WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Wikidata item

Print/export
Download as PDF
Printable version

In other projects
Wikimedia Commons
Wikibooks

Artificial intelligence

From Wikipedia, the free encyclopedia

"AI" redirects here. For other uses, see AI (disambiguation) and Artificial intelligence (disambiguation).

Artificial intelligence (**AI**), is intelligence demonstrated by machines, unlike the **natural intelligence** displayed by humans and animals, which involves consciousness and emotionality. The distinction between the former and the latter categories is often revealed by the acronym chosen. 'Strong' AI is usually labelled as AGI (Artificial General Intelligence) while attempts to emulate 'natural' intelligence have been called ABI (Artificial Biological Intelligence). Leading AI textbooks define the field as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals.[3] Colloquially, the term "artificial intelligence" is often used to describe machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving".[4]

Part of a series on
Artificial intelligence
Major goals [show]
Approaches [show]
Philosophy [show]
History [show]
Technology [show]
Glossary [show]
V · T · E

As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the AI effect.[5] A quip in Tesler's Theorem says "AI is whatever hasn't been done yet."[6] For instance, optical character recognition is frequently excluded from things considered to be AI,[7] having become a routine technology.[8] Modern machine capabilities generally classified as AI include successfully understanding human speech,[9] competing at the highest level in strategic game systems (such as chess and Go),[10] autonomously operating cars, intelligent routing in content delivery networks, and military simulations.[11]

Artificial intelligence was founded as an academic discipline in 1955, and in the years since has experienced several waves of optimism,[12][13] followed by disappointment and the loss of funding (known as an "AI winter"),[14][15] followed by new approaches, success and renewed funding.[13][16] After AlphaGo successfully defeated a professional Go player in 2015, artificial intelligence once again attracted widespread global attention.[17] For most of its history, AI research has been divided into sub-fields that often fail to communicate with each other.[18] These sub-fields are based on technical considerations, such as particular goals (e.g. "robotics" or "machine learning"),[19] the use of particular tools ("logic" or artificial neural networks), or deep philosophical differences.[22][23][24] Sub-fields have also been based on social factors (particular institutions or the work of particular researchers).[18]

The traditional problems (or goals) of AI research include reasoning, knowledge representation, planning, learning, natural language processing, perception and the ability to move and manipulate objects.[19] General intelligence is among the field's long-term goals.[25] Approaches include statistical methods, computational intelligence, and traditional symbolic AI. Many tools are used in AI, including versions of search and mathematical optimization, artificial neural networks, and methods based on statistics, probability and economics. The AI field draws upon computer science, information engineering, mathematics, psychology, linguistics, philosophy, and many other fields.

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Wikidata item

Print/export
Download as PDF

Deep learning

From Wikipedia, the free encyclopedia

Deep learning (also known as **deep structured learning**) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised.[1][2][3]

Deep-learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, machine vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance.[4][5][6]

Artificial neural networks (ANNs) were inspired by information processing and distributed communication nodes in biological systems. ANNs have various differences from biological brains. Specifically, neural networks tend to be static and symbolic, while the biological brain of most living organisms is dynamic (plastic) and analog.[7][8][9]

The adjective "deep" in deep learning comes from the use of multiple layers in the network. Early work showed that a linear perceptron cannot be a universal classifier, and then that a network with a nonpolynomial activation function with one hidden layer of unbounded width can on the other hand so be. Deep learning is a modern variation which is concerned with an unbounded number of layers of bounded size, which permits practical application and optimized implementation, while retaining theoretical universality under mild conditions. In deep learning the layers are also permitted to be heterogeneous and to deviate widely from biologically informed connectionist models, for the sake of efficiency, trainability and understandability, whence the "structured" part.

Part of a series on
Machine learning and data mining

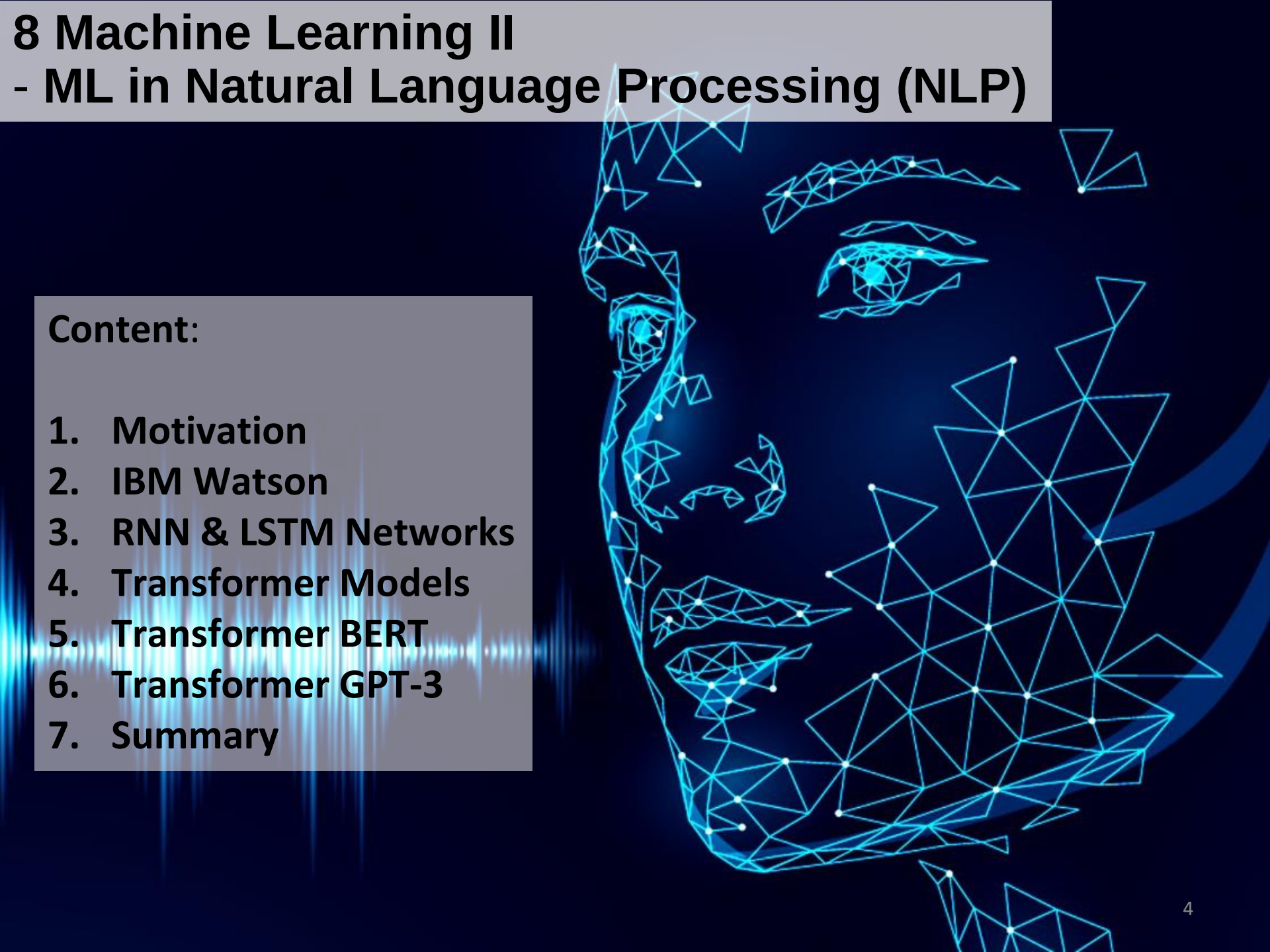
Problems	[show]
Supervised learning (classification • regression)	[show]
Clustering	[show]
Dimensionality reduction	[show]
Structured prediction	[show]
Anomaly detection	[show]
Artificial neural network	[show]
Reinforcement learning	[show]
Theory	[show]
Machine-learning venues	[show]
Glossary of artificial intelligence	[show]
Related articles	[show]

V • T • E

Part of a series on
Artificial intelligence

Major goals	[show]

8 Machine Learning II
- ML in Natural Language Processing (NLP)

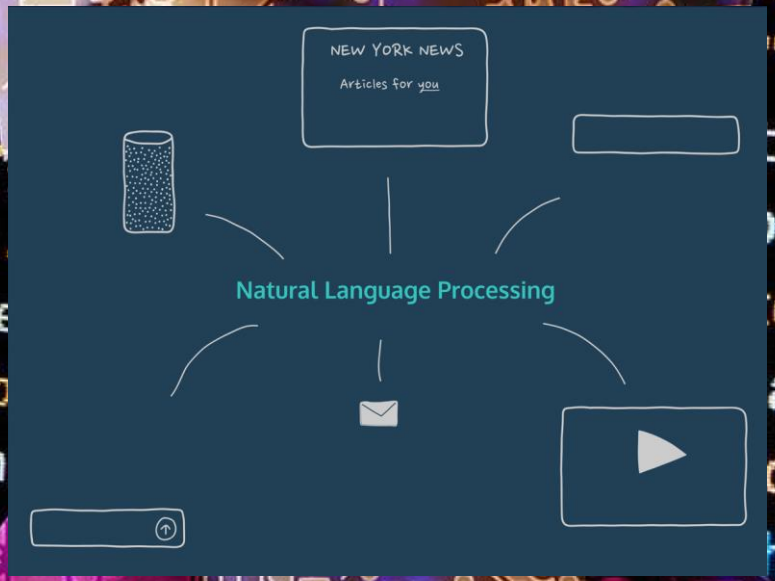


Content:

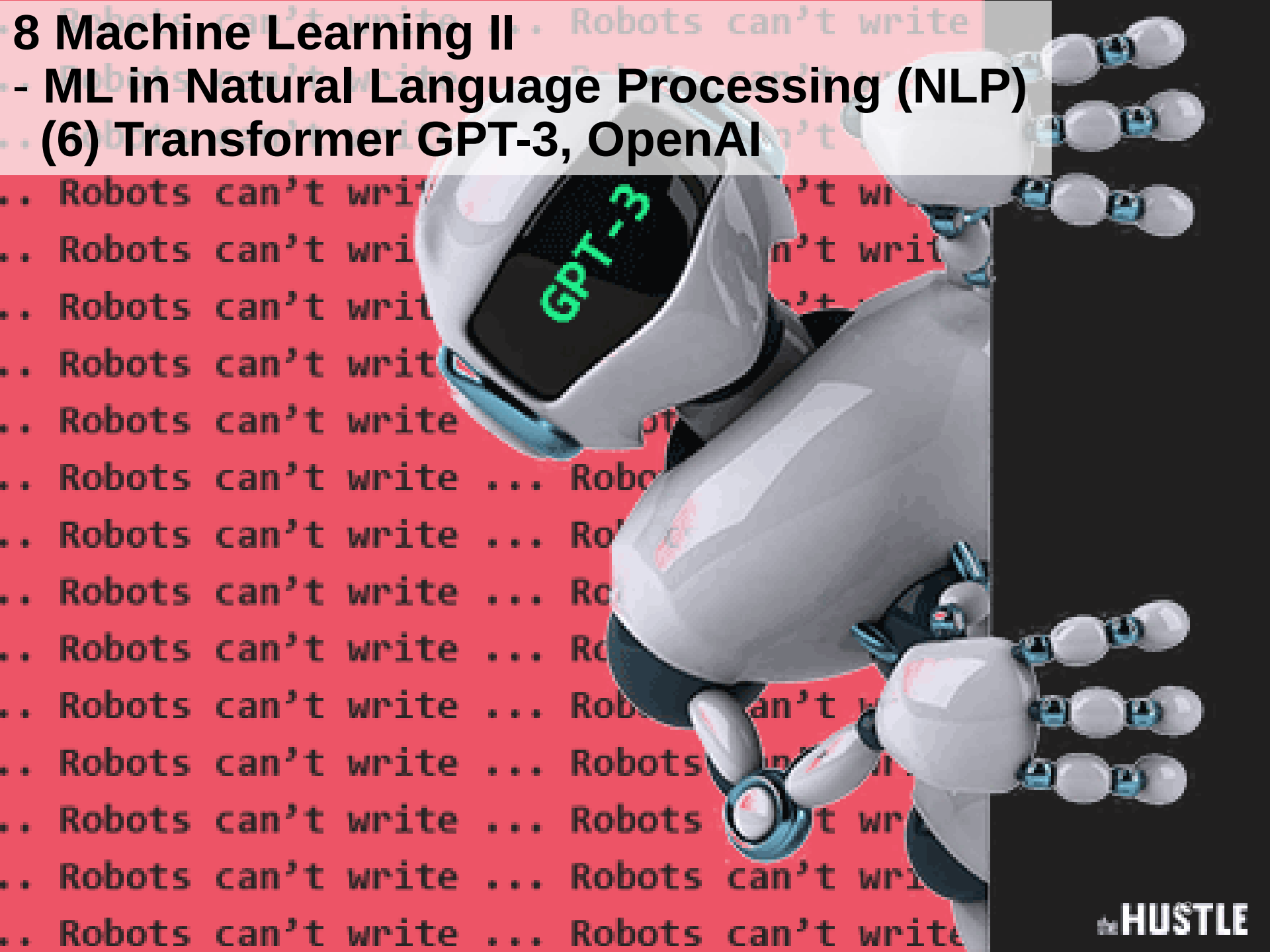
1. Motivation
2. IBM Watson
3. RNN & LSTM Networks
4. Transformer Models
5. Transformer BERT
6. Transformer GPT-3
7. Summary

8 Machine Learning II
- ML in Natural Language Processing (NLP)
(5) Transformer BERT, Google

8 Machine Learning II
- ML in Natural Language Processing (NLP)
(1) Motivation & Intro NLP



8 Machine Learning II
- ML in Natural Language Processing (NLP)
(6) Transformer GPT-3, OpenAI



GPT-3 vs BERT?

Both, **GPT-3 and BERT** have been **relatively new** for the industry. Their state-of-the-art performance has made them the **winners among other models** in the **natural language processing field**.
Being trained on 175 billion parameters, **GPT-3** becomes **470 times bigger in size** than **BERT-Large**.

While **BERT** requires an **elaborated fine-tuning process**, **GPT-3's** allows the users to **reprogram it using instructions and access it**.

Case in point — for **sentiment analysis** or **question answering** tasks, to use **BERT**, the users **have to train the model** on a separate layer on sentence encodings. However, **GPT-3** uses a **few-shot learning process** on the input token to predict the output result.

GPT-3 Vs BERT For NLP Tasks
analyticsindiamag.com/gpt-3-vs-bert-for-nlp-tasks/

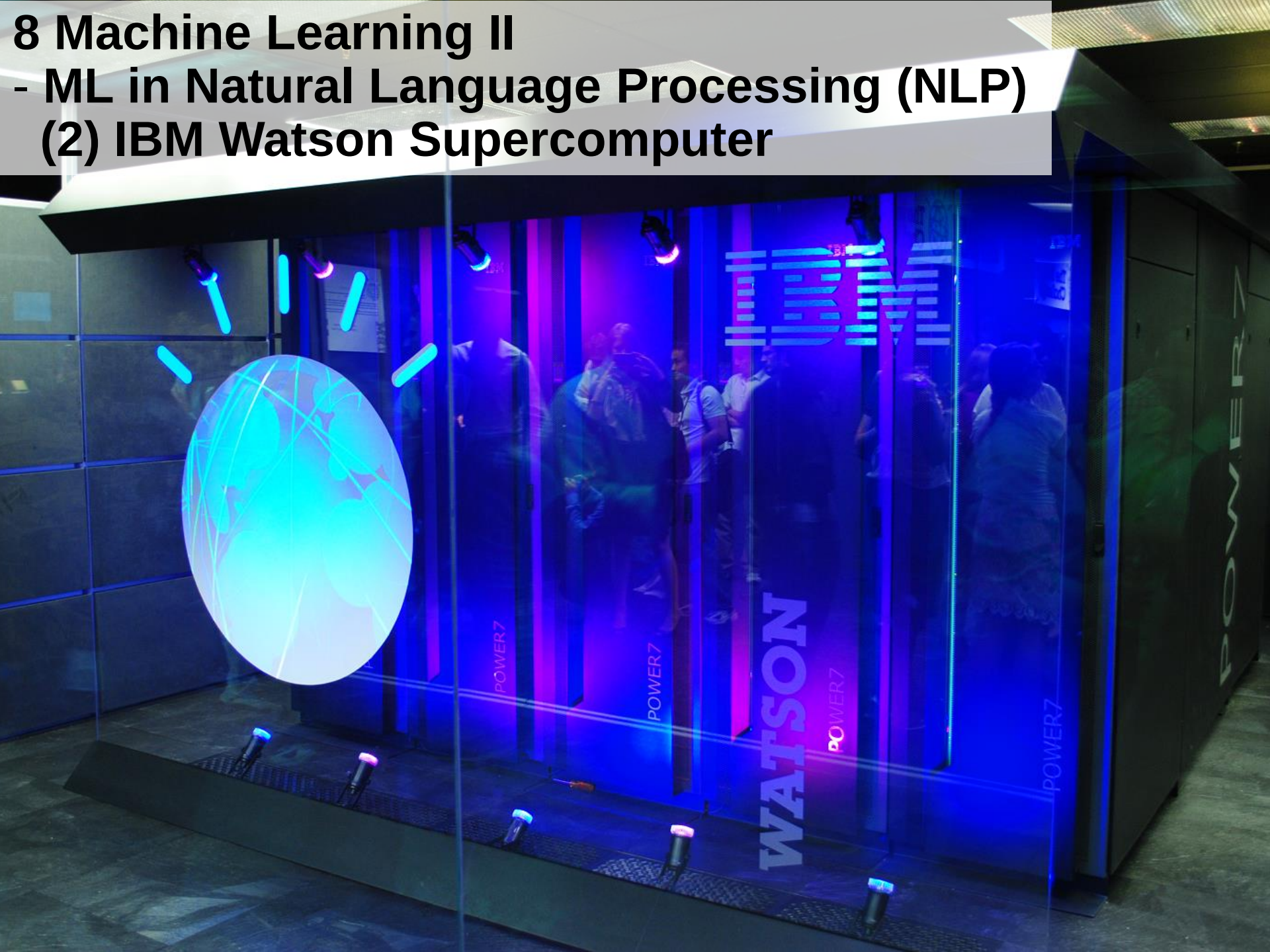
8 Machine Learning II
- ML in Natural Language Processing (NLP)
(7) Summary

Content:

1. **Motivation**
2. **IBM Watson**
3. **RNN & LSTM Networks**
4. **Transformer Models**
5. **Transformer BERT**
6. **Transformer GPT-3**
7. **Summary**

- Vor **Transformer** wurden für **NLP RNNs** und **LSTMs** eingesetzt.
- **Transformer** bauen auf dem **Attention-Mechanismus** auf und erlauben Parallel-verarbeitung.
- Die aktuell führenden **Transformer** sind **BERT** und **GPT**.

8 Machine Learning II
- ML in Natural Language Processing (NLP)
(2) IBM Watson Supercomputer



Natural Language Processing

Die Abkürzung **NLP** steht für **Natural Language Processing** und beschreibt **Techniken und Methoden** zur **maschinellen Verarbeitung natürlicher Sprache**.

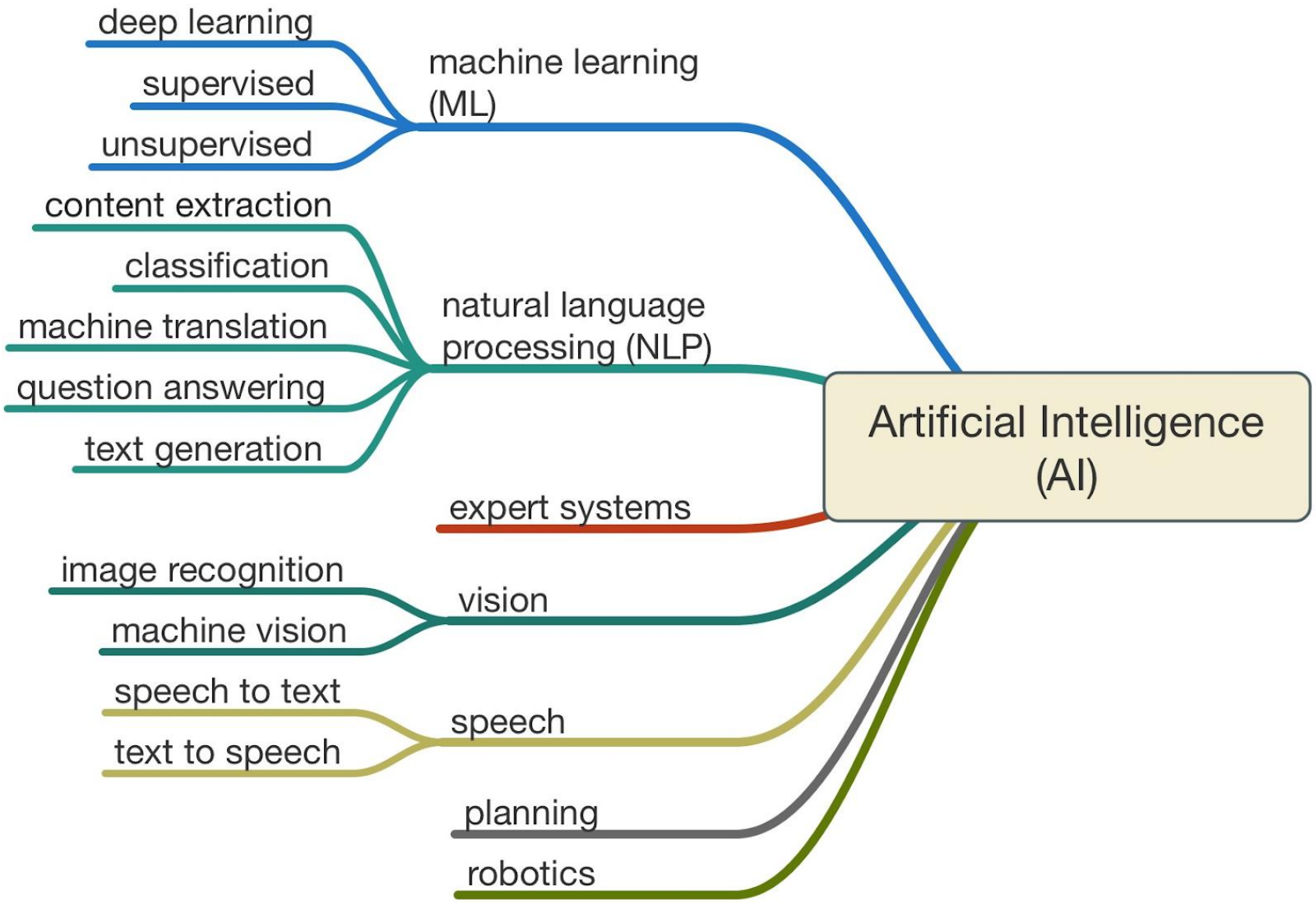
Ziel ist eine **direkte Kommunikation** zwischen **Mensch** und **Computer** auf Basis der **natürlichen Sprache**.

NLP muss **Lösungen** schaffen, um sowohl **gesprochene** als auch **geschriebene Sprache** zu erkennen, zu analysieren und den Sinn zur weiteren Verarbeitung zu extrahieren. Hierfür ist ein **Verständnis** nicht nur von einzelnen Wörtern und Sätzen, sondern das Erfassen von **kompletten Textzusammenhängen** und Sachverhalten notwendig.

Eine **Herausforderung** für das **Natural Language Processing** stellt die **Komplexität** der **menschlichen Sprache** und deren **Mehrdeutigkeit** dar.

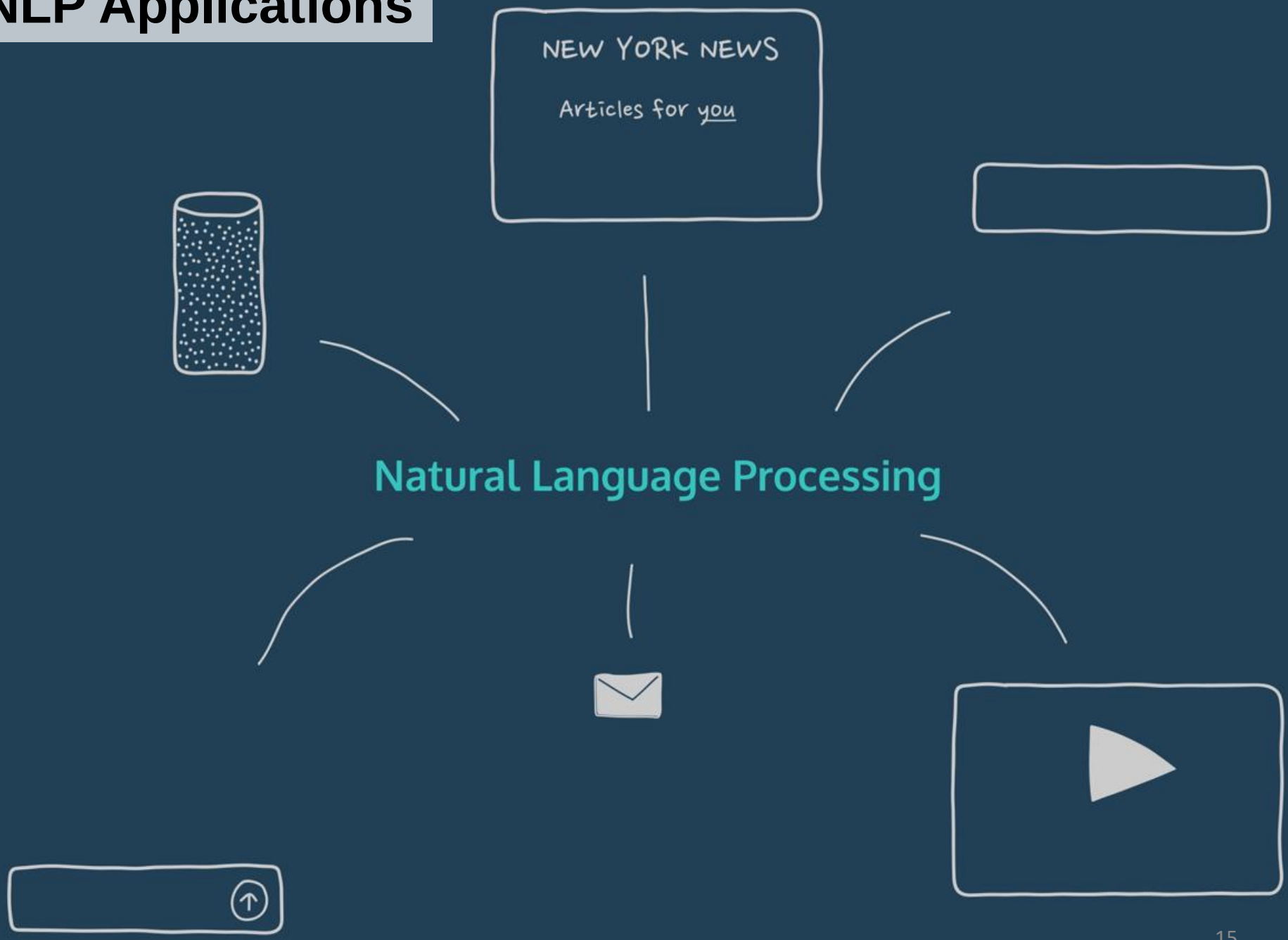
Was ist Natural Language Processing?
www.bigdata-insider.de/was-ist-natural-language-processing-a-590102/

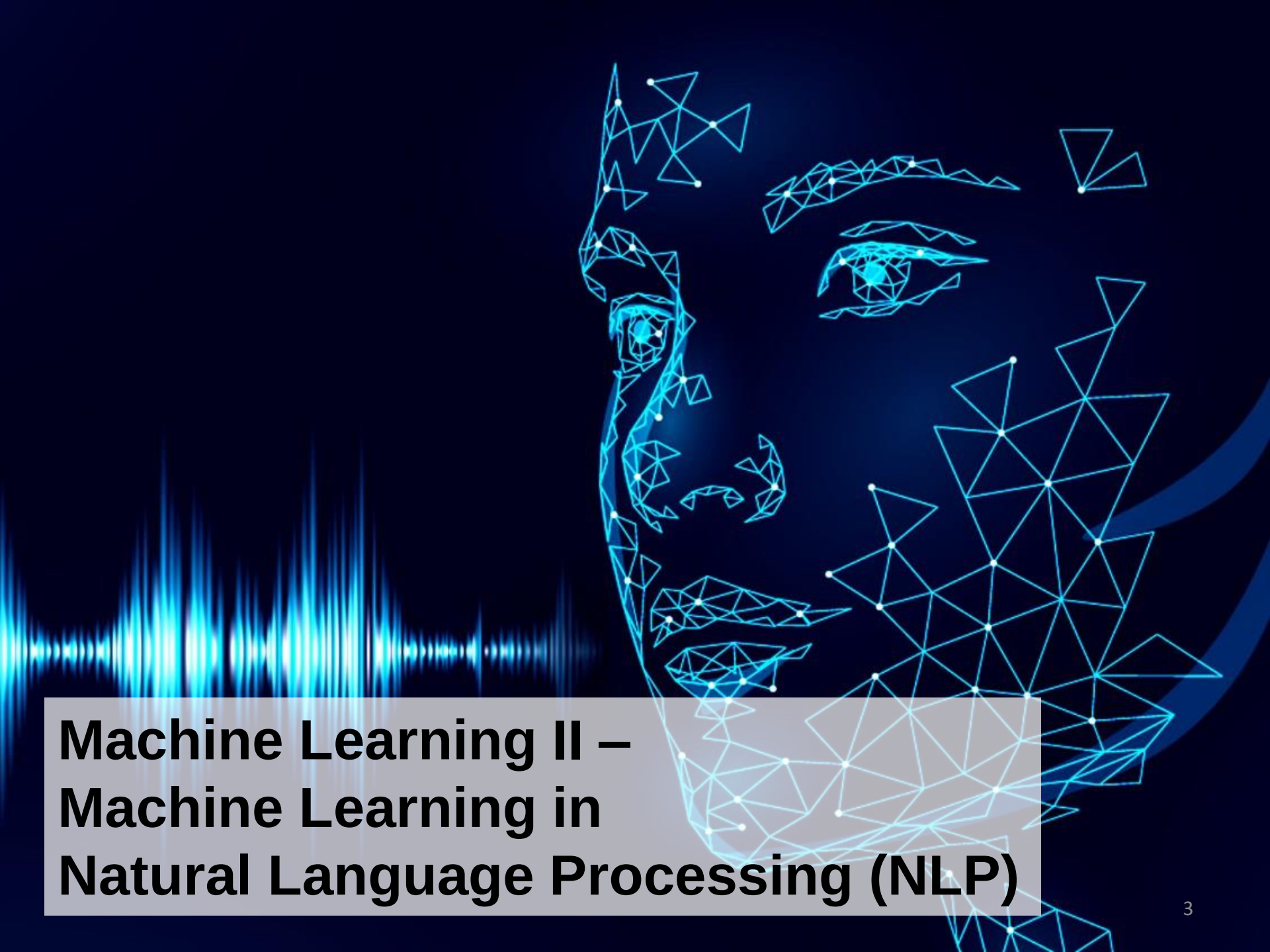
Overview **Natural Language Processing**



The Complete Roadmap for AI, Data Scientist Aspirants, 2020
tekrajawasthi15.medium.com/the-complete-roadmap-to-be-a-data-scientist-9a07721b88fd

NLP Applications





Machine Learning II – Machine Learning in Natural Language Processing (NLP)

IBM Watson Supercomputer

Watson is an **IBM supercomputer** that combines **artificial intelligence** (AI) and sophisticated **analytical software** for performance as a **"question answering" machine**.

The Watson supercomputer processes at a **rate of 80 teraflops** (trillion floating-point operations per second). Watson accesses **90 servers** with a combined **data store of over 200 million pages of information**, which it processes against **six million logic rules**.
The device and its data are **self-contained** in a **space** that could accommodate **10 refrigerators**.

Watson's key components include:

- **Apache UIMA** (Unstructured Information Management Architecture) frameworks, infrastructure and other elements required for the **analysis of unstructured data**.
- **Apache's Hadoop**, a free, Java-based programming framework that supports the **processing of large data sets in a distributed computing environment**.
- **SUSE Enterprise Linux Server 11**, the fastest available **Power7 processor operating system**.
- **2,880 processor cores**.
- **15 terabytes of RAM**.
- **500 gigabytes** of **preprocessed information**.
- **IBM'sDeepQA software**, which is designed for information retrieval that incorporates **natural language processing** and **machine learning**.

Quelle: http://whatis.techtarget.com/definition/IBM-Watson-supercomputer

Transformer Networks (machine learning model)

Transformer (machine learning model)

From Wikipedia, the free encyclopedia

The **Transformer** is a deep learning model introduced in 2017, used primarily in the field of natural language processing (NLP).[1]

Like recurrent neural networks (RNNs), Transformers are designed to handle sequential data, such as natural language, for tasks such as translation and text summarization. However, unlike RNNs, Transformers do not require that the sequential data be processed in order. For example, if the input data is a natural language sentence, the Transformer does not need to process the beginning of it before the end. Due to this feature, the Transformer allows for much more parallelization than RNNs and therefore reduced training times.[1]

Transformers have rapidly become the model of choice for NLP problems,[2] replacing older recurrent neural network models such as the long short-term memory (LSTM). Since the Transformer model facilitates more parallelization during training, it has enabled training on larger datasets than was possible before it was introduced. This has led to the development of pretrained systems such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), which have been trained with huge general language datasets, such as Wikipedia Corpus, and can be fine-tuned to specific language tasks.[3][4]

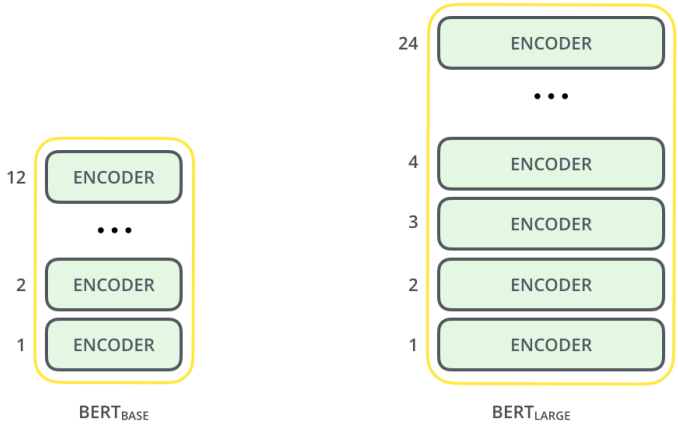
BERT Transformer, Google 2018

BERT (language model)

From Wikipedia, the free encyclopedia

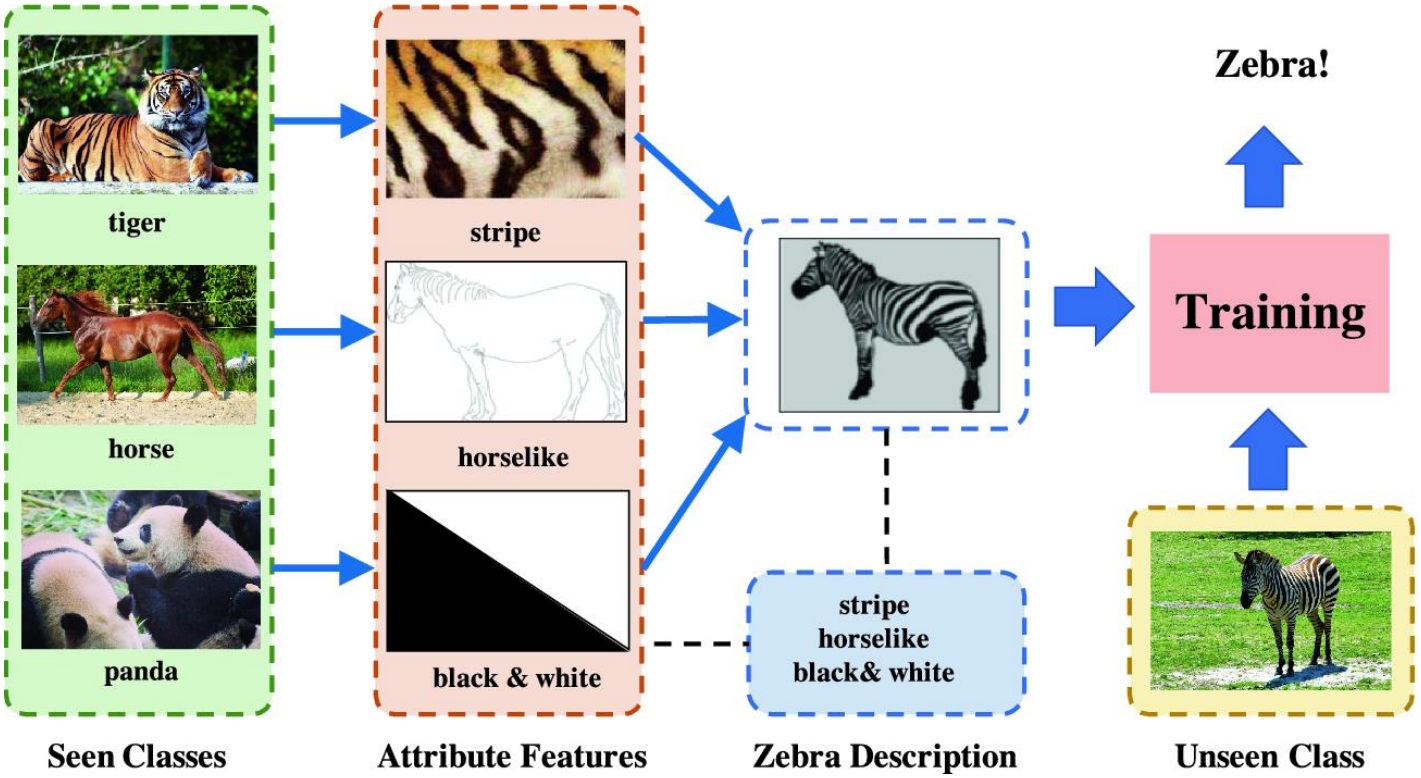
Bidirectional Encoder Representations from Transformers (**BERT**) is a Transformer-based machine learning technique for natural language processing (NLP) pre-training developed by Google. BERT was created and published in 2018 by Jacob Devlin and his colleagues from Google.[1][2] As of 2019, Google has been leveraging BERT to better understand user searches.[3]

The original English-language BERT model comes with two pre-trained general types:[1] (1) the $\text{BERT}_{\text{BASE}}$ model, a 12-layer, 768-hidden, 12-heads, 110M parameter neural network architecture, and (2) the $\text{BERT}_{\text{LARGE}}$ model, a 24-layer, 1024-hidden, 16-heads, 340M parameter neural network architecture; both of which were trained on the BooksCorpus[4] with 800M words, and a version of the English Wikipedia with 2,500M words.



Zero-Shot Learning (Kernidee)

In **Machine Learning**, **Zero-Shot Learning** is a **problem setup** where, **at test stage**, a learner **recognizes object from classes not previously seen at training stage**. This problem is widely studied in **computer vision**, **natural language processing** and **machine perception**.



en.wikipedia.org/wiki/Zero-shot_learning
deepai.org/publication/zero-shot-learning-by-generating-pseudo-feature-representations

GPT-3 Transformer, OpenAI 2020

GPT-3

From Wikipedia, the free encyclopedia

Generative Pre-trained Transformer 3 (**GPT-3**) is an autoregressive language model that uses deep learning to produce human-like text. It is the third-generation language prediction model in the GPT-n series (and the successor to GPT-2) created by OpenAI, a San Francisco-based artificial intelligence research laboratory.[2] GPT-3's full version has a capacity of 175 billion machine learning parameters. GPT-3, which was introduced in May 2020, and was in beta testing as of July 2020,[3] is part of a trend in natural language processing (NLP) systems of pre-trained language representations.[1] Before the release of GPT-3, the largest language model was Microsoft's Turing NLG, introduced in February 2020, with a capacity of 17 billion parameters or less a tenth of GPT-3s.[4]

The quality of the text generated by GPT-3 is so high that it is difficult to distinguish from that written by a human, which has both benefits and risks.[4] Thirty-one OpenAI researchers and engineers presented the original May 28, 2020 paper introducing GPT-3. In their paper, they warned of GPT-3's potential dangers and called for research to mitigate risk.[1]:34 David Chalmers, an Australian philosopher, described GPT-3 as "one of the most interesting and important AI systems ever produced."[5]

Microsoft announced on September 22, 2020 that it had licensed "exclusive" use of GPT-3; others can still use the public API to receive output, but only Microsoft has control of the source code.[6]