Report Topic : Big data

Student : Al-Shiama Ashraf Anwer Abd-Hamed

Seat Number : 200  
Group : 2

Section : 9

B.N : 13

Big data

*Big data is a field that treats ways to analyze, systematically extract information from, or*

*otherwise deal with data sets that are too large or complex to be dealt with by traditional*

*data-processing application software. Data with many fields (columns) offer greater statistical power,*

*while data with higher complexity (more attributes or columns) may lead to a higher* [*false discovery*](https://en.wikipedia.org/wiki/False_discovery_rate)

[*rate*](https://en.wikipedia.org/wiki/False_discovery_rate)*. Big data analysis challenges include* [*capturing data*](https://en.wikipedia.org/wiki/Automatic_identification_and_data_capture)*,* [*data storage*](https://en.wikipedia.org/wiki/Computer_data_storage)*,* [*data analysis*](https://en.wikipedia.org/wiki/Data_analysis)*, search,* [*sharing*](https://en.wikipedia.org/wiki/Data_sharing)*,*

[*transfer*](https://en.wikipedia.org/wiki/Data_transmission)*,* [*visualization*](https://en.wikipedia.org/wiki/Data_visualization)*,* [*querying*](https://en.wikipedia.org/wiki/Query_language)*, updating,* [*information privacy*](https://en.wikipedia.org/wiki/Information_privacy)*, and data source. Big data was originally*

*associated with three key concepts: volume, variety, and velocity.The analysis of big data presents challenges in sampling, and thus previously allowing for only observations and sampling. Therefore, big data often includes data with sizes that exceed the capacity of traditional software to process within an acceptable time and value.*

*Current usage of the term big data tends to refer to the use of* [*predictive analytics*](https://en.wikipedia.org/wiki/Predictive_analytics)*,* [*user behavior analytics*](https://en.wikipedia.org/wiki/User_behavior_analytics)*, or certain other advanced data analytics methods that extract* [*value*](https://en.wikipedia.org/wiki/Data_valuation) *from big data, and seldom to a particular size of data set. "There is little doubt that the quantities of data now available are indeed large, but that's not the most relevant characteristic of this new data ecosystem." Analysis of data sets can find new correlations to "spot business trends, prevent diseases, combat crime and so on".Scientists, business executives, medical practitioners, advertising and* [*governments*](https://en.wikipedia.org/wiki/Government_database) *alike regularly meet difficulties with large data-sets in areas including* [*Internet searches*](https://en.wikipedia.org/wiki/Web_search_engine)*,* [*fintech*](https://en.wikipedia.org/wiki/Fintech)*, healthcare analytics, geographic information systems, urban informatics, and* [*business informatics*](https://en.wikipedia.org/wiki/Business_informatics)*. Scientists encounter limitations in* [*e-Science*](https://en.wikipedia.org/wiki/E-Science) *work, including* [*meteorology*](https://en.wikipedia.org/wiki/Meteorology)*,* [*genomics*](https://en.wikipedia.org/wiki/Genomics)*,**[[6]](https://en.wikipedia.org/wiki/Big_data" \l "cite_note-6)* [*connectomics*](https://en.wikipedia.org/wiki/Connectomics)*, complex physics simulations, biology, and environmental research.*

*The size and number of available data sets has grown rapidly as data is collected by devices such as mobile devices, cheap and numerous information-sensing* [*Internet of things*](https://en.wikipedia.org/wiki/Internet_of_things) *devices, aerial (*[*remote sensing*](https://en.wikipedia.org/wiki/Remote_sensing)*), software logs,* [*cameras*](https://en.wikipedia.org/wiki/Digital_camera)*, microphones,* [*radio-frequency identification*](https://en.wikipedia.org/wiki/Radio-frequency_identification) *(RFID) readers and* [*wireless sensor networks*](https://en.wikipedia.org/wiki/Wireless_sensor_networks)*. The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s; as of 2012, every day 2.5* [*exabytes*](https://en.wikipedia.org/wiki/Exabyte) *(2.5×260 bytes) of data are generated. Based on an IDC report prediction, the global data volume was predicted to grow exponentially from 4.4 zettabytes to 44 zettabytes between 2013 and 2020. By 2025, IDC predicts there will be 163 zettabytes of data. One question for large enterprises is determining who should own big-data initiatives that affect the entire organization.*

*Relational database management systems and desktop statistical software packages used to visualize data often have difficulty processing and analyzing big data. The processing and analysis of big data may require "massively parallel software running on tens, hundreds, or even thousands of servers". What qualifies as "big data" varies depending on the capabilities of those analyzing it and their tools. Furthermore, expanding capabilities make big data a moving target. "For some organizations, facing hundreds of gigabytes of data for the first time may trigger a need to reconsider data management options. For others, it may take tens or hundreds of terabytes before data size becomes a significant consideration.*

# Applications

Big data has increased the demand of information management specialists so much so that [Software AG](https://en.wikipedia.org/wiki/Software_AG), [Oracle Corporation](https://en.wikipedia.org/wiki/Oracle_Corporation), [IBM](https://en.wikipedia.org/wiki/IBM), [Microsoft](https://en.wikipedia.org/wiki/Microsoft), [SAP](https://en.wikipedia.org/wiki/SAP_AG), [EMC](https://en.wikipedia.org/wiki/EMC_Corporation), [HP](https://en.wikipedia.org/wiki/Hewlett-Packard), and [Dell](https://en.wikipedia.org/wiki/Dell) have spent more than $15 billion on software firms specializing in data management and analytics. In 2010, this industry was worth more than $100 billion and was growing at almost 10 percent a year: about twice as fast as the software business as a whole.Developed economies increasingly use data-intensive technologies. There are 4.6 billion mobile-phone subscriptions worldwide, and between 1 billion and 2 billion people accessing the internet.[5] Between 1990 and 2005, more than 1 billion people worldwide entered the middle class, which means more people becamemore literate, which in turn led to information growth. The world's effective capacity to exchange information through telecommunication networks was 281 [petabytes](https://en.wikipedia.org/wiki/Petabytes) in 1986, 471

[petabytes](https://en.wikipedia.org/wiki/Petabytes) in 1993, 2.2 exabytes in 2000, 65 [exabytes](https://en.wikipedia.org/wiki/Exabytes) in 2007[10] and predictions put the amount of internet traffic at 667 exabytes annually

by 2014.[5] According to one estimate, one-third of the globally stored information is in the form of alphanumeric text and still image

data,[53] which is the format most useful for most big data applications. This also shows the potential of yet unused data (i.e. in the form of video and audio content).

While many vendors offer off-the-shelf solutions for big data, experts recommend the development of in-house solutions custom-tailored to



Bus wrapped with [SAP](https://en.wikipedia.org/wiki/SAP_AG) big data parked outside [IDF13](https://en.wikipedia.org/wiki/Intel_Developer_Forum).

solve the company's problem at hand if the company has sufficient technical capabilities.[54]

## Government

The use and adoption of big data within governmental processes allows efficiencies in terms of cost, productivity, and innovation,[55] but does not come without its flaws. Data analysis often requires multiple parts of government (central and local) to work in collaboration and create new and innovative processes to deliver the desired outcome. A common government organization that makes use of big data is the National Security Administration ([NSA](https://en.wikipedia.org/wiki/National_Security_Agency)), who monitor the activities of the Internet constantly in search for potential patterns of suspicious or illegal activities their system may pick up.

[Civil registration and vital statistics](https://en.wikipedia.org/wiki/Civil_registration_and_vital_statistics) (CRVS) collects all certificates status from birth to death. CRVS is a source of big data for governments.

## International development

Research on the effective usage of information and communication technologies for development (also known as "ICT4D") suggests that big data technology can make important contributions but also present unique challenges to [international development](https://en.wikipedia.org/wiki/International_development).[56][57] Advancements in big data analysis offer cost-effective opportunities to improve decision-making in critical development areas such as health care, employment, [economic productivity](https://en.wikipedia.org/wiki/Economic_productivity), crime, security, and [natural disaster](https://en.wikipedia.org/wiki/Natural_disaster) and resource management.[58][59][60] Additionally, user-generated data offers new opportunities to give the unheard a voice.[61] However, longstanding challenges for developing regions such as inadequate technological infrastructure and economic and human resource scarcity exacerbate existing concerns with big data such as privacy, imperfect methodology, and interoperability issues.[58] The challenge of "big data for development"[58] is currently evolving toward the application of this data through machine learning, known as "artificial intelligence for development (AI4D).[62]

### Benefits

A major practical application of big data for development has been "fighting poverty with data".[63] In 2015, Blumenstock and colleagues estimated predicted poverty and wealth from mobile phone metadata [64] and in 2016 Jean and colleagues combined satellite imagery and machine learning to predict poverty.[65] Using digital trace data to study the labor market and the digital economy in Latin America, Hilbert and colleagues [66][67] argue that digital trace data has several benefits such as:

Thematic coverage: including areas that were previously difficult or impossible to measure Geographical coverage: our international sources provided sizable and comparable data for almost all countries, including many small countries that usually are not included in international inventories

Level of detail: providing fine-grained data with many interrelated variables, and new aspects, like network connections

Timeliness and timeseries: graphs can be produced within days of being collected

### Challenges

At the same time, working with digital trace data instead of traditional survey data does not eliminate the traditional challenges involved when working in the field of international quantitative analysis. Priorities change, but the basic discussions remain the same. Among the main challenges are:

Representativeness. While traditional development statistics is mainly concerned with the representativeness of random survey samples, digital trace data is never a random sample.

Generalizability. While observational data always represents this source very well, it only represents what it represents, and nothing more. While it is tempting to generalize from specific observations of one platform to broader settings, this is often very deceptive.

Harmonization. Digital trace data still requires international harmonization of indicators. It adds the challenge of so-called "data-fusion", the harmonization of different sources.

Data overload. Analysts and institutions are not used to effectively deal with a large number of variables, which is efficiently done with interactive dashboards. Practitioners still lack a standard workflow that would allow researchers, users and policymakers to efficiently and effectively.[66]

## Healthcare

Big data analytics has helped healthcare improve by providing personalized medicine and prescriptive analytics, clinical risk intervention and predictive analytics, waste and care variability reduction, automated external and internal reporting of patient data, standardized medical terms and patient registries and fragmented point solutions.[68][69][70][71] Some areas of improvement are more aspirational than actually implemented. The level of data generated within [healthcare systems](https://en.wikipedia.org/wiki/Health_system) is not trivial. With the added adoption of mHealth, [eHealth and wearable technologies the volume of data will continue to increase. This includes electronic health record data, imaging data, patient generated data, sensor data, and other forms of difficult to process data.](https://en.wikipedia.org/wiki/Electronic_health_record) There is now an even greater need for such environments to pay greater attention to data and information quality.[72] "Big data very often means '[dirty data](https://en.wikipedia.org/wiki/Dirty_data)' and the fraction of data inaccuracies increases with data volume growth." Human inspection at the big data scale is impossible and there is a desperate need in health service for intelligent tools for accuracy and believability control and handling of information missed.[73] While extensive information in healthcare is now electronic, it fits under the big data umbrella as most is unstructured and difficult to use.[74] The use of big data in healthcare has raised significant ethical challenges ranging from risks for individual rights, privacy and [autonomy](https://en.wikipedia.org/wiki/Autonomy), to transparency and trust.[75]

Big data in health research is particularly promising in terms of exploratory biomedical research, as data-driven analysis can move forward more quickly than hypothesis-driven research.[76] Then, trends seen in data analysis can be tested in traditional, hypothesis-driven follow up biological research and eventually clinical research.

[A related application sub-area, that heavily relies on big data, within the healthcare field is that of computer- aided diagnosis in medicine. [77] For instance, for](https://en.wikipedia.org/wiki/Computer-aided_diagnosis) [epilepsy](https://en.wikipedia.org/wiki/Epilepsy) [monitoring it is customary to create 5 to 10 GB of](https://en.wikipedia.org/wiki/Computer-aided_diagnosis) data daily. [78] Similarly, a single uncompressed image of breast [tomosynthesis](https://en.wikipedia.org/wiki/Tomosynthesis) averages 450 MB of data. These are just few of the many examples where [computer-aided diagnosis](https://en.wikipedia.org/wiki/Computer-aided_diagnosis) uses big data. For this reason, big data has been recognized as one of the seven key challenges that computer-aided diagnosis systems need to overcome in order to reach the next level of performance. [80]

## Education

A [McKinsey Global Institute](https://en.wikipedia.org/wiki/McKinsey_%26_Company) study found a shortage of 1.5 million highly trained data professionals and managers[44] and a number of universities[81] including [University of Tennessee](https://en.wikipedia.org/wiki/University_of_Tennessee) and [UC Berkeley](https://en.wikipedia.org/wiki/UC_Berkeley), have created masters programs to meet this demand. Private boot camps have also developed programs to meet that demand, including free programs like [The Data Incubator](https://en.wikipedia.org/wiki/The_Data_Incubator) or paid programs like [General Assembly](https://en.wikipedia.org/wiki/General_Assembly).[82] In the specific field of marketing, one of the problems stressed by Wedel and Kannan[83] is that marketing has several sub domains (e.g., advertising, promotions, product development, branding) that all use different types of data. Because one-size-fits-all analytical solutions are not desirable, business schools should prepare marketing managers to have wide knowledge on all the different techniques used in these subdomains to get a big picture and work effectively with analysts.

## Internet of Things (IoT)

Big data and the IoT work in conjunction. Data extracted from IoT devices provides a mapping of device inter-connectivity. Such mappings have been used by the media industry, companies, and governments to more accurately target their audience and increase media efficiency. The IoT is also increasingly adopted as a meansof gathering sensory data, and this sensory data has been used in medical,[88] manufacturing[89] and transportation[90] contexts.

[Kevin Ashton](https://en.wikipedia.org/wiki/Kevin_Ashton), the digital innovation expert who is credited with coining the term,[91] defines the Internet of things in this quote: "If we had computers that knew everything there was to know about things—using data they gathered without any help from us—we would be able to track and count everything, and greatly reduce waste, loss, and cost. We would know when things needed replacing, repairing, or recalling, and whether they were fresh or past their best."

## Information technology

Especially since 2015, big data has come to prominence within [business operations](https://en.wikipedia.org/wiki/Business_operations) as a tool to help employees work more efficiently and streamline the collection and distribution of [information technology](https://en.wikipedia.org/wiki/Information_technology) (IT). The use of big data to resolve IT and data collection issues within an enterprise is called [IT operations analytics](https://en.wikipedia.org/wiki/IT_operations_analytics) (ITOA).[92] By applying big data principles into the concepts of [machine intelligence](https://en.wikipedia.org/wiki/Machine_intelligence) and deep computing, IT departments can predict potential issues and move to provide solutions before the problems even happen.[92] In this time, ITOA businesses were also beginning to play a major role in [systems management](https://en.wikipedia.org/wiki/Systems_management) by offering platforms that brought individual [data silos](https://en.wikipedia.org/wiki/Data_silos) together and generated insights from the whole of the system rather than from isolated pockets of data.

# **Screen Shots**

# 

# **Code**

<!DOCTYPE html>

<html xmlns="http://www.w3.org/1999/xhtml">

<head>

<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />

<link rel="stylesheet" type="text/css" href="style.css" />

<title>Artificial intelligence</title>

</head>

<body>

<div id="container">

<div id="mainpic"></div>

<div id="menu">

<ul>

<li class="menuitem"><a href="index.html">Home</a></li>

<li class="menuitem"><a href="apps.html">Applications</a></li>

<li class="menuitem"><a href="platform.html">History</a></li>

<li class="menuitem"><a href="types.html">Types</a></li>

<li class="menuitem">

<a href="deepLearning.html">Deep Learning</a>

</li>

</ul>

</div>

<div id="content">

<h2>Deep Learning</h2>

<p>

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.Deep-learning architectures such as

deep neural networks, deep belief networks, graph neural networks,

recurrent neural networks and convolutional neural networks have been

applied to fields including computer vision, speech recognition,

natural language processing, 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. 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 analogue.[8][9][10] The adjective

"deep" in deep learning refers to the use of multiple layers in the=

connectionist models, for the sake of efficiency, trainability and

understandability, whence the "structured" part.

</p>

</div>

</div>

</body>

</html>