- artificial neural networks (ANNs)
- deep learning
- convolutional neural networks (CNNs)

machine learning:

- learning from examples
- generalizing after the end of the learning phase

Terms and Definitions

- algorithm
- artificial intelligence (AI)
- machine learning
- artificial neural networks (ANNs)
- deep learning
- convolutional neural networks (CNNs)

artificial neural networks (ANNs):

- *one* method for machine learning
- neurons receiving input signals, weighting them, and calculating an output signal each

Terms and Definitions

- algorithm
- artificial intelligence (AI)
- machine learning
- artificial neural networks (ANNs)
- deep learning
- convolutional neural networks (CNNs)

deep learning:

- at least 2 hidden layers between input and output layer
- in practice: several hundred hidden layers

Terms and Definitions

- algorithm
- artificial intelligence (AI)
- machine learning
- artificial neural networks (ANNs)
- deep learning
- convolutional neural networks (CNNs)

convolutional neural networks (CNNs):

- class of deep neural networks specifically designed for pixel data and image recognition
- inspired by the behavior of the visual cortex of the brain

What Is Intelligence? | Poldi and Nora

What Is Intelligence? | Poldi and Nora

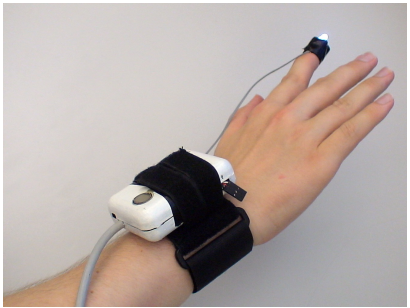
“Guten Morgen
mein kleiner Spatz!”

What Is Intelligence? | Poldi and Nora

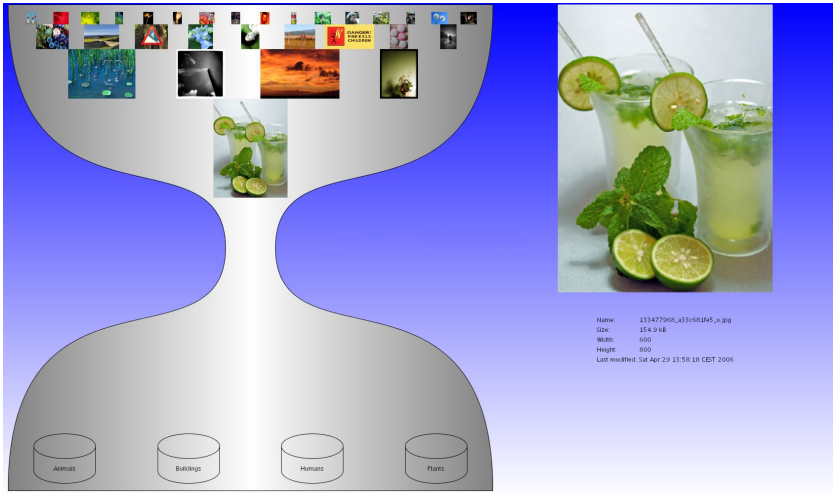
“Komm zum Herrchen!”

What Is Artificial Intelligence? | Gesture-based Image Selection UI

- Gesture-based Image Selection User Interface for Large Screen Interaction Environments, Bachelor Thesis, Computer Graphics Lab, ETH Zurich, 2006
- 3 position coordinates (cameras)
- 2 acceleration coordinates (accelerometers)



Demo | Gesture-based Image Selection User Interface



What AI Can Do

- recognize faces and objects in images (Facebook)
- estimate ages of people (How old do I look?[1])
- recognize speech (Siri)
- filter personal preferences
- recommend videos, music, and products (Netflix)
- write newspaper articles (robot journalism)
- filter out spam
- play games (Google DeepMind)
- drive and park cars autonomously
- route ridesharing fleets (Uber, Lyft)
- optimize picker routes (Zalando)
- check symptoms (Ada)
- diagnose certain types of cancer[2]
- predict crimes (Precobs (Germany), COMPAS (United States))

[1] *https://www.how-old.net*

[2] A. Grävemeyer, KI erkennt Krebs – Neuronale Netze in der Radiologie, c't 14/2018

What AI Can*not* Do

- understand society, ethics, money, competition, love, or sexual desire
- differentiate between originals and parodies (GEMA)
- form the singularity in the future
- be as intelligent as humans
- learn like humans

- $\Longrightarrow$ AI systems are geeks.
- $\Longrightarrow$ Every system masters exactly one thing.
- $\Longrightarrow$ Specialists cannot unite to form an “artificial superintelligence”.

Demo | MNIST Database | Training Phase

- handwritten digits
- 60,000 training images
- 10,000 test images
- `exec(open('mnist_database.py').read())`

[3] R. Schwaiger et al., Neuronale Netze programmieren mit Python, Rheinwerk Verlag, 2019

How ANNs Work

1. Set parameters of every neuron to random values.
2. Give feedback to the network after each training example.
3. Readjust all parameters according to their contribution to the error.
4. Repeat steps 2 and 3 until the error rate is below a certain threshold.
5. Freeze all parameters after the end of the training phase.

Demo | Neural Network Playground

playground.tensorflow.org

Demo | MNIST Database | Test Phase

```
model.evaluate(testImages, testLabels)
```

3 Main Ingredients of Deep Learning

1. masses of data
2. indexed data (e.g., tagged objects in images)
3. computing power (e.g., cloud, data centers, or specialized hardware)

Demo | Pretrained Inception-v3 Convolutional Neural Network

- `inception_v3.py`
- `picture.jpg`

Weaknesses of Deep Learning

- thousands or millions of training data (e.g., tagged images)
- very time-consuming and CPU-intensive[4]
- biases and errors (see next slide)[5]
- manipulations (see slide after next)
- segregation and combination problems[6]
- exploding and vanishing gradients, overfitting, and saturation[7]

[4] C. Heitzmann, Warum künstliche Intelligenz so schwierig ist – Einführung in die Komplexität von Algorithmen, JavaSPEKTRUM 3/2019, *link.simplexacode.ch/mysk*

[5] P. Merkert, Statistik ist nicht Denken, c't 24/2018

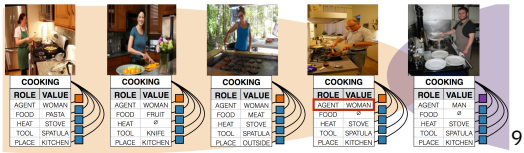
[6] P. Merkert, Statistik ist nicht Denken, c't 24/2018

[7] R. Schwaiger et al., Neuronale Netze programmieren mit Python, Rheinwerk Verlag, 2019

Weaknesses of Deep Learning | Biases and Errors

- recognize huskies only with snow in the background
- recognize trains only based on platforms and tracks, not locomotives or cars
- classify black people as gorillas (Google)
- favor male over female job applicants (Amazon)
- tag people in kitchen situations as female (imSitu)
- produce false-positives with deadly consequences (SKYNET)


[8]



[9]

[8] Pixabay

[9] J. Zhao et al., Men Also Like Shopping, University of Virginia, 2017

Weaknesses of Deep Learning | Manipulations

- trick antivirus software by appending specific strings[10]
- noise patterns and optical illusions

11

12

[10] J. Schmidt, Antivirus: Cylance-KI böse ausgetrickst, heise Security, 2019

[11] M. E. Hvass Pedersen, TensorFlow Tutorial #11, Adversarial Examples, *github.com/Hvass-Labs/*

[12] M. E. Hvass Pedersen, TensorFlow Tutorial #11, Adversarial Examples, *github.com/Hvass-Labs/*

Weaknesses of Deep Learning | Manipulations

- Face Swap




[13]

[13] all pictures: *www.dailymail.co.uk/femail/article-4261400/The-biggest-face-swap-FAILS.html*

3 Main Reasons for Flawed Deep Learning

1. no contextual understanding
2. too little traceability
3. too much tinkering

Reasons for Flawed Deep Learning | No Contextual Understanding

- AI can perfectly find correlations between X and Y.
- AI can*not* find the causation of any relationship between X and Y.
- in general: Correlation does not imply causation.
- examples from daily life:
  - more wind $\longleftrightarrow$ faster windmills
  - children watching TV $\longleftrightarrow$ children being more violent
  - increase in ice cream sales $\longleftrightarrow$ increase in drowning deaths

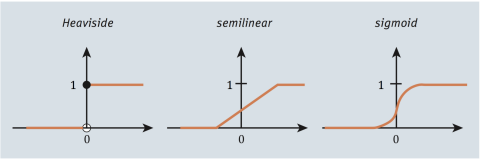
Reasons for Flawed Deep Learning | Too Little Traceability

- “black box” because / in spite of loads of parameters
- no traceability, no insight, and no debugging tools
- biased systems because of biased training data
- $\Longrightarrow$ “explainable AI”
- examples:
  - decision-making of cars right before a crash
  - medical decisions
  - rejected loans[14]

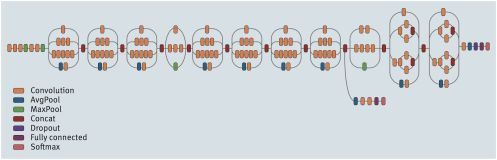
[14] cf. GDPR, Article 22, “Automated individual decision-making, including profiling”

Reasons for Flawed Deep Learning | Too Much Tinkering

- initializations
- network types
- activation functions
- gradient optimizations
- early stoppings
- dropouts



15



16

[15] R. Schwaiger et al., Neuronale Netze programmieren mit Python, Rheinwerk Verlag, 2019
[16] R. Schwaiger et al., Neuronale Netze programmieren mit Python, Rheinwerk Verlag, 2019

Demo | Convolution Matrices

setosa.io/ev/image-kernels

Conclusion

1. There is and never will be “the AI”, but very specific AI implementations that excel within their very specific domains, e.g.:
   - speech recognition
   - computer vision
   - logistics
2. Both AI systems and human experts will be important and support each other in the future.
3. AI implementations can only be as good as their developers understand the underlying problems and solutions.

Questions?

SimplexaCode