

# OECD, Wikipedia: 21<sup>st</sup> Century Skills



PARTNERSHIP FOR  
21ST CENTURY SKILLS

## 21<sup>st</sup> Century Skills: How can you prepare students for the new Global Economy?

### 21st century skills

From Wikipedia, the free encyclopedia

Charles Fadel  
Global Lead,  
Cisco Systems

OECD/CERI  
Paris, May 20

21st century skills comprise [skills](#), [abilities](#), and learning [dispositions](#) that have been identified as being required for success in 21st century society and workplaces by educators, business leaders, [academics](#), and [governmental agencies](#). This is part of a growing international movement focusing on the skills required for students to master in preparation for success in a rapidly changing, digital society. Many of these skills are also associated with [deeper learning](#), which is based on mastering skills such as [analytic reasoning](#), [complex problem solving](#), and [teamwork](#). These skills differ from traditional academic skills in that they are not primarily content knowledge-based.<sup>[1][2][3]</sup>



KIPEDIA  
e Enzyklopädie

seite

importale

er Artikel

nen

verbessern

Artikel anlegen

portal

änderungen

n

uge

uf diese Seite

ngen an

en Seiten

seiten

enter Link

informationen

citieren

a-Datenobjekt

n/exportieren

stellen

F herunterladen

# Data Science

**Data Science** (von englisch *data* „Daten“ und *science* „Wissenschaft“, im Deutschen auch **Datenwissenschaft**) bezeichnet generell die Extraktion von **Wissen** aus Daten.<sup>[1][2]</sup>

Data Science ist ein interdisziplinäres Wissenschaftsfeld, welches wissenschaftlich fundierte Methoden, Prozesse, Algorithmen und Systeme zur Extraktion von Erkenntnissen, Mustern und Schlüssen sowohl aus strukturierten als auch unstrukturierten Daten ermöglicht<sup>[3][4]</sup>.

Der Studiengang *Data Science* verwendet Techniken und Theorien aus den Fächern Mathematik, **Statistik** und Informationstechnologie, einschließlich der **Signalverarbeitung**, verwendet Wahrscheinlichkeitsmodelle des **maschinellen Lernens**, des statistischen Lernens, der Programmierung, der Datentechnik, der **Mustererkennung**, der **Prognostik**, der Modellierung von Unsicherheiten und der Datenlagerung.

Personen, die im Bereich Data Science arbeiten, werden als **Data Scientist** bzw. **Datenwissenschaftler** bezeichnet, wobei meist speziellere oder Spezialisierungen anderer, übergeordneter Berufsbezeichnungen üblich sind (z. B. Statistiker, Informatiker).

## Inhaltsverzeichnis [Verbergen]

- 1 Geschichte
- 2 Das Berufsfeld
  - 2.1 Anforderungen
  - 2.2 Aufgabenbereich
  - 2.3 Ausbildungsmöglichkeiten
  - 2.4 Branchen
- 3 Siehe auch
- 4 Literatur
- 5 Weblinks



[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](#)

# Data science

From Wikipedia, the free encyclopedia

(Redirected from [Data Science](#))

*Not to be confused with [information science](#).*

! **Data science** is an [inter-disciplinary](#) field that uses scientific methods, processes, algorithms and systems to extract [knowledge](#) and insights from many structural and [unstructured data](#).<sup>[1][2]</sup> Data science is related to [data mining](#), [machine learning](#) and [big data](#).

! Data science is a "concept to unify [statistics](#), [data analysis](#) and their related methods" in order to "understand and analyze actual phenomena" with data.<sup>[3]</sup> It uses techniques and theories drawn from many fields within the context of [mathematics](#), [statistics](#), [computer science](#), [domain knowledge](#) and [information science](#). Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science ([empirical](#), [theoretical](#), [computational](#) and now [data-driven](#)) and asserted that "everything about science is changing because of the impact of information technology" and the [data deluge](#).<sup>[4][5]</sup>

## Contents [hide] !

- 1 Foundations
  - 1.1 Relationship to statistics
- 2 Etymology
  - 2.1 Early usage
  - 2.2 Modern usage
- 3 Careers in data science
  - 3.1 Educational path
  - 3.2 Specializations and associated careers
- 4 Impacts of data science
- 5 Technologies and techniques

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



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

Article Talk

Not logged in Talk Contributions Create account Log in

Read Edit View history

Search Wikipedia



# Enterprise resource planning

From Wikipedia, the free encyclopedia

**Enterprise resource planning (ERP)** is the integrated management of main business processes, often in real time and mediated by software and technology.

ERP is usually referred to as a category of [business management software](#)—typically a suite of integrated [applications](#)—that an organization can use to collect, store, manage, and interpret data from many [business activities](#).

ERP provides an integrated and continuously updated view of core business processes using common [databases](#) maintained by a [database management system](#). ERP systems track business resources—cash, raw materials, production capacity—and the status of business commitments: orders, purchase orders, and payroll. The applications that make up the system share data across various departments (manufacturing, purchasing, sales, [accounting](#), etc.) that provide the data.<sup>[1]</sup> ERP facilitates information flow between all business functions and manages connections to outside stakeholders.<sup>[2]</sup>

Enterprise system software is a multibillion-dollar industry that produces components supporting a variety of business functions. IT investments have, as of 2011, become one of the largest categories of capital expenditure in United States-based businesses. Though early ERP systems focused on large enterprises, smaller enterprises increasingly use ERP systems.<sup>[3]</sup>

The ERP system integrates varied organizational systems and facilitates error-free transactions and production, thereby enhancing the organization's efficiency. However, developing an ERP system differs from traditional system development.<sup>[4]</sup> ERP systems run on a variety of [computer hardware](#) and [network](#) configurations, typically using a [database](#) as an [information repository](#).<sup>[5]</sup>

## Business administration

### Management of a business

[Accounting](#) [show]

[Business entities](#) [show]

[Corporate governance](#) [show]

[Corporate law](#) [show]

[Corporate title](#) [show]

[Economics](#) [show]

[Finance](#) [show]

[Types of management](#) [show]

[Organization](#) [show]

[Trade](#) [show]



V · T · E

Part of a series of articles on

## Machine industry



### Manufacturing methods

[Batch production](#) · [Job production](#) ·  
[Flow production](#) · [Lean manufacturing](#) ·



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

Article Talk

Not logged in Talk Contributions Create account Log in

Read Edit View history

Search Wikipedia



# Customer relationship management

From Wikipedia, the free encyclopedia

**Customer relationship management (CRM)** is one of many different approaches that allow a company to manage and analyse its own interactions with its past, current and potential customers.<sup>[1]</sup> It uses data analysis about customers' history with a company to improve business relationships with customers, specifically focusing on customer retention and ultimately driving sales growth.<sup>[2]</sup>

One important aspect of the CRM approach is the systems of CRM compile data from a range of different communication channels, including a company's website, telephone, email, live chat, marketing materials and more recently, social media.<sup>[3]</sup> Through the CRM approach and the systems used to facilitate it, businesses learn more about their target audiences and how to best cater for their needs.

## Contents [hide]

- 1 Software history
- 2 Types
  - 2.1 Strategic
  - 2.2 Operational
  - 2.3 Analytical
  - 2.4 Collaborative
  - 2.5 Customer Data Platform
- 3 Components
- 4 Effect on customer satisfaction
  - 4.1 Customer benefits
  - 4.2 Examples
- 5 Customer profile

## Business administration

### Management of a business

[Accounting](#) [show]

[Business entities](#) [show]

[Corporate governance](#) [show]

[Corporate law](#) [show]

[Corporate title](#) [show]

[Economics](#) [show]

[Finance](#) [show]

[Types of management](#) [show]

[Organization](#) [show]

[Trade](#) [show]



[Business and economics portal](#)

V · T · E



WIKIPEDIA  
The Free Encyclopedia

Article Talk

Not logged in Talk Contributions Create account Log in

Read Edit View history

Search Wikipedia



# Business intelligence

From Wikipedia, the free encyclopedia

**Business intelligence (BI)** comprises the strategies and technologies used by enterprises for the [data analysis of business information](#).<sup>[1]</sup> BI technologies provide historical, current, and predictive views of [business operations](#). Common functions of business intelligence technologies include reporting, [online analytical processing](#), [analytics](#), [dashboard development](#), [data mining](#), [process mining](#), [complex event processing](#), [business performance management](#), [benchmarking](#), [text mining](#), [predictive analytics](#), and [prescriptive analytics](#). BI technologies can handle large amounts of structured and sometimes unstructured data to help identify, develop, and otherwise create new strategic [business opportunities](#). They aim to allow for the easy interpretation of these [big data](#). Identifying new opportunities and implementing an effective strategy based on [insights](#) can provide [businesses](#) with a competitive market advantage and long-term stability.<sup>[2]</sup>

Business intelligence can be used by enterprises to support a wide range of business decisions ranging from operational to strategic. Basic operating decisions include [product positioning](#) or [pricing](#). Strategic business decisions involve priorities, goals, and directions at the broadest level. In all cases, BI is most effective when it combines data derived from the market in which a company operates (external data) with data from company sources internal to the business such as financial and operations data (internal data). When combined, external and internal data can provide a complete picture which, in effect, creates an "intelligence" that cannot be derived from any singular set of data.<sup>[3]</sup> Among myriad uses, business intelligence tools empower organizations to gain insight into new markets, to assess demand and suitability of products and services for different [market segments](#), and to gauge the impact of marketing efforts.<sup>[4]</sup>

BI applications use data gathered from a [data warehouse](#) (DW) or from a [data mart](#), and the concepts of BI and DW combine as "BI/DW"<sup>[5]</sup> or as "BIDW". A data warehouse contains a copy of analytical data that facilitate [decision support](#).

## Business administration

### Management of a business

[Accounting](#) [show]

[Business entities](#) [show]

[Corporate governance](#) [show]

[Corporate law](#) [show]

[Corporate title](#) [show]

[Economics](#) [show]

[Finance](#) [show]

[Types of management](#) [show]

[Organization](#) [show]

[Trade](#) [show]



[Business and economics portal](#)

V · T · E

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

Wikipedia Commons

Contents [hide]

48



# Business analytics

From Wikipedia, the free encyclopedia

**Business analytics (BA)** refers to the skills, technologies, practices for continuous iterative exploration and investigation of past business performance to gain insight and drive business planning.<sup>[1]</sup> Business analytics focuses on developing new insights and understanding of business performance based on data and statistical methods. In contrast, business intelligence traditionally ? focuses on using a consistent set of metrics to both measure past performance and guide business planning, which is also based on data and statistical methods.<sup>[citation needed]</sup>

Business analytics makes extensive use of analytical modeling and numerical analysis, including explanatory and predictive modeling,<sup>[2]</sup> and fact-based management to drive decision making. It is therefore closely related to management science. ?

Analytics may be used as input for human decisions or may drive fully automated decisions. Business intelligence is querying, reporting, online analytical processing (OLAP), and "alerts."

In other words, querying, reporting, OLAP, it is alert tools can answer questions such as what happened, how many, how often, where the problem is, and what actions are needed. Business analytics can answer questions like why is this happening, what if these trends continue, what will happen next (predict), and what is the best outcome that can happen (optimize).<sup>[3]</sup> ?

## Contents [hide]

- 1 Examples of application
- 2 Types of analytics
- 3 Basic domains with applications
- 4 History
- 5 Challenges
- 6 Competing on analytics
- 7 See also
- 8 References
- 9 Further reading

## Types of analytics [edit]

- Decision analytics: supports human decisions with visual analytics that the user models to reflect reasoning.<sup>[5]</sup>
- Descriptive analytics: gains insight from historical data with reporting, scorecards, clustering etc.
- Predictive analytics: employs predictive modelling using statistical and machine learning techniques
- Prescriptive analytics: recommends decisions using optimization, simulation, etc.





WIKIPEDIA  
The Free Encyclopedia

Not logged in Talk Contributions Create account Log in

Article Talk

Read Edit View history

Search Wikipedia



# Predictive analytics

From Wikipedia, the free encyclopedia

**Predictive analytics** encompasses a variety of [statistical](#) techniques from [data mining](#), [predictive modelling](#), and [machine learning](#), that analyze current and historical facts to make [predictions](#) about future or otherwise unknown events.<sup>[1][2]</sup>

In business, predictive models exploit [patterns](#) found in historical and transactional data to identify risks and opportunities. Models capture relationships among many factors to allow assessment of risk or potential associated with a particular set of conditions, guiding [decision-making](#) for candidate transactions.<sup>[3]</sup>

The defining functional effect of these technical approaches is that predictive analytics provides a predictive score (probability) for

each individual (customer, in order to determine, information marketing, credit risk assessment, enforcement.

Predictive analytics is used for travel,<sup>[10]</sup> mobility,<sup>[11]</sup> health, other fields.

One of the best-known applications is a bank's assessment of a customer's [credit history](#), to predict whether the customer will default on future credit payments on time.

## Definition [\[edit\]](#)

Predictive analytics is an area of statistics that deals with [extracting information](#) from data and using it to predict [trends](#) and behavior patterns. The enhancement of predictive web analytics calculates statistical [probabilities](#) of future events online. [Predictive analytics](#) statistical techniques include [data modeling](#), [machine learning](#), [AI](#), [deep learning](#) algorithms and [data mining](#).<sup>[18]</sup> Often the unknown event of interest is in the future, but predictive analytics can be applied to any type of unknown whether it be in the past, present or future. For example, identifying suspects after a crime has been committed, or credit card fraud as it occurs.<sup>[19]</sup> The core of predictive analytics relies on capturing relationships between [explanatory variables](#) and the predicted variables from past occurrences, and exploiting them to predict the unknown outcome. It is important to note, however, that the accuracy and usability of results will depend greatly on the level of data analysis and the quality of assumptions.

Cont.

1 Definition

2 Types

- 2.1 Predictive models
- 2.2 Descriptive models
- 2.3 Decision models

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

Languages





# Gartner

From Wikipedia, the free encyclopedia

**Gartner, Inc.**, officially known as **Gartner**, is a global research and advisory firm providing information, advice, and tools for leaders in IT, finance, HR, customer service and support, communications, legal and compliance, marketing, sales, and supply chain functions.<sup>[3]</sup> Its headquarters are in Stamford, Connecticut, United States. The firm changed its name from **Gartner Group, Inc** to Gartner in 2000.<sup>[4]</sup> It is a member of the S&P 500.<sup>[5]</sup>

Research provided by Gartner has historically targeted CIOs, senior IT, marketing, and supply chain managers. The acquisition of **CEB, Inc.** has expanded its range and breadth of offerings to support every business function in every industry and enterprise size.<sup>[6][3]</sup> Gartner clients include large corporations, government agencies, technology companies, and the investment community. Its client base consists of over 15,000 organizations in over 100 countries.<sup>[3]</sup> The company's products and services include Research, Executive Programs, Consulting, and Conferences. Founded in 1979, Gartner has over 15,000 employees located in 100+ offices worldwide.<sup>[7]</sup>

## Contents [hide]

- 1 History
- 2 See also
- 3 References
- 4 External links

## History [edit]

The company was founded in 1979 by Gideon Gartner. Originally a private

## Gartner, Inc.

# Gartner®



Gartner headquarters in Stamford

Type	Public company
Traded as	NYSE: IT
	S&P 500 component
Industry	Research & Advisory Services
Founded	1979; 41 years ago by Gideon Gartner in Stamford, Connecticut
Headquarters	Stamford, Connecticut <small>[citation needed]</small>
Key people	Gene Hall (CEO) Craig Safian (CFO) Mike Diliberto (CIO)
Products	Research Consulting Conferences
Revenue	▲ US\$4.25 billion (2019) <small>[153]</small>

# Was ist Salesforce?

Salesforce.com

Unternehmen



Salesforce.com ist ein internationaler Anbieter von Cloud-Computing-Lösungen für Unternehmen. Das Unternehmen mit Hauptsitz in San Francisco stellt Geschäftsanwendungen über das Internet zur Verfügung. [Wikipedia](#)

**CEO:** Marc Benioff (Nov. 2001–)

**Kundenservice:** 00 353 1 440 3500

**Aktienkurs:** FOO (FRA) 205,85 € +1,70 (+0,83 %)

29. Okt., 15:11 MEZ - Haftungsausschluss

**Umsatz:** 17,1 Milliarden USD (2020)

**Gründung:** Februar 1999, Kalifornien, Vereinigte Staaten

**Tochtergesellschaften:** Pardot, Tableau Software, Heroku, MEHR

**Gründer:** Marc Benioff, Parker Harris, Halsey Minor

Haftungsausschluss

Figure 1. Magic Quadrant for the CRM Customer Engagement Center



# Was ist Tableau?

## Tableau Software <

Unternehmen



Tableau Software ist ein Hersteller von Visualisierungs-Software aus dem US-amerikanischen Seattle. Schwerpunkt der Software sind Datenvisualisierung und Reporting. Tableau Software wurde 2003 als Ausgründung aus der Stanford University gegründet. Heute nutzen mehr als 39.000 Unternehmen Produkte von Tableau Software. [Wikipedia](#)

**Gründung:** Januar 2003, Mountain View, Kalifornien, Vereinigte Staaten

**Umsatz:** 1,2 Milliarden USD (2018)

**CEO:** Adam Selipsky (16. Sept. 2016–)

**Vertrieb:** 069 589976700

**Dachorganisation:** [Salesforce.com](#)

**Gründer:** Pat Hanrahan, Christian Chabot, Andrew Beers, Chris Stolte

## Salesforce closes \$15.7B Tableau deal

Ron Miller @ron\_miller / 10:44 PM GMT+2 • August 1, 2019

Comment





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

Article Talk

Not logged in Talk Contributions Create account Log in

Read

Edit

View history

Search Wikipedia



# RapidMiner

From Wikipedia, the free encyclopedia

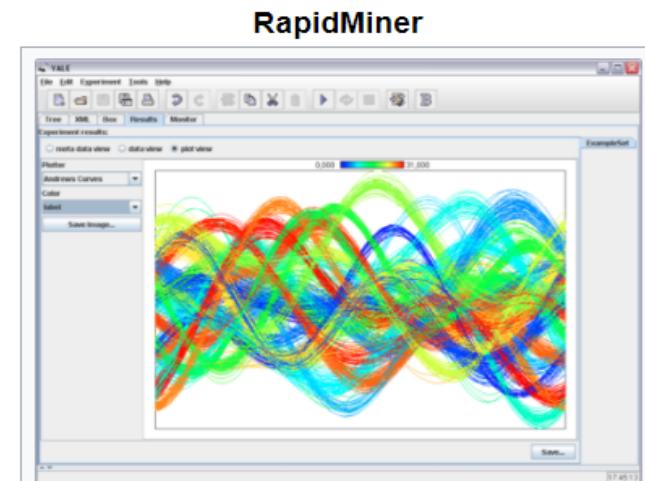
**RapidMiner** is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics. It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, model validation and optimization.<sup>[1]</sup> RapidMiner is developed on an open core model.

## Contents [hide]

- 1 History
- 2 Description
- 3 Products
- 4 Adoption
- 5 Developer
- 6 References
- 7 External links

## History [edit]

RapidMiner, formerly known as YALE (Yet Another Learning Environment), was developed starting in 2001 by Ralf Klinkenberg, Ingo Mierswa, and Simon Fischer at the Artificial Intelligence Unit of



RapidMiner

<b>Developer(s)</b>	RapidMiner
<b>Initial release</b>	2006; 14 years ago
<b>Stable release</b>	9.6 / 2 March 2020; 7 months ago
<b>Operating system</b>	Cross-platform
<b>Type</b>	Data science, machine learning, predictive analytics
<b>License</b>	Professional and Enterprise Editions are Proprietary; Free Edition (10,000 rows and 1 logical processor limit) is available as AGPL
<b>Website</b>	<a href="http://rapidminer.com">rapidminer.com</a>

# OECD, Wikipedia: 21<sup>st</sup> Century Skills

## 21st Century Skills

How today's students can stay competitive  
in a changing job market

### Learning Skills



critical thinking



creativity



collaboration



communication

### Literacy Skills



information



media



technology

### Life Skills



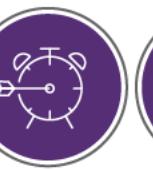
flexibility



leadership



initiative



productivity



social skills

## 21st century skills

From Wikipedia, the free encyclopedia

**21st century skills** comprise **skills**, **abilities**, and learning **dispositions** that have been identified as being required for success in 21st century society and workplaces by educators, business leaders, **academics**, and **governmental agencies**. This is part of a growing international movement focusing on the skills required for students to master in preparation for success in a rapidly changing, **digital society**. Many of these skills are also associated with **deeper learning**, which is based on mastering skills such as **analytic reasoning**, **complex problem solving**, and **teamwork**. These skills differ from traditional academic skills in that they are not primarily content knowledge-based.<sup>[1][2][3]</sup>



KIPEDIA  
e Enzyklopädie

seite

importale

er Artikel

nen

verbessern

Artikel anlegen

portal

änderungen

n

uge

uf diese Seite

ngen an

en Seiten

seiten

enter Link

informationen

titieren

a-Datenobjekt

n/exportieren

stellen

F herunterladen

# Data Science

**Data Science** (von englisch *data* „Daten“ und *science* „Wissenschaft“, im Deutschen auch **Datenwissenschaft**) bezeichnet generell die Extraktion von **Wissen** aus Daten.<sup>[1][2]</sup>

Data Science ist ein interdisziplinäres Wissenschaftsfeld, welches wissenschaftlich fundierte Methoden, Prozesse, Algorithmen und Systeme zur Extraktion von Erkenntnissen, Mustern und Schlüssen sowohl aus strukturierten als auch unstrukturierten Daten ermöglicht<sup>[3][4]</sup>.

Der Studiengang *Data Science* verwendet Techniken und Theorien aus den Fächern Mathematik, **Statistik** und Informationstechnologie, einschließlich der **Signalverarbeitung**, verwendet Wahrscheinlichkeitsmodelle des **maschinellen Lernens**, des statistischen Lernens, der Programmierung, der Datentechnik, der **Mustererkennung**, der **Prognostik**, der Modellierung von Unsicherheiten und der Datenlagerung.

Personen, die im Bereich Data Science arbeiten, werden als **Data Scientist** bzw. **Datenwissenschaftler** bezeichnet, wobei meist speziellere oder Spezialisierungen anderer, übergeordneter Berufsbezeichnungen üblich sind (z. B. Statistiker, Informatiker).

## Inhaltsverzeichnis [Verbergen]

- 1 Geschichte
- 2 Das Berufsfeld
  - 2.1 Anforderungen
  - 2.2 Aufgabenbereich
  - 2.3 Ausbildungsmöglichkeiten
  - 2.4 Branchen
- 3 Siehe auch
- 4 Literatur
- 5 Weblinks



[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](#)

# Data science

From Wikipedia, the free encyclopedia

(Redirected from [Data Science](#))

*Not to be confused with [information science](#).*

! **Data science** is an [inter-disciplinary](#) field that uses scientific methods, processes, algorithms and systems to extract [knowledge](#) and insights from many structural and [unstructured data](#).<sup>[1][2]</sup> Data science is related to [data mining](#), [machine learning](#) and [big data](#).

! Data science is a "concept to unify [statistics](#), [data analysis](#) and their related methods" in order to "understand and analyze actual phenomena" with data.<sup>[3]</sup> It uses techniques and theories drawn from many fields within the context of [mathematics](#), [statistics](#), [computer science](#), [domain knowledge](#) and [information science](#). Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science ([empirical](#), [theoretical](#), [computational](#) and now [data-driven](#)) and asserted that "everything about science is changing because of the impact of information technology" and the [data deluge](#).<sup>[4][5]</sup>

## Contents [hide] !

- 1 Foundations
  - 1.1 Relationship to statistics
- 2 Etymology
  - 2.1 Early usage
  - 2.2 Modern usage
- 3 Careers in data science
  - 3.1 Educational path
  - 3.2 Specializations and associated careers
- 4 Impacts of data science
- 5 Technologies and techniques

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]

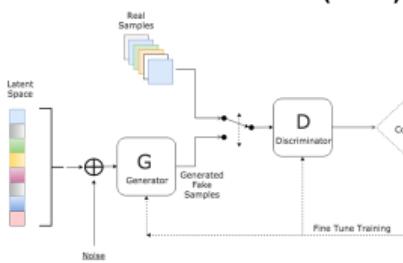
V · T · E

# 8 Machine Learning III

## - Specialized Areas in Machine Learning

### (4) Generative Adversarial Networks GAN

#### Generative Adversarial Networks (GAN)



A Generative Adversarial Network (GAN) is a class of machine learning frameworks designed by Goodfellow and his colleagues in 2014.

Two neural networks contest with each other in a game. Given a training set, this technique learns to generate new data with the same statistics as the training set.

The core idea of a GAN is based on the "indirect" training through the discriminator, which itself is also being updated dynamically. This basically means that the generator is not trained to minimize the distance to a specific image, but rather to fool the discriminator. This enables the model to learn in an unsupervised manner. (Wikipedia)

#### StyleGAN: Motivation Style Transfer



Intuitive Guide to Neural Style Transfer, 2019

[towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-neural-style...](https://towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-neural-style...)

# Fill in the Blank.

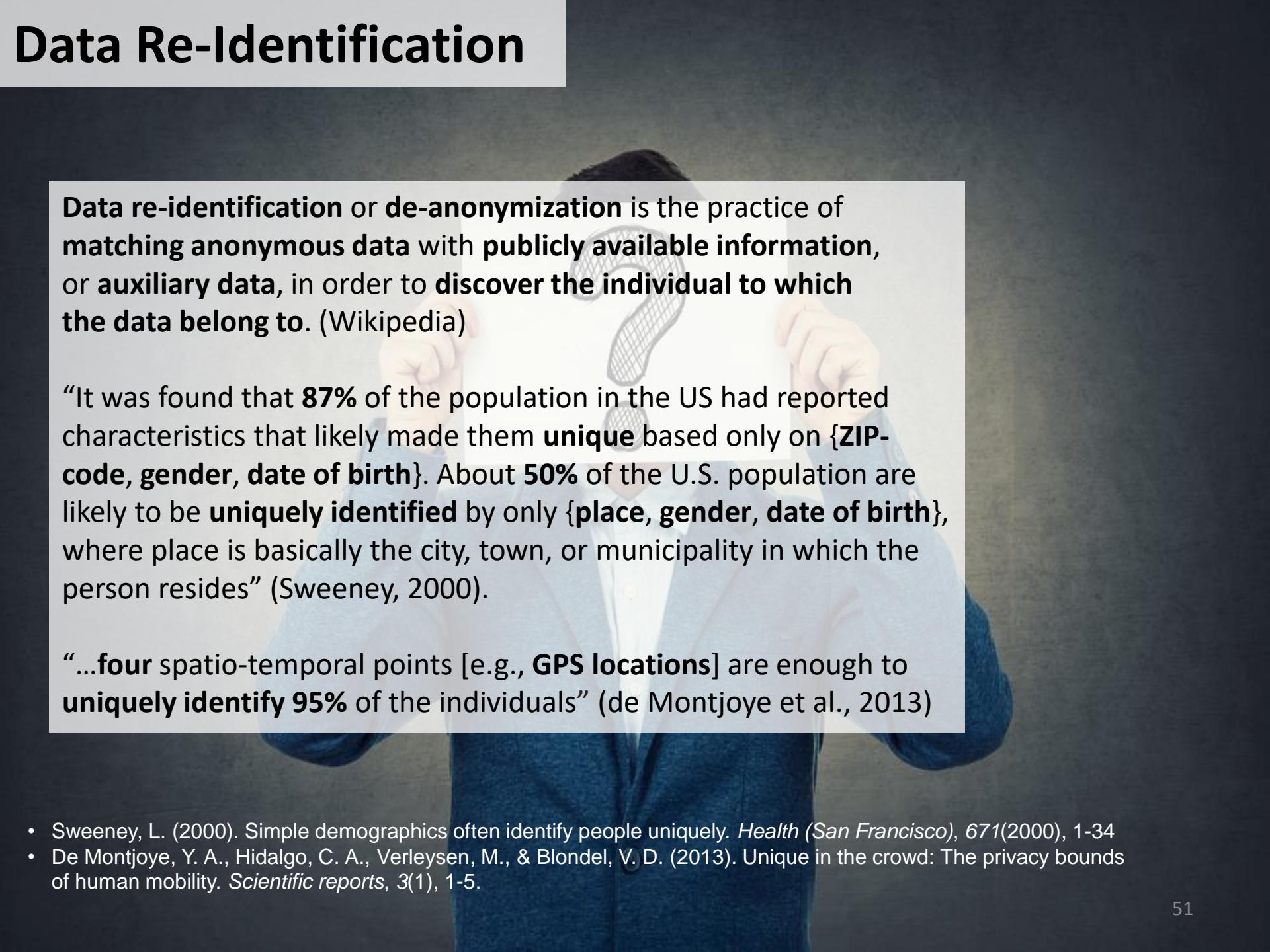
Schwierigkeitsgrad	Art des Wissens	Abfragewissen (Vorlesung)	Anwendungswissen (Literatur)
Einfach			
Mittel			
Schwierig			

1 is an area of machine learning concerned  
with how intelligent agents ought to take actions in 2 in order to  
3 the notion of cumulative 4

1 is one of three basic Machine Learning Paradigms,  
alongside Supervised Learning and Unsupervised Learning. (Wikipedia)

- a) Reinforcement learning | environment | maximize | mistakes
- b) Reinforcement learning | environment | minimize | rewards
- c) Reinforcement learning | process | generate | loss
- d) Reinforcement learning | process | minimize | loss
- e) Reinforcement learning | environment | maximize | rewards

# Data Re-Identification



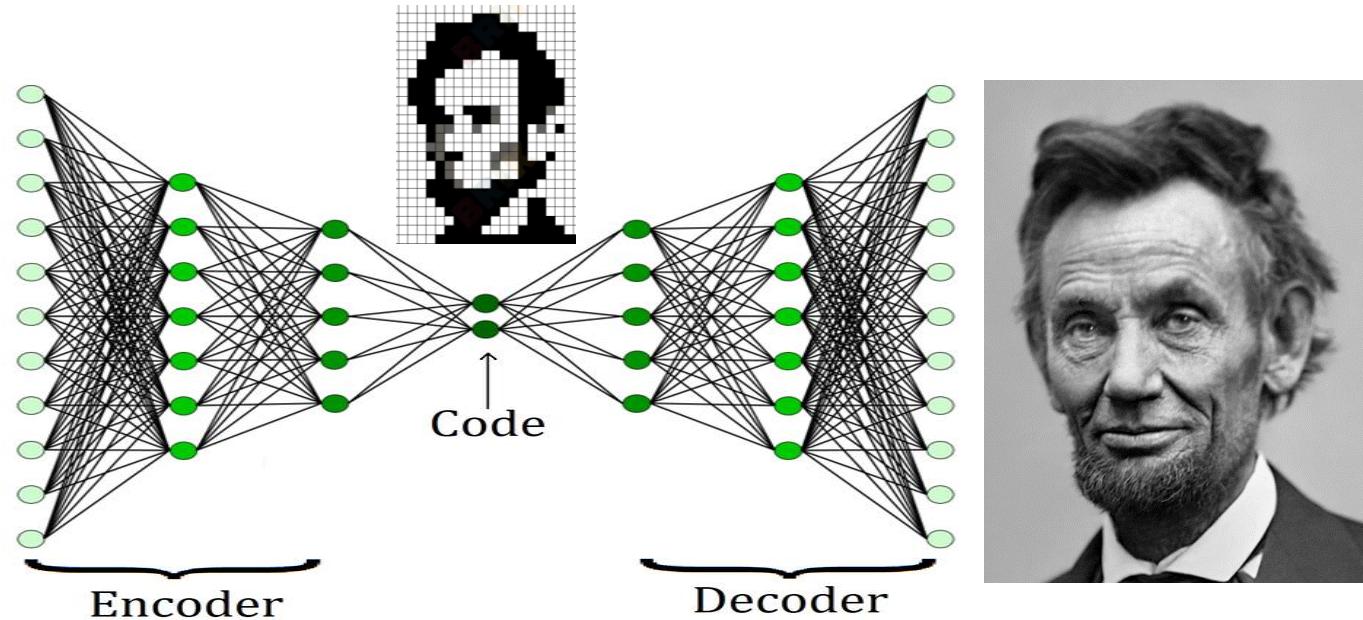
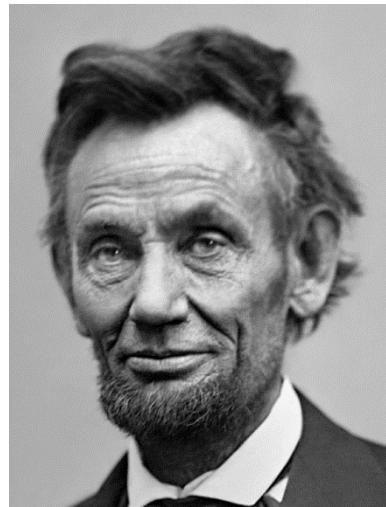
**Data re-identification or de-anonymization** is the practice of matching anonymous data with publicly available information, or auxiliary data, in order to discover the individual to which the data belong to. (Wikipedia)

“It was found that **87%** of the population in the US had reported characteristics that likely made them **unique** based only on {**ZIP-code, gender, date of birth**}. About **50%** of the U.S. population are likely to be **uniquely identified** by only {**place, gender, date of birth**}, where place is basically the city, town, or municipality in which the person resides” (Sweeney, 2000).

“...four spatio-temporal points [e.g., **GPS locations**] are enough to **uniquely identify 95%** of the individuals” (de Montjoye et al., 2013)

- Sweeney, L. (2000). Simple demographics often identify people uniquely. *Health (San Francisco)*, 671(2000), 1-34
- De Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3(1), 1-5.

# Autoencoder



An **Autoencoder** is a type of artificial **neural network** used to **learn efficient data codings** in an **unsupervised manner**.

The aim of an **Autoencoder** is to learn a **representation (encoding)** for a **set of data**, typically for **dimensionality reduction**, by training the network to **ignore signal “noise”**.

Along with the **reduction side (Encoder)**, a **reconstructing side (Decoder)** is learnt, where the **Autoencoder** tries to **generate** from the **reduced encoding** a **representation** as **close as possible** to its **original input**, hence its name. ([Wikipedia](#))

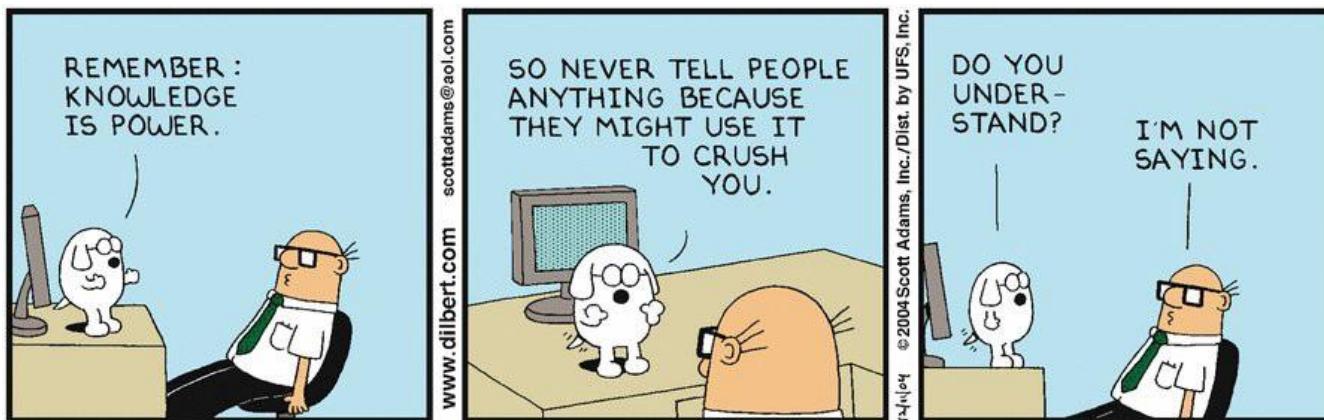
# Transfer Learning

## Transfer learning

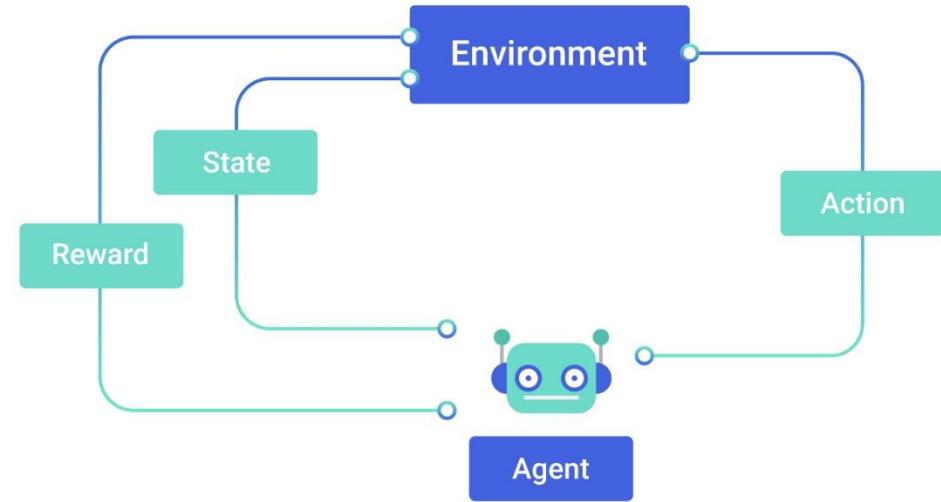
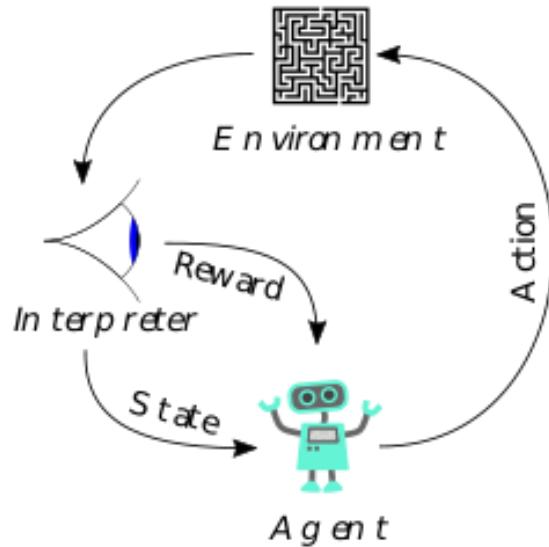
From Wikipedia, the free encyclopedia

**Transfer learning (TL)** is a research problem in machine learning (ML) that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem.<sup>[1]</sup> For example, knowledge gained while learning to recognize cars could apply when trying to recognize trucks.

This area of research bears some relation to the long history of psychological literature on [transfer of learning](#), although formal ties between the two fields are limited. From the practical standpoint, reusing or transferring information from previously learned tasks for the learning of new tasks has the potential to significantly improve the sample efficiency of a [reinforcement learning](#) agent.<sup>[2]</sup>



# Reinforcement Learning (RL)



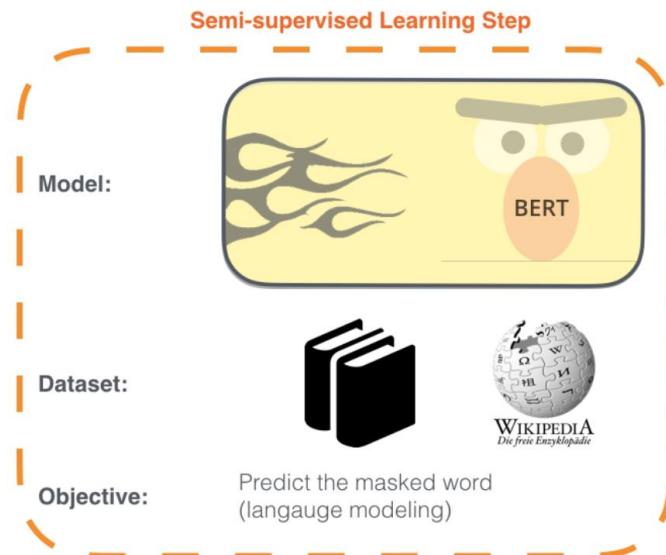
**Reinforcement Learning (RL)** is an area of machine learning concerned with **how intelligent agents** ought to **take actions** in an **environment** in order to **maximize** the notion of cumulative **reward** (**learn from mistakes / experience**).

**Reinforcement Learning** is one of three basic Machine Learning Paradigms, alongside **Supervised Learning** and **Unsupervised Learning**. (Wikipedia)

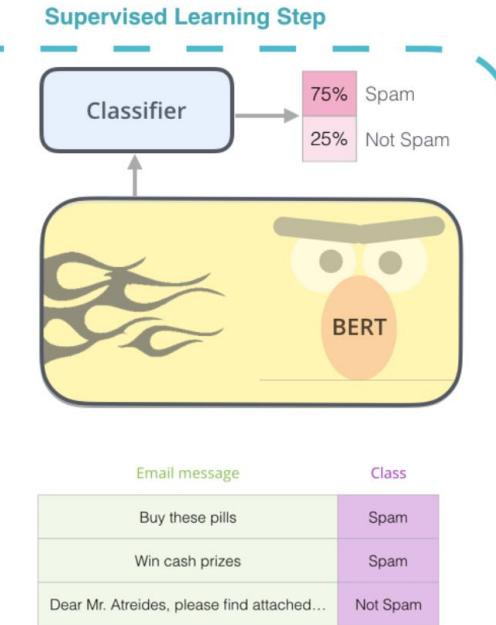
# Transfer Learning with BERT Transformer

1 - **Semi-supervised** training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.



2 - **Supervised** training on a specific task with a labeled dataset.



The two steps of how BERT is developed. You can download the model pre-trained in step 1 (trained on un-annotated data), and only worry about fine-tuning it for step 2. [Source for book icon].

Eine Form von **Transfer Learning**:  
Fokus auf **Fine-Tuning, Pre-Training** übernehmen

# Cognitive Biases (Kognitive Verzerrungen)

A photograph of a vast, calm ocean under a clear, light blue sky. The water is a deep turquoise-blue, with small white caps visible on the surface. The horizon is straight and level, meeting the sky at the center of the frame.

Eine kognitive Verzerrung ist ein kognitionspsychologischer Sammelbegriff für systematische fehlerhafte Neigungen beim Wahrnehmen, Erinnern, Denken und Urteilen. Sie bleiben meist unbewusst und basieren auf kognitiven Heuristiken. (Wikipedia)

# Data Governance raises data quality.

From Wikipedia, the free encyclopedia

**Data governance** is a term used on both a macro and a micro level. The former is a political concept and forms part of international relations and Internet governance; the latter is a data management concept and forms part of corporate data governance.

Part of a series on

## Governance

### Models

[Bad](#) · [Collaborative](#) · [Good](#) · [Multistakeholder](#)

### Micro level [edit]

Here the focus is on an individual company. Here **data governance** is a **data management** concept concerning the capability that enables an organization to ensure that high **data quality** exists throughout the complete lifecycle of the data, and data controls are implemented that support business objectives. The key focus areas of data governance include availability, usability, consistency,<sup>[2]</sup> **data integrity** and **data security** and includes establishing processes to ensure effective data management throughout the enterprise such as accountability for the adverse effects of poor data quality and ensuring that the data which an enterprise has can be used by the entire organization.

A **data steward** is a role that ensures that data governance processes are followed and that guidelines enforced, as well as recommending improvements to data governance processes.

Data governance encompasses the people, processes, and **information technology** required to create a consistent and proper handling of an organization's data across the **business enterprise**. It provides all data management practices with the necessary foundation, strategy, and structure needed to ensure that data is managed as an asset and transformed into meaningful information.<sup>[3]</sup> Goals may be defined at all levels of the enterprise and doing so may aid in acceptance of processes by those who will use them. Some goals include

- Increasing consistency and confidence in **decision making**
- Decreasing the risk of regulatory fines
- Improving **data security**, also defining and verifying the requirements for data distribution policies<sup>[4]</sup>
- Maximizing the income generation potential of data
- Designating accountability for information quality
- Enable better planning by supervisory staff
- Minimizing or eliminating re-work
- Optimize staff effectiveness
- Establish process performance baselines to enable improvement efforts
- Acknowledge and hold all gain



# General Data Protection Regulation

From Wikipedia, the free encyclopedia

The **General Data Protection Regulation (GDPR)** is a regulation in EU law on data protection and privacy in the European Union (EU) and the European Economic Area (EEA). It also addresses the transfer of personal data outside the EU and EEA areas. The GDPR's primary aim is to give control to individuals over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU.<sup>[1]</sup> Superseding the Data Protection Directive 95/46/EC, the regulation contains provisions and requirements related to the processing of personal data of individuals (formally called *data subjects* in the GDPR) who are located in the EEA, and applies to any enterprise—regardless of its location and the data subjects' citizenship or residence—that is processing the personal information of individuals inside the EEA.

Controllers and processors of personal data must put in place appropriate technical and organizational measures to implement the data protection principles. Business processes that handle personal data must be designed and built with consideration of the principles and provide safeguards to protect data (for example, using pseudonymization or full anonymization where appropriate). Data controllers must design information systems with privacy in mind. For instance, using the highest-possible privacy settings by default, so that the datasets are not publicly available by default and cannot be used to identify a subject. No personal data may be processed unless this processing is done under one of the six lawful bases specified by the regulation (consent, contract, public task, vital interest, legitimate interest or legal requirement). When the processing is based on consent the data subject has the right to revoke it at any time.

Data controllers must clearly disclose any data collection, declare the lawful basis and purpose for data processing, and state how long data is being retained and if it is being shared with any third parties or outside of the EEA. Firms have the obligation to protect data of employees and consumers to the degree where only the necessary data is extracted with minimum interference with data privacy from employees, consumers, or third parties. Firms should have internal controls and regulations for various departments such as audit, internal controls, and operations. Data subjects have the right to request a portable copy of the data collected by a controller in a common format, and the right to have their data erased under certain circumstances. Public authorities, and businesses whose core activities consist of regular or systematic processing of personal data, are required to employ a data protection officer (DPO), who is responsible for managing compliance with the GDPR. Businesses must report data breaches to national supervisory authorities within 72 hours if they have an adverse effect on user privacy. In some cases, violators of the GDPR may be

(EU) 2016/679



## Title

Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (Data Protection Directive)

## Made by

European Parliament and Council of the European Union

## Journal reference

L 100000 May 2016, p. 1–88

## History

Date made 14 April 2016

Implementation 25 May 2018 date

## Preparative texts

Commission COM/2012/010 final – 2012/0010 proposal (COD)

## Other legislation

Replaces Data Protection Directive

## Current legislation



# California Consumer Privacy Act

From Wikipedia, the free encyclopedia

The **California Consumer Privacy Act (CCPA)** is a state statute intended to enhance privacy rights and consumer protection for residents of California, United States. The bill was passed by the California State Legislature and signed into law by Jerry Brown, Governor of California, on June 28, 2018, to amend Part 4 of Division 3 of the California Civil Code.<sup>[2]</sup> Officially called **AB-375**, the act was introduced by Ed Chau, member of the California State Assembly, and State Senator Robert Hertzberg.<sup>[3][4]</sup>

Amendments to the CCPA, in the form of Senate Bill 1121, were passed on September 13, 2018.<sup>[5][6]</sup> Additional substantive amendments were signed into law on October 11, 2019.<sup>[7]</sup> The CCPA became effective on January 1, 2020.<sup>[8]</sup>

<b>Contents</b> [hide]
1 Intentions of the Act
1.1 Compliance
1.2 Responsibility and accountability
1.3 Sanctions and remedies
2 Definition of personal data
3 History
4 See also
5 References
6 Further reading
7 External links

## California Consumer Privacy Act



California State Legislature

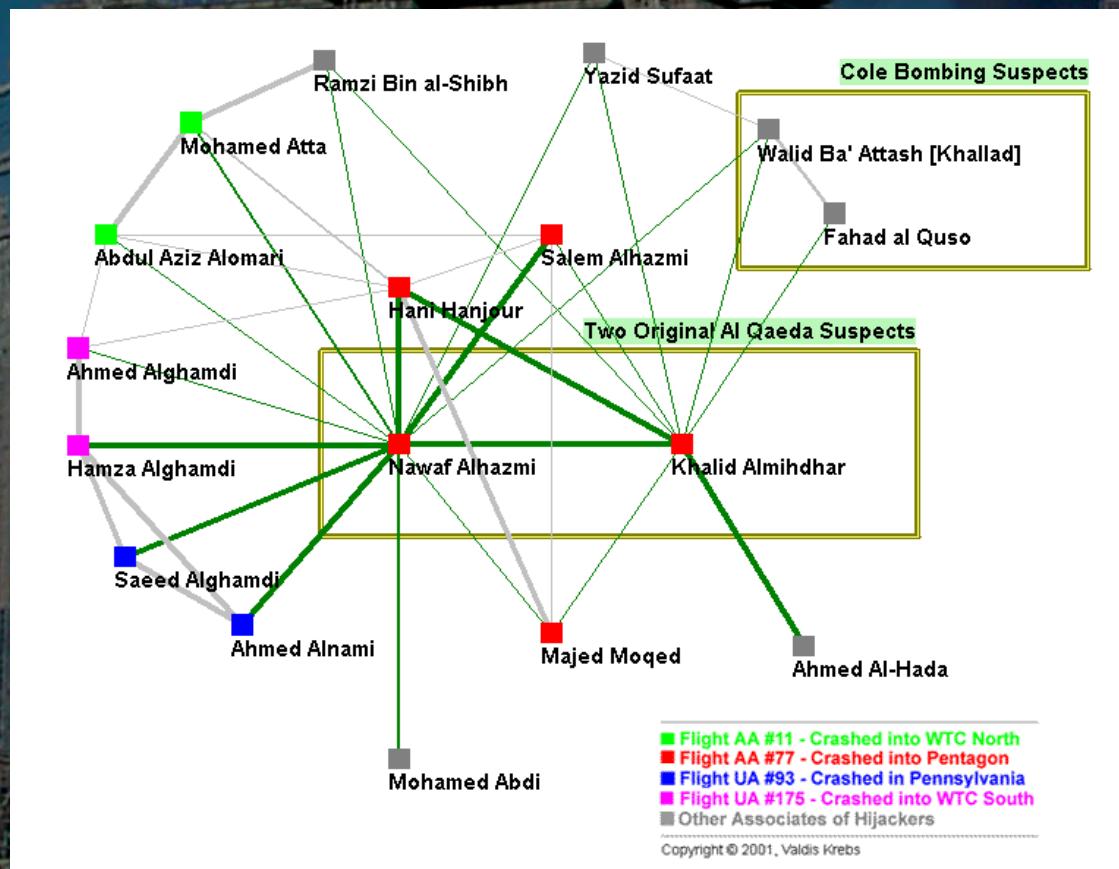
Full name	California Consumer Privacy Act of 2018 <sup>[1]</sup>
Status	Passed
Introduced	January 3, 2018
Signed into law	June 28, 2018
Governor	Jerry Brown
Code	California Civil Code
Section	1798.100
Resolution	AB-375 (2017–2018 Session)
Website	Assembly Bill No. 375 <sup>[2]</sup>

## Intentions of the Act [edit]

The intentions of the Act are to provide California residents with the right to:

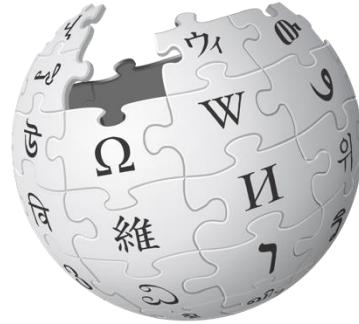
1. Know what **personal data** is being collected about them.
2. Know whether their **personal data** is sold or disclosed and to whom.
3. Say no to the sale of **personal data**.
4. Access their **personal data**.
5. Request a business to delete any **personal information** about a consumer collected from that consumer.<sup>[9]</sup>
6. Not be discriminated against for exercising their **privacy rights**.

# Die Attentäter vom 9. September 2001 standen alle in direktem Kontakt mit den Verdächtigen.



Name	Age	Nationality	Flight
Mohamed Atta	33	Egypt	American Airlines Flight 11
Abdulaziz al-Omari	22	Saudi Arabia	
Waleed al-Shehri	28		
Satam al-Suqami	25		
Marwan al-Shehhi	23	UAE	United Airlines Flight 175
Fayez Banihad	24		
Mohand al-Shehri	22		
Hamza al-Ghamdi	20		
Ahmed al-Ghamdi	22		American Airlines Flight 77
Hani Hanjour	29	Saudi Arabia	
Khalid al-Mihdhar	26		
Majed Moqed	24		
Nawaf al-Hazmi	25		United Airlines Flight 93
Salem al-Hazmi	20		
Ziad Jarrah	26	Lebanon	
Ahmed al-Haznawi	20	Saudi Arabia	
Ahmed al-Nami	24		
Saeed al-Ghamdi	21		

# Was sagt Wikipedia über Shitstorms?



**Shitstorm** ... bezeichnet das **lawinenartige Auftreten negativer Kritik** gegen Personen oder Unternehmen im Rahmen von sozialen Netzwerken, Blogs oder Kommentarfunktionen von Internetseiten bis hin zur Schmähkritik.

... Dabei richtet sich „in kurzem Zeitraum eine **subjektiv große Anzahl von kritischen Äußerungen** ..., von denen sich zumindest ein Teil vom ursprünglichen Thema ablöst und [die] stattdessen **aggressiv, beleidigend, bedrohend oder anders attackierend** geführt [werden]“ gegen [alle und jeden].

Eine Rube-Goldberg-Maschine ist eine **Nonsense-Maschine**, die eine Aufgabe in zahlreiche unnötige und komplizierte Einzelschritte zerlegt. (Wikipedia)



Tapscott 2016, Blockchain Revolution:

„**Financial transactions** are comparable to **rube goldberg machines** – deliberately over-engineered but performing just a simple task.“

# Data Integrity (Wikipedia)

## Data integrity

---

From Wikipedia, the free encyclopedia

**Data integrity** is the maintenance of, and the assurance of, the accuracy and consistency of data over its entire life-cycle,<sup>[1]</sup> and is a critical aspect to the design, implementation and usage of any system which stores, processes, or retrieves data. The term is broad in scope and may have widely different meanings depending on the specific context – even under the same general umbrella of [computing](#). It is at times used as a proxy term for [data quality](#),<sup>[2]</sup> while [data validation](#) is a pre-requisite for data integrity.<sup>[3]</sup> Data integrity is the opposite of [data corruption](#).<sup>[4]</sup> The overall intent of any data integrity technique is the same: ensure data is recorded exactly as intended (such as a database correctly rejecting mutually exclusive possibilities), and upon later [retrieval](#), ensure the data is the same as it was when it was originally recorded. In short, data integrity aims to prevent unintentional changes to information. Data integrity is not to be confused with [data security](#), the discipline of protecting data from unauthorized parties.

Any unintended changes to data as the result of a storage, retrieval or processing operation, including malicious intent, unexpected hardware failure, and [human error](#), is failure of data integrity. If the changes are the result of unauthorized access, it may also be a failure of data security. Depending on the data involved this could manifest itself as benign as a single pixel in an image appearing a different color than was originally recorded, to the loss of vacation pictures or a business-critical database, to even catastrophic loss of human life in a [life-critical system](#).

## Proof of work

From Wikipedia, the free encyclopedia



This article may require [cleanup](#) to meet Wikipedia's [quality standards](#). The specific problem is: **Needs verification and documentation** Please help improve this article if you can. (May 2015) ([Learn how and when to remove this template message](#))

**Proof of work (PoW)** is a form of [cryptographic zero-knowledge proof](#) in which one party (the *prover*) proves to others (the *verifiers*) that a certain amount of computational effort has been expended for some purpose. Verifiers can subsequently confirm this expenditure with minimal effort on their part. The concept was invented by [Cynthia Dwork](#) and [Moni Naor](#) in 1993 as a way to deter [denial-of-service attacks](#) and other service abuses such as [spam](#) on a network by requiring some work from a service requester, usually meaning processing time by a computer. The term "proof of work" was first coined and formalized in a 1999 paper by [Markus Jakobsson](#) and [Ari Juels](#).<sup>[1][2]</sup> Proof of work was later popularized by [Bitcoin](#) as a foundation for [consensus](#) in permissionless [blockchains](#) and [cryptocurrencies](#), in which miners compete to append blocks and mint new currency, each miner experiencing a success probability proportional to the amount of computational effort they have provably expended. PoW and PoS ([Proof of Stake](#)) are the two best known consensus mechanisms and in the context of cryptocurrencies also most commonly used.<sup>[3]</sup>

## Smart contract



From Wikipedia, the free encyclopedia



This article has multiple issues. Please help [improve it](#) or discuss these [\[hide\]](#) issues on the [talk page](#). (*Learn how and when to remove these template messages*)

- This article **possibly contains original research.** (December 2016)
- This article's **factual accuracy is disputed.** (August 2019)

A **smart contract** is a [computer program](#) or a [transaction protocol](#) which is intended to automatically execute, control or document legally relevant events and actions according to the terms of a [contract](#) or an agreement.<sup>[1][2][3][4]</sup> The objectives of smart contracts are the reduction of need in trusted intermediaries, arbitrations and enforcement costs, fraud losses, as well as the reduction of malicious and accidental exceptions.<sup>[5][2]</sup>

Vending machines are mentioned as the oldest piece of technology equivalent to smart contract implementation.<sup>[3]</sup> 2014's [white paper](#) about the [cryptocurrency Ethereum](#)<sup>[6]</sup> describes the [Bitcoin protocol](#) as a weak version of the smart contract concept as defined by computer scientist, lawyer and cryptographer [Nick Szabo](#). Since Ethereum, various cryptocurrencies support scripting languages which allow for more advanced smart contracts between untrusted parties.<sup>[7]</sup> Smart contracts should be distinguished from [smart legal contracts](#). The latter refers to a traditional natural language legally-binding agreement which has certain terms expressed and implemented in machine-readable code.<sup>[8][9][10]</sup>

# Benfordsches Gesetz (Wikipedia)

Das **Benfordsche Gesetz**, auch **Newcomb-Benford's Law** (NBL), beschreibt eine **Gesetzmäßigkeit** in der **Verteilung** der **führenden Ziffern** von **Zahlen** in **empirischen Datensätzen**.

Je niedriger der **zahlenmäßige Wert** einer **Ziffernsequenz** bestimmter Länge an einer bestimmten Stelle einer Zahl ist, desto **wahrscheinlicher** ist ihr **Auftreten**.

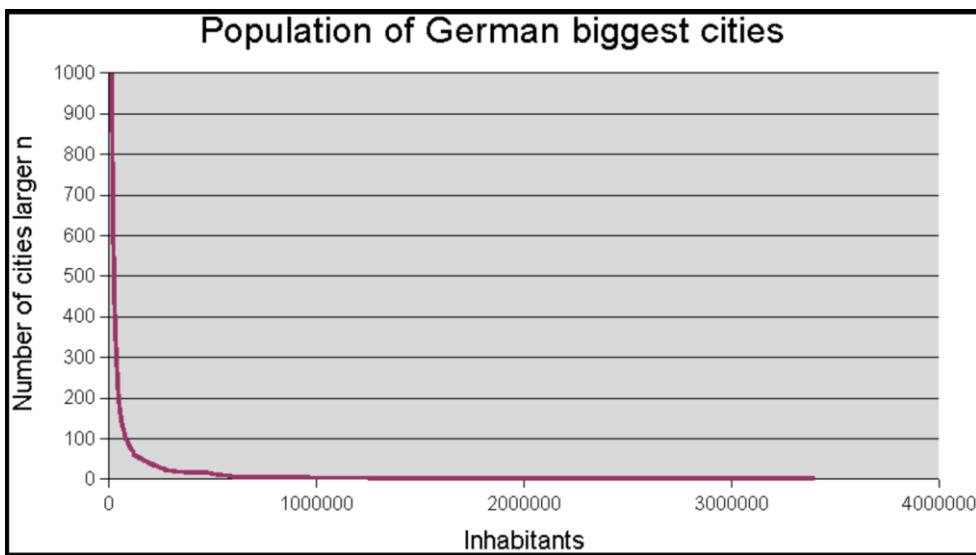
Für Zahlen des Zehnersystems z.B.:

**Zahlen** mit der **Anfangsziffer 1** treten etwa **6,6-mal** so häufig auf wie Zahlen mit der **Anfangsziffer 9**.

Führende Ziffer	Wahrscheinlichkeit
1	30,1 %
2	17,6 %
3	12,5 %
4	9,7 %
5	7,9 %
6	6,7 %
7	5,8 %
8	5,1 %
9	4,6 %

# Benfordsches Gesetz (Wikipedia)

## Am Beispiel der 998 größten Städte Deutschlands



Ziffer	Gemessen	Erwartet
1	340	300,4
2	320	175,7
3	133	124,7
4	87	96,7
5	50	79,0
6	24	66,8
7	20	57,9
8	12	51,1
9	12	45,7

# Welches **Unicorn** wird derzeit laut Wikipedia am höchsten bewertet?

- 
- A photograph showing several students in a classroom setting, focused on writing in their notebooks. The student in the foreground has long blonde hair and is wearing a patterned scarf.
- a) SpaceX
  - b) DiDi
  - c) Bytedance
  - d) Grab
  - e) Ant Financial

Schwierigkeitsgrad	Art des Wissens	
	Abfragewissen (Vorlesung)	Anwendungswissen (Literatur)
Einfach	Green	Yellow
Mittel	Yellow	Red
Schwierig	Red	Red

# List of unicorn startup companies

From Wikipedia, the free encyclopedia



Company	Valuation (US \$billion)	Valuation date	Industry	Country/region
Ant Financial	150	April 2018 <sup>[12]</sup>	Technology	China
Bytedance	100	May 2020 <sup>[13]</sup>	Internet	China
DiDi	51.6	December 2018 <sup>[14]</sup>	Transportation	China
SpaceX	46	August 2020 <sup>[15]</sup>	Aerospace	US
Stripe	36.0	April 2020 <sup>[16]</sup>	Financial services	US
Lufax	30.0	June 2018 <sup>[17]</sup>	Personal finance	China
JUUL Labs	24.0	October 2019 <sup>[18]</sup>	Electronic Cigarettes	US

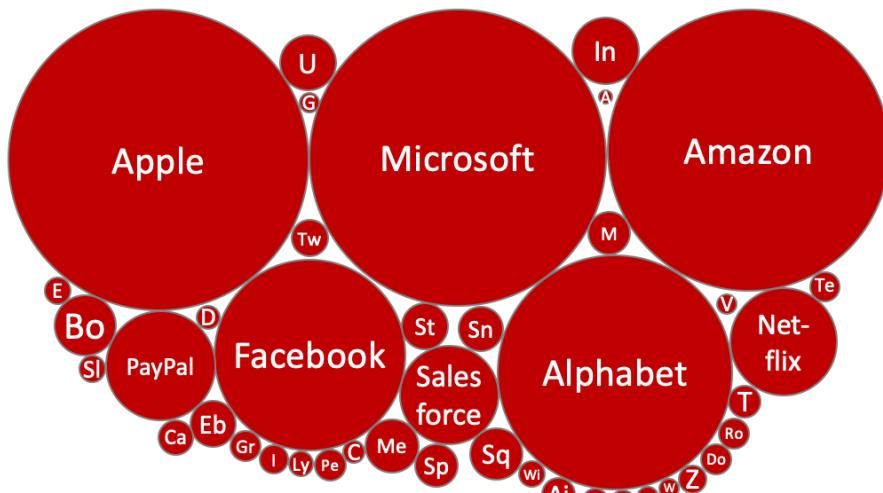
# Unicorns	Total valuation	Date	Reported by
82		Late January 2015	<i>Forbes</i> , CB Insights <sup>[8]</sup>
229	US\$1.3 trillion	January 2016	<i>VentureBeat</i> <sup>[9]</sup>
208	US\$761 billion	December 2016	<i>TechCrunch</i> <sup>[6]</sup>
224	US\$771.9 billion	April 2017	<i>TechCrunch</i>
193	US\$665 billion	April 2017	<i>Inc magazine</i> <sup>[10]</sup>
266	US\$861 billion	August 2018	CB Insights <sup>[4]</sup>
394	US\$1.2 trillion	August 2019	CB Insights <sup>[11]</sup>
495	US\$1.566 trillion	November 2020	CB Insights <sup>[11]</sup>

# Plattform-Ökonomie: Top 100

## Top 100 Plattformen der Welt (Juni 2020)

Börsenwert / Bewertung jüngste bekannte Finanzierung |  
Gesamtwert 10,8 Bio. Dollar (Stand Juni 2020)

### Amerika



Airbnb	Ebay	Lyft	Roku	Teladoc
Alteryx	Etsy	Match	Slack	Twilio
Booking	Expedia	MercadoLibre	Snap	Twitter
Carvana	Grainger	Uber	Splunk	Vroom
Chegg	Grubhub	Peloton	Square	Wikipedia
DoorDash	Instacart	Pinterest	Stripe	Wish
Dropbox	Intuit			Zillow

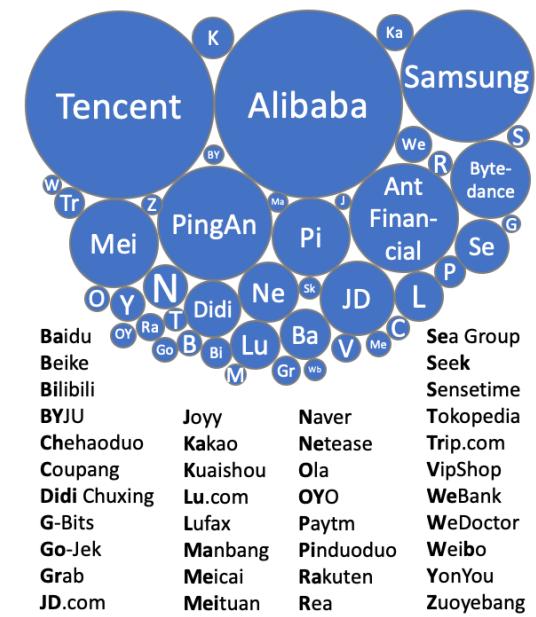
### Europa



### Afrika



### Asien-Pazifik



Anteile in Prozent



© Copyright 2020 Dr. Holger Schmidt | Hamidreza Hosseini | Netzoekonom.de | TU Darmstadt | Ecodynamics.io | Plattform-Index.com



- Erscheinungsjahr: **2001**
- Medientyp: **digital**
- Anzahl Sprachen: **301**
- **302.456 aktive Editoren**
- ca. **5.746.871 Artikel im englischen Wikipedia (ca. 40 Mio. insgesamt)**
- **nicht kommerziell**

# Welche Enzyklopädie ist objektiver?

Is Wikipedia Biased?

By SHANE GREENSTEIN AND FENG ZHU\*

\* Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL, 60208, greenstein@kellogg.northwestern.edu and Marshall School of Business, University of Southern California, Los Angeles, CA, 90089-0808, fzhu@marshall.usc.edu. We thank Megan Busse, Michelle Deveraux, Gil Penchina, Scott Stern, Monica Sun, Joel Waldfogel and many seminar participants for their comments. We are responsible for all remaining errors.

This achievement is astonishing in light of the resources deployed. Wikipedia achieved its size and high profile with minimal staff. Wikipedia is part of a not-for-profit organization. Donations entirely fund the

- Studie 2014 zur **Objektivität englischsprachiger Enzyklopädie-Artikel über politischen Themen.**
- Vergleich von **Encyclopedia Britannica und Wikipedia.**
- **Wikipedia Artikel sind subjektiver**, wobei die Subjektivität mit der **Anzahl von Editoren** abnimmt.

# Statistische Auswertungen der Inhalte von Wikipedia finden Sie unter „Wikipedia:Statistics“



WIKIPEDIA  
The Free Encyclopedia

Project page Talk

Read Edit View history

Not logged in Talk Contributions Create account Log in

Search Wikipedia



## Wikipedia:Statistics

From Wikipedia, the free encyclopedia

"WP:ST" redirects here. You may also be looking for the [list of stub templates](#) or [WikiProject Star Trek](#).

"WP:STAT" and "WP:STATS" redirect here. For the policy regarding use of statistics in articles, see [WP:NOTSTATSBOOK](#). For article readership data, see [Wikipedia:Pageview statistics](#). For mathematical science WikiProject, see [Wikipedia:WikiProject Statistics](#).



This is an information page.

It describes the editing community's established practice on some aspect or aspects of Wikipedia's norms and customs. It is not one of Wikipedia's policies or guidelines, as it has not been thoroughly vetted by the community.

Shortcuts  
[WP:ST](#)  
[WP:STAT](#)  
[WP:STATS](#)

Main page

Contents

Featured content

Current events

Random article

Donate to Wikipedia

Wikipedia store

Interaction

Help

About Wikipedia

Community portal

Recent changes

Contact page

Tools

What links here

Related changes

Upload file

Special page

Permanent link

Page information

Wikidata item

Print/export

Create a book

Download as PDF

Printable version

In other projects

Wikimedia Commons

Wikibooks

Wikinews

Wikiversity

Wikivoyage

Wiktionary

Languages



• Ein Wikipedia Artikel wird über 1.7 edits per second, performed by editors from all over the world.

Currently, the English Wikipedia includes 5,746,451 articles and it averages 559 new articles per day. This amount of data can be analysed in a huge number of ways. The best way to get an idea of the bigger picture is with [statistics](#).

im Schnitt 150 Mal editiert und von 43 Editoren (Users) erstellt.

Contents [hide]

1 Statistics

1.1 General

1.1.2 Images

1.1.4 Page views

1.1.5 Deletion and vandalism statistics

2 Frameworks

3 Datasets

4 Wikistats

5 Archived analysis

6 See also

8 References

9 External links

• Ca. 1000 Artikel werden jeden

Tag von den Admins gelöscht.

• Das Rating für den jeweiligen

Artikel kann in dem Reiter

„Talk“ eingesehen werden.

### English Wikipedia (update)

Articles	5,746,451
Pages	46,278,906
Files	880,100
Edits	862,777,011
Users	34,885,971
Admins	1,202
Active users <sup>[note 1]</sup>	133,506

See more

### WikiStats



Main

General info  
Status statistics  
WikiStats project  
WikiProject edit counters

#### General statistics

Awareness  
Editors  
Total size  
Size comparisons  
Growth  
Records  
Zipf's law comparison

#### Breakdowns

Editors by number of edits  
Editors by article count  
Time between edits  
Administrator actions

V · T · E

# Was Wikipedia sagt...

- „Im Endspurt verlieren die Schlagleute, aufgrund der Erschöpfung, hin und wieder die Kontrolle über ihr Boot und fahren aus ihrer Bahn heraus.“
- „Auf Regatten feuern Steuermanner ihre Mannschaft während des Rennens an, geben Kommandos für Zwischenspurs und liefern Informationen zum Renngeschehen. “
- „Von einem Steuermann kann ... ein ganzes Rennen abhängen. Gute Steuermanner sind daher sehr gefragt und werden bei einem Sieg auch mit einer Medaille ausgezeichnet.“

# Weltbestzeiten im Rudern

Boots-klasse	Beschreibung		Weltbestzeit	Ort	Jahr	Nation
M1x	Einer (Skiff) 1 Ruderer		06:30,74	Polen	2017	NZL
M2x	Zweier 2 Ruderer		05:59,72	Holland	2014	HRV

Quelle: Wikipedia, Distanz 2000 Meter



# Weltbestzeiten im Rudern

Boots-klasse	Beschreibung		Weltbestzeit	Ort	Jahr	Nation
M1x	Einer (Skiff) 1 Ruderer		06:30,74	Polen	2017	NZL
M2x	Zweier 2 Ruderer		05:59,72	Holland	2014	HRV
M4-	Vierer 4 Ruderer		05:37,86	Schweiz	2012	UK
M4+	Vierer 4 Ruderer 1 Steuermann		05:58,96	Wien	1991	GER

Quelle: Wikipedia, Distanz 2000 Meter



# Weltbestzeiten im Rudern

Boots-klasse	Beschreibung		Weltbestzeit	Ort	Jahr	Nation
M1x	Einer (Skiff) 1 Ruderer		06:30,74	Polen	2017	NZL
M2x	Zweier 2 Ruderer		05:59,72	Holland	2014	HRV
M4-	Vierer 4 Ruderer		05:37,86	Schweiz	2012	UK
M4+	Vierer 4 Ruderer 1 Steuermann		05:58,96	Wien	1991	GER
M8+	Achter 8 Ruderer 1 Steuermann		05:18,68	Polen	2017	GER

# Wikipedia gibt einen sehr guten Überblick

The screenshot shows a web browser window with the URL [de.wikipedia.org/wiki/Kanban\\_\(Softwareentwicklung\)](https://de.wikipedia.org/wiki/Kanban_(Softwareentwicklung)). The page title is "Kanban (Softwareentwicklung)". The main content area describes Kanban as a method in software development where the number of parallel tasks is limited to achieve shorter lead times and identify bottlenecks. A sidebar on the left contains links to various Wikipedia categories and tools.

**Inhaltsverzeichnis [Verbergen]**

- 1 Wurzeln und Geschichte
- 2 Kanban-Prinzipien und -Praktiken
  - 2.1 Grundprinzipien
  - 2.2 Kernpraktiken
- 3 Kanban Flight Levels – Wie passt Kanban ins Unternehmen?
- 4 Verhältnis zu anderen Vorgehensweisen in der Softwareentwicklung
  - 4.1 Wasserfall
  - 4.2 Scrum
    - 4.2.1 Gemeinsamkeiten von Kanban und Scrum
    - 4.2.2 Unterschiede zwischen Kanban und Scrum
- 5 Kaizen
- 6 Value, Flow und Waste Elimination
- 7 Priorisierung
- 8 Service Level Agreements (SLA)
- 9 Anwendungsbereiche
- 10 Varianzen
- 11 Tracking
- 12 Literatur
- 13 Weblinks

Laut Wikipedia versteht man unter „**Project Management**“:

“**Project management** is the **practice** of **initiating, planning, executing, controlling, and closing** the work of a **team** to achieve **specific goals** and meet **specific success criteria** at the **specified time**.

A **project** is a **temporary endeavor** designed to produce a unique **product, service or result** with a **defined beginning** and **end** (usually **time-constrained**, and often constrained by **funding or staffing**) undertaken to meet **unique goals** and **objectives**.

The **temporary nature** of projects stands in **contrast** with **business as usual** (or operations), which are repetitive, permanent, or semi-permanent functional activities...”

## Emerging technologies

From Wikipedia, the free encyclopedia

*For specific emerging technologies, see the [List of emerging technologies](#)*

**Emerging technologies** are [technologies](#) whose development, practical applications, or both are still largely unrealized, such that they are figuratively [emerging](#) into prominence from a background of nonexistence or obscurity. These technologies are [new](#), such as various applications of biotechnology including [gene therapy](#) (which date to circa 1990 but even today have large undeveloped potential). Emerging technologies are often perceived as capable of changing the [status quo](#).

Emerging technologies are characterized by radical novelty (in application even if not in origins), relatively fast growth, coherence, prominent impact, and uncertainty and ambiguity. In other words, an emerging technology can be defined as "a radically novel and relatively fast growing technology characterised by a certain degree of coherence persisting over time and with the potential to exert a considerable impact on the socio-economic domain(s) which is observed in terms of the composition of actors, institutions and patterns of interactions among those, along with the associated knowledge production processes. Its most prominent impact, however, lies in the future and so in the emergence phase is still somewhat uncertain and ambiguous."<sup>[1]</sup>

Emerging technologies include a variety of technologies such as [educational technology](#), [information technology](#), [nanotechnology](#), [biotechnology](#), [cognitive science](#), [psychotechnology](#), [robotics](#), and [artificial intelligence](#).<sup>[note 1]</sup>

### History of technology

- [By technological eras](#) [show]
- [By historical regions](#) [show]
- [By type of technology](#) [show]
- [Technology timelines](#) [show]
- [Article indices](#) [show]

V · T · E

# Laut Wikipedia...


[Article](#)
[Talk](#)
[Read](#)
[Edit](#)
[View history](#)
[Search Wikipedia](#)


**WIKIPEDIA**  
The Free Encyclopedia

[Main page](#)  
[Contents](#)  
[Featured content](#)  
[Current events](#)  
[Random article](#)  
[Donate to Wikipedia](#)  
[Wikipedia store](#)

[Interaction](#)

[Help](#)  
[About Wikipedia](#)  
[Community portal](#)  
[Recent changes](#)  
[Contact page](#)

[Tools](#)

[What links here](#)  
[Related changes](#)  
[Upload file](#)  
[Special pages](#)  
[Permanent link](#)  
[Page information](#)  
[Wikidata item](#)  
[Cite this page](#)

[Print/export](#)

[Create a book](#)  
[Download as PDF](#)  
[Printable version](#)

[In other projects](#)

[Wikimedia Commons](#)

[Languages](#)

[Deutsch](#)

[Español](#)

[Français](#)

[Bahasa Indonesia](#)

[Italiano](#)

## X (company)

From Wikipedia, the free encyclopedia

*This article is about the semi-secret research subsidiary of Alphabet Inc.. For the Canadian television series, see [X Company](#). For the online banking company, [X.com](#), see [PayPal](#). For other uses, see [X \(disambiguation\)](#).*

**X Development LLC.** (formerly [Google X](#)) is an [American](#) semi-secret research and development facility and organization founded by [Google](#) in January 2010,<sup>[1][2]</sup> which now operates as a subsidiary of [Alphabet Inc.](#)<sup>[3]</sup> X has its headquarters about a mile and a half from [Alphabet's](#) corporate headquarters, the [Googleplex](#), in [Mountain View, California](#).<sup>[4][5]</sup>

Work at X is overseen by entrepreneur scientist [Astro Teller](#), as CEO and "Captain of Moonshots".<sup>[6][7][8]</sup> The lab started with the development of [Google's](#) self-driving car.<sup>[8]</sup>

On October 2, 2015, after the complete restructuring of Google into [Alphabet](#), Google X became an independent [Alphabet](#) company and was renamed X.

- Von **Google 2010 gegründete, teils geheime Forschungseinrichtung**
- **CEO: Astro Teller**
- **Erstes Projekt war das „Selbstfahrende Auto“**

### Contents [hide]

#### 1 Projects

- 1.1 Glass
- 1.2 Makani
- 1.3 FSOC
- 1.4 Barge
- 1.5 Graduated projects
  - 1.5.1 Waymo (self-driving car)
  - 1.5.2 Loon
  - 1.5.3 Wing
  - 1.5.4 Others
- 1.6 Explorations
- 1.7 Other projects

### X Development LLC.



Formerly	<a href="#">Google X (2010–2015)</a>
Type	<a href="#">Subsidiary of Alphabet Inc.</a>
Industry	Research and development
Founded	January 2010; 8 years ago (as Google X, under Google)
Founders	October 2, 2015; 3 years ago (as X, under Alphabet Inc.)
Sebastian Thrun	Yoky Matsuoka
Astro Teller	
Headquarters	Mountain View, California, United States
Products	Artificial intelligence Computational neuroscience Health care Biotechnology
Parent	<a href="#">Google (2010–2015)</a> <a href="#">Alphabet Inc. (2015–present)</a>
Subsidiaries	<a href="#">Subsidiaries</a> [show]
Website	<a href="#">x.company</a> ↗

# Wikipedia zu DeepMind

## DeepMind

From Wikipedia, the free encyclopedia

**DeepMind Technologies Limited** is a British artificial intelligence company founded in September 2010.

Acquired by Google in 2014, the company has created a neural network that learns how to play video games in a fashion similar to that of humans,<sup>[4]</sup> as well as a Neural Turing machine,<sup>[5]</sup> or a neural network that may be able to access an external memory like a conventional Turing machine, resulting in a computer that mimics the short-term memory of the human brain.<sup>[6][7]</sup>

The company made headlines in 2016 after its AlphaGo program beat a human professional Go player for the first time in October 2015<sup>[8]</sup> and again when AlphaGo beat Lee Sedol the world champion in a five-game match, which was the subject of a documentary film.<sup>[9]</sup>

A more generic program, AlphaZero, beat the most powerful programs playing go, chess and shogi (Japanese chess) after a few hours of play against itself using reinforcement learning.<sup>[10]</sup>

### Contents [hide]

- 1 History
- 2 Machine learning
  - 2.1 Deep reinforcement learning

### DeepMind Technologies Limited



# DeepMind

Type of business	Subsidiary
Founded	23 September 2010; 7 years ago <sup>[1]</sup>
Headquarters	6 Pancras Square, <sup>[2]</sup> London N1C 4AG, UK
Founder(s)	Demis Hassabis, Shane Legg, Mustafa Suleyman
CEO	Demis Hassabis
Industry	Artificial Intelligence
Employees	400 <sup>[3]</sup>
Parent	Independent (2010–2014) Google Inc. (2014–present) Alphabet Inc. (2015–present)
Website	<a href="http://www.deepmind.com">www.deepmind.com</a>

# Wikipedia zu AlphaGo

## AlphaGo

From Wikipedia, the free encyclopedia

AlphaGo is a computer program that plays the board game Go.<sup>[1]</sup> It was developed by Alphabet Inc.'s Google DeepMind in London.

In October 2015, AlphaGo became the first computer Go program to beat a human professional Go player without handicaps on a full-sized 19×19 board.<sup>[2][3]</sup> In March 2016, it beat Lee Sedol in a five-game

match, the first time a computer Go program has beaten a 9-dan professional without handicaps.<sup>[4]</sup> Although it lost to Lee Sedol in the fourth game, Lee resigned the final game, giving a final score of 4 games to 1 in favour of AlphaGo. In recognition of the victory, AlphaGo was awarded an honorary 9-dan by the Korea Baduk Association.<sup>[5]</sup> The lead up and the challenge match with Lee Sedol were documented in a documentary film also titled *AlphaGo*,<sup>[6]</sup> directed by Greg Kohs. It was chosen by *Science* as one of the Breakthrough of the Year runners-up on 22 December 2016.<sup>[7]</sup>

At the 2017 Future of Go Summit, AlphaGo beat Ke Jie, the world No.1 ranked player at the time, in a three-game match. After this, AlphaGo was awarded professional 9-dan by the Chinese Weiqi Association.<sup>[8]</sup> After the match between AlphaGo and Ke Jie, AlphaGo retired while DeepMind continues AI research in other areas.<sup>[9]</sup>

AlphaGo uses a Monte Carlo tree search algorithm to find its moves based on knowledge previously "learned" by machine learning, specifically by an artificial neural network (a deep learning method) by extensive training, both from human and computer play.<sup>[10]</sup>





# 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.<sup>[3]</sup> 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".<sup>[4]</sup>

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

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

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.<sup>[19]</sup> **General intelligence** is among the field's long-term goals.<sup>[25]</sup> 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.

Part of a series on

## Artificial intelligence

<b>Major goals</b>	<a href="#">[show]</a>
<b>Approaches</b>	<a href="#">[show]</a>
<b>Philosophy</b>	<a href="#">[show]</a>
<b>History</b>	<a href="#">[show]</a>
<b>Technology</b>	<a href="#">[show]</a>
<b>Glossary</b>	<a href="#">[show]</a>

V · T · E

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

Wikiquotes



# Machine learning

From Wikipedia, the free encyclopedia

**Machine learning (ML)** is the study of computer algorithms that improve automatically through experience.<sup>[1]</sup> It is seen as a subset of **artificial intelligence**. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so.<sup>[2]</sup> Machine learning algorithms are used in a wide variety of applications, such as **email filtering** and **computer vision**, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

A subset of machine learning is closely related to **computational statistics**, which focuses on making predictions using computers; but not all machine learning is statistical learning. The study of **mathematical optimization** delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on **exploratory data analysis** through unsupervised learning.<sup>[4][5]</sup> In its application across business problems, machine learning is also referred to as **predictive analytics**.

## Contents [hide]

- 1 Overview
  - 1.1 Machine learning approaches
- 2 History and relationships to other fields
  - 2.1 Artificial intelligence
  - 2.2 Data mining
  - 2.3 Optimization
  - 2.4 Generalization
  - 2.5 Statistics
- 3 Theory
- 4 Approaches
  - 4.1 Types of learning algorithms
  - 4.2 Models
  - 4.3 Training models
- 5 Applications

Part of a series on  
**Machine learning  
and  
data mining**

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

V · T · E



# 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.<sup>[1][2][3]</sup>

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.<sup>[4][5][6]</sup>

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.<sup>[7][8][9]</sup>

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**

<b>Problems</b>	<a href="#">[show]</a>
<b>Supervised learning</b>	<a href="#">[show]</a>
(classification • regression)	
<b>Clustering</b>	<a href="#">[show]</a>
<b>Dimensionality reduction</b>	<a href="#">[show]</a>
<b>Structured prediction</b>	<a href="#">[show]</a>
<b>Anomaly detection</b>	<a href="#">[show]</a>
<b>Artificial neural network</b>	<a href="#">[show]</a>
<b>Reinforcement learning</b>	<a href="#">[show]</a>
<b>Theory</b>	<a href="#">[show]</a>
<b>Machine-learning venues</b>	<a href="#">[show]</a>
<b>Glossary of artificial intelligence</b>	<a href="#">[show]</a>
<b>Related articles</b>	<a href="#">[show]</a>

V · T · E

Part of a series on  
**Artificial intelligence**

<b>Major goals</b>	<a href="#">[show]</a>
--------------------	------------------------

V · T · E

## Artificial intelligence



From Wikipedia, the free encyclopedia

*"AI" redirects here. For other uses, see [AI \(disambiguation\)](#) and [Artificial intelligence \(disambiguation\)](#).*

**Artificial intelligence (AI, also machine intelligence, MI)** is intelligence displayed by machines, in contrast with the **natural intelligence (NI)** displayed by humans and other animals. In computer science AI research is defined as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of success at some goal.<sup>[1]</sup> Colloquially, the term "artificial intelligence" is applied when a machine mimics "cognitive" functions that humans associate with other **human minds**, such as "learning" and "problem solving".<sup>[2]</sup> See [glossary of artificial intelligence](#).

The scope of AI is disputed: as machines become

Part of a series on  
**Science**

**Formal** [show]

**Physical** [show]

**Life** [show]

**Social** [show]

**Applied** [show]

**Interdisciplinary** [show]

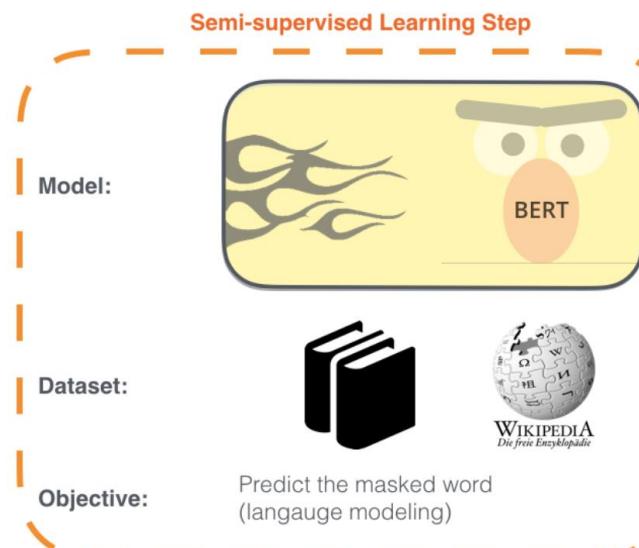
**Philosophy · History** [show]

**Glossaries of Science and**<sup>9</sup> [show]  
Technology

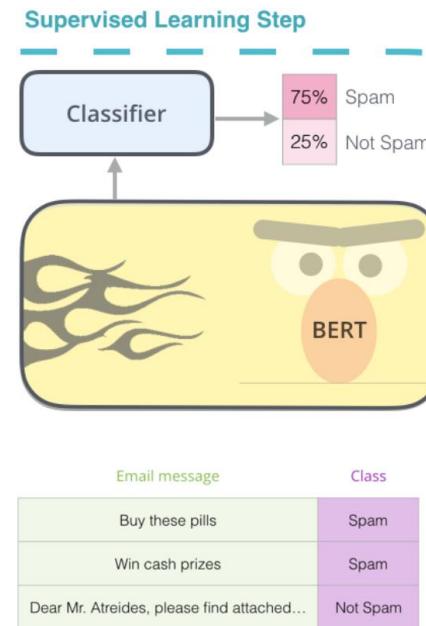
# BERT Transformer, Google 2018

1 - Semi-supervised training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.



2 - Supervised training on a specific task with a labeled dataset.



The two steps of how BERT is developed. You can download the model pre-trained in step 1 (trained on un-annotated data), and only worry about fine-tuning it for step 2. [Source for book icon].

**Das Training läuft in zwei Schritten: Pre-Training und Fine-Tuning (eine Form von Transfer Learning)**

## Watson (computer)

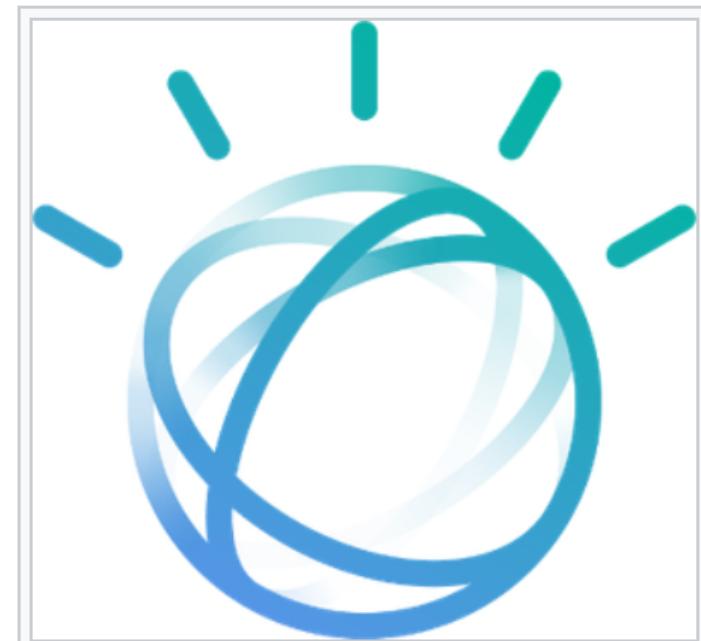


From Wikipedia, the free encyclopedia

"IBM Watson" redirects here. For the IBM laboratory, see [Thomas J. Watson Research Center](#).

**Watson** is a [question answering](#) computer system capable of answering questions posed in [natural language](#),<sup>[2]</sup> developed in IBM's DeepQA project by a research team led by [principal investigator David Ferrucci](#).<sup>[3]</sup> Watson was named after IBM's first CEO, industrialist [Thomas J. Watson](#).<sup>[4][5]</sup> The computer system was specifically developed to answer questions on the [quiz show](#) *Jeopardy!*<sup>[6]</sup> and, in 2011, the Watson computer system competed on *Jeopardy!* against former winners Brad Rutter and [Ken Jennings](#)<sup>[4][7]</sup> winning the first place prize of \$1 million.<sup>[8]</sup>

Watson had access to 200 million pages of structured and



Watson's [avatar](#), inspired by the IBM "smarter planet" logo<sup>[1]</sup>



# 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).<sup>[1]</sup>

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.<sup>[1]</sup>

Transformers have rapidly become the model of choice for NLP problems,<sup>[2]</sup> 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.<sup>[3][4]</sup>

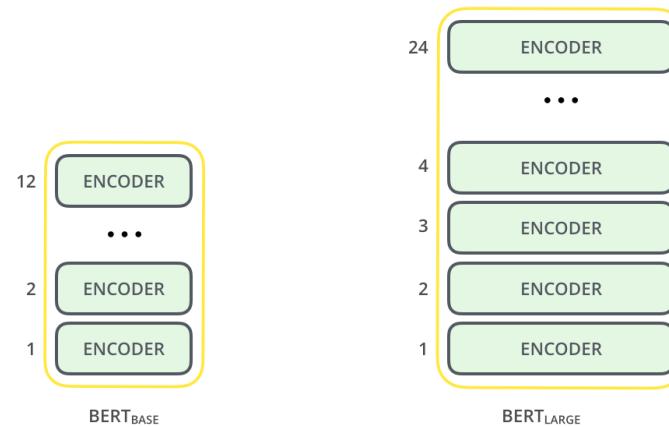
# BERT Transformer, Google 2018

## 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.<sup>[1][2]</sup> As of 2019, Google has been leveraging BERT to better understand user searches.<sup>[3]</sup>

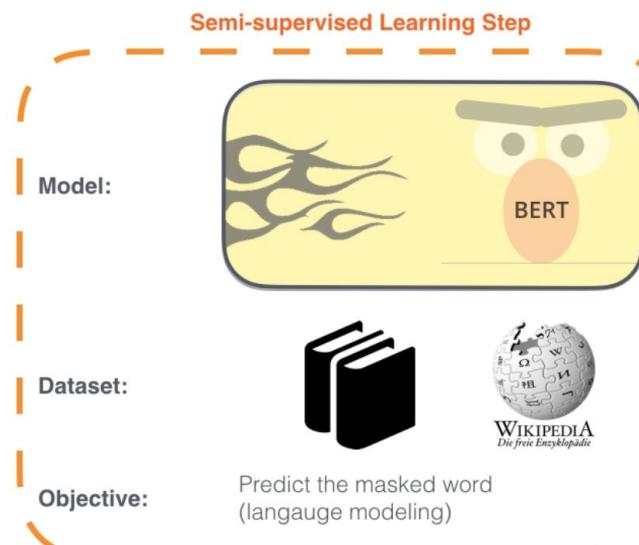
The original English-language BERT model comes with two pre-trained general types:<sup>[1]</sup> (1) the [BERT<sub>BASE</sub>](#) model, a 12-layer, 768-hidden, 12-heads, 110M parameter neural network architecture, and (2) the [BERT<sub>LARGE</sub>](#) model, a 24-layer, 1024-hidden, 16-heads, 340M parameter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



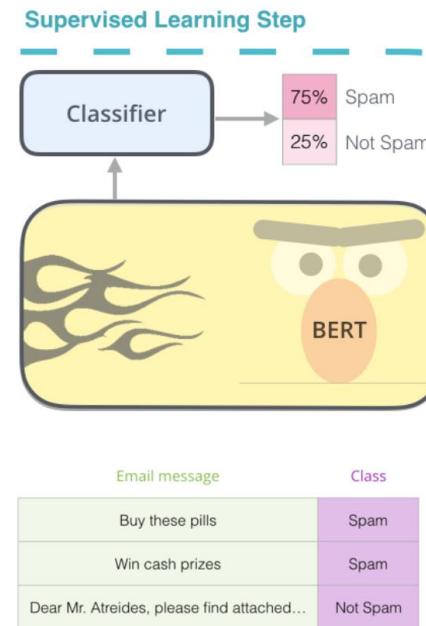
# BERT Transformer, Google 2018

1 - Semi-supervised training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.



2 - Supervised training on a specific task with a labeled dataset.



The two steps of how BERT is developed. You can download the model pre-trained in step 1 (trained on un-annotated data), and only worry about fine-tuning it for step 2. [Source for book icon].

Das **Training** läuft in zwei Schritten: **Pre-Training** und **Fine-Tuning** (eine Form von **Transfer Learning**)

# 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.<sup>[2]</sup> 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,<sup>[3]</sup> is part of a trend in [natural language processing](#) (NLP) systems of pre-trained language representations.<sup>[1]</sup> 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.<sup>[4]</sup>

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.<sup>[4]</sup> 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.<sup>[1]:34</sup> [David Chalmers](#), an Australian philosopher, described GPT-3 as "one of the most interesting and important AI systems ever produced."<sup>[5]</sup>

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.<sup>[6]</sup>

# GPT-3: Trainingsdaten

Since Neural Networks are **compressed/compiled version** of the training data, the size of the dataset has to scale accordingly with the size of the model. GPT-3 175B is trained with 499 Billion tokens. Here is the breakdown of the data:

Dataset	# Tokens (Billions)
Total	499
Common Crawl (filtered by quality)	410
WebText2	19
Books1	12
Books2	55
Wikipedia	3