



DIGITALISATION FOR DEVELOPMENT. A TOOLKIT FOR
DEVELOPMENT COOPERATION PRACTITIONERS
INTERNATIONAL PARTNERSHIPS (INTPA)

Big Data and Artificial Intelligence

InfoSheet n°5

This InfoSheet is part of a series on digitalisation and relevance to EU International Partnerships and development cooperation programmes. The Toolkit is designed to provide key definitions, main opportunities and challenges for global development presented by digital transformation, case studies and suggested further reading. Learn more on [Cap4Dev](#)

The world is experiencing a data revolution. The volume of data produced by individuals, organizations and businesses is constantly increasing.

For those who wish to play a role in the data economy, issues ranging from connectivity to processing and storage of data, computing power and cybersecurity have to be tackled. Moreover, governance structures have to be developed for handling data and there needs to be an increase in the pools of quality data available for use and re-use. Such issues are taken on board through the approach of the European Union (EU) in an ever more competitive and geopolitically contested environment, where data are at the centre of economic and societal transformation.

The [EU Data Strategy](#) and the [White Paper on AI](#) are of relevance here and have been mentioned in the accompanying **Infosheet on policy and regulation**.¹

The EU's human-centric approach is important in the context of International Partnerships, and as a counterbalance to avoid a 'Big Data dystopia'. Recent initiatives include the Gaia-X initiative in Europe, and the EU-AU Data Flagship both of which are briefly presented here.

The Gaia-X initiative intends to create a European cloud data infrastructure to foster digital independence in the context of geopolitical competitors. Gaia-X will support the development of a digital ecosystem in Europe, which will generate innovation and new data-driven services and applications. It aims at interoperability and portability of infrastructure, data and services, and the establishment of a high degree of trust for users.

Through the EU-AU Data Flagship, the European Commission and the African Union Commission will strive

to develop a data framework based on shared values and principles and with the main goals of protecting citizens' rights, ensuring data sovereignty and supporting the creation of the African Single Digital Market. The joint data framework will focus entirely on creating data harmonisation on the African continent and facilitating investments in public data infrastructure and private data technologies and services.

Data are the building block of Artificial Intelligence (AI). Data and AI are improving everyday lives and are already being extensively used in development cooperation activities. By leveraging Big Data, AI can, for instance:

- contribute to climate change mitigation and adaptation, for example by improving weather prediction, ensuring security of energy distribution, as well as tracking sources of air pollution;
- increase the efficiency of farming, by enabling early detection of crop diseases and forecasting market supply and demand;
- improve urban development and planning, including urban mobility, by analyzing data to better understand how cities and services can be modified for the benefit of their residents;
- enhance private sector development by improving supply chains and industry logistics; and
- make health diagnosis more precise, detecting and classifying diseases, enabling better prevention and even treatment.

This infosheet provides an introduction to Big Data and AI and gives some examples of use cases relevant to development cooperation.

¹ The EU is proposing a Data Act in 2021, which could foster business-to-government data sharing for the public interest among other issues.

What are Big Data?

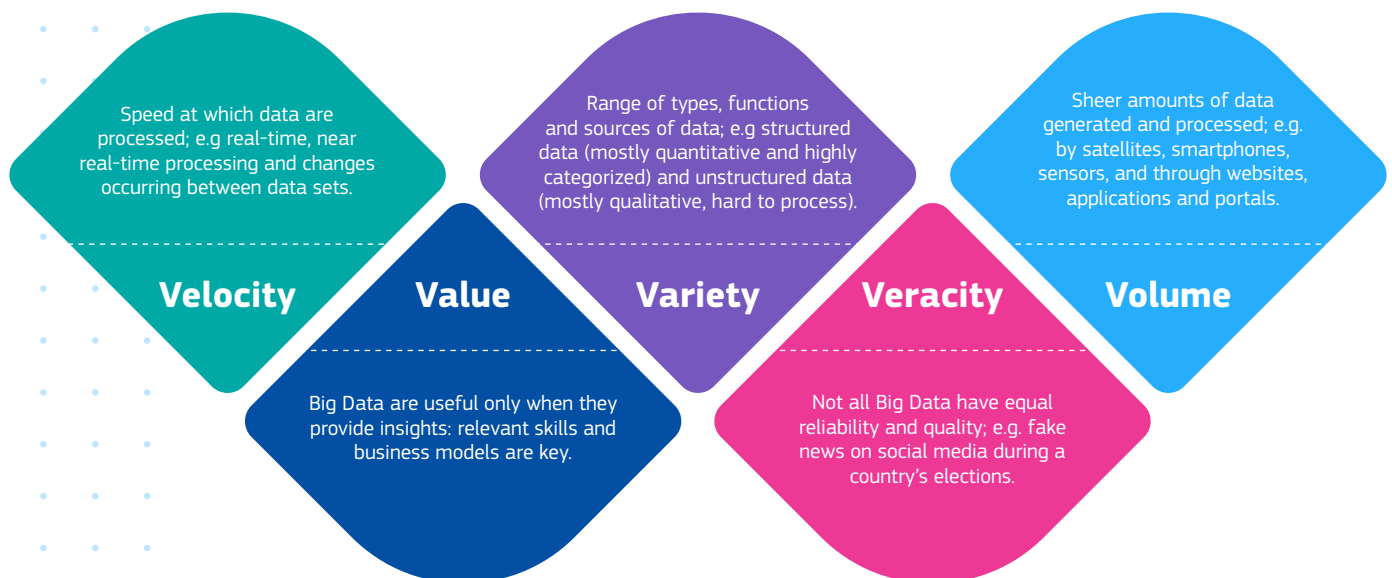
Big Data is an umbrella term referring to the large amounts of digital data that are continuously being generated by the global population. Data can be created by individuals, organizations, governments, objects and machines covering any kind of interaction with technology, be it direct (a person who posts a picture on a social network) or indirect (an image of the same person captured by a satellite). There is virtually no difference between what we call Big Data and any other type of data. At the end of the day, **all active and passive interactions** that each person has with digital technologies are translated into digital numbers (zeros and ones). In this way, Big Data incorporates text messages, images and videos, voice messages, GPS signals and geospatial data, electronic purchases and mobile-banking transactions, online searches and browsing history, data from satellites, cars or fitness watches, network and time-series data, etc.

Data can be stored both in local storage systems and on the Internet in cloud-based data centres. Because Big Data are extremely large in volume, highly diverse and complex, and growing in size very quickly, they cannot be managed with traditional data management tools.

The Internet of Things (IoT)

The International Data Corporation (IDC) estimates that there will be 41.6 billion IoT devices by 2025. IDC noted that the 79.4 zettabytes (ZB) of data will consist of everything from small items that generate data, like machine health and status, as well as large unstructured data from video surveillance cameras.²

THE 5 Vs OF BIG DATA

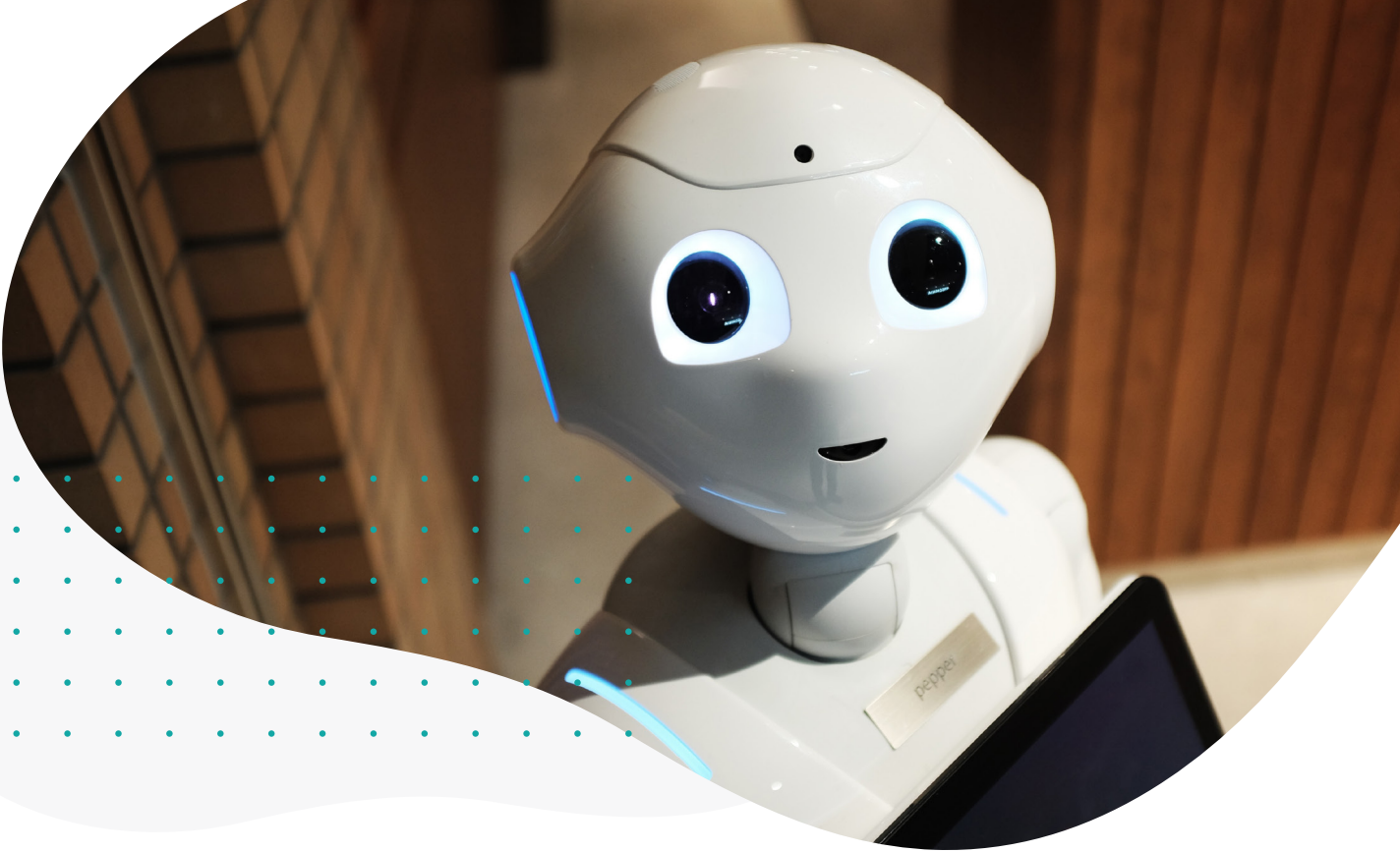


Big Data can be privately owned or have **varying degrees of access control**. Today, companies manage and analyze an enormous amount of data sets and these can bring valuable insights to support decision-making or provide market intelligence. Considered as **"the new gold"**, Big Data have become a commodity and a key factor of production – but only if processed correctly. Companies often monetize data assets by offering data-driven products and applications. For instance, Facebook does not simply sell its user data in raw form. Instead, it has developed an enormously more profitable business around **data-driven advertising** that allows shopping and entertainment brands to target Facebook users, based on the data these users provide about their opinions and preferences. Many of today's biggest IT companies have been accused of monopolizing data and abusing market dominance (much related to data) and there is currently a surge in interest for regulatory scrutiny. Considering this from a global perspective, it can have significant **impacts on social trends and political events**, including people's perception of events in the world including attempts to influence national presidential and parliamentary elections.

In order to avoid a myriad of abuses made possible by Big Data, and for governments to benefit appropriately from relevant revenues and insights, it is key to foster **laws and policies that regulate data** at the national, regional and international level. Public sector institutions have also started integrating **data analysis to drive policymaking** using new tools. For instance, the potential of Big Data in giving information about **changes in landscape and behaviour** over time allows for predictive modelling and other types of data analysis, which enable public institutions to focus more on prevention, instead of just response and mitigation. Hence, the strong need to have public-private partnerships that prove sustainable over time and have clear frameworks that regulate roles, needs and expectations on all sides.³ This is extremely relevant when it comes to understanding **how Big Data can help address development issues**, measuring the impact of relevant strategies and programmes, as well as producing effective public policies. A key concept to be addressed in countries where data governance is lacking or still at its early stages is **data sovereignty**. Data sovereignty is a country-specific requirement that data are subject to laws and governance structures of the country in which they are collected and processed, and must remain within national borders.

² Business wire, 18 June 2019: [Growth Connected to IoT](#)

³ United Nations, [Big Data for Sustainable Development](#)



What is AI?

AI refers to any machine or algorithm that is capable of observing its environment, learning from experience, and, based on the knowledge gained, adjusting to new inputs and performing human-like tasks. To perform tasks computers generally require step-by-step instructions, called algorithms. A general distinction is made between weak and strong AI. **Weak AI** solves an assigned problem with self-optimising algorithms. **Strong AI** has intellectual skills comparable or even superior to human intelligence. AI research and applications are currently at their most advanced in the area of weak AI.⁴ A typical application of AI in this area is **machine learning**, which is used to analyse and process big data. While doing this, AI can detect relationships and patterns that elude the perception of humans and predict future patterns and behaviour. This allows for the use of Big Data to automate and enhance complex descriptive and predictive analytical tasks that would be extremely labour intensive and time-consuming if performed by humans.

5G (and subsequently 6G) network architecture supports AI processing, speeding up services on the cloud. In turn, AI analyzes and learns from the same data faster.⁵

Big Data feed the machine learning process. In turn, the faster and more efficient ways to analyze data provided by AI allow us to get deeper insights from larger amounts of data. There is a **two-way relationship between Big Data and AI**: Big Data are of no use without being analyzed, but AI strongly depends on the quality of Big Data. Since AI and machine-learning success are subject to the data available, we cannot derive **reliable insights from AI** if we cannot rely on the data to be analyzed. Challenges related to the availability of reliable data and the subsequent missed opportunities in AI are particularly relevant in **partner countries**, in which the national statistical system and digital economy are still developing.

Together with these issues, there are several potential risks related to use of **AI**. These include ethics, human rights and democracy. AI can cause intentional and unintentional harm, including threats to fundamental rights such as privacy and non-discrimination, relating to gender or ethnic groups, intrusion in our private lives, or being used for criminal purposes (as well as for disinformation and cyberattacks). Problems are not only related to misleading data but also to other factors like the opacity of the system, leading to a **black box effect**, and the conscious **manipulation of behaviour** that some entities might decide to implement. In addition, it is vital to consider the increasing complexity of **neural networks**, which may incorporate thousands of layers, and make it more and more difficult to understand the decision-making process of AI, and therefore to verify if ethical rules have been respected. AI may also transform the nature and practice of conflict. Not only can it increase the effectiveness of deployment of weapons systems, which can operate autonomously, but AI also seeks to dramatically improve the speed and accuracy of military logistics, intelligence and situational awareness, battlefield planning and overall operations.

Key references on AI and ethics:

European Commission
[High-Level Expert Group on Artificial Intelligence](#)

European Parliament
[The Ethics of Artificial Intelligence: Issues and Initiatives](#)

EU Agency for Fundamental Rights (FRA)
[Artificial Intelligence, Big Data and Fundamental Rights](#)

UNESCO
[Preliminary Study on the Ethics of Artificial Intelligence](#)

OECD
[Ethical guidelines](#)

⁴ German Federal Ministry for Economic Cooperation and Development (BMZ), [Glossary – Digitalisation in Development cooperation](#)

⁵ Deepsig, [How AI Improves 5G Wireless Capabilities](#)

During its history AI has gone through several fluctuations in its development:

- 1950s** ● The Turing test is proposed: if a machine can trick humans into thinking it is human, it has demonstrated human intelligence. The term AI is coined to describe “the science and engineering of making intelligent machines”.
- 1960s** ● AI development and machine learning accelerates, as computers become more accessible and powerful.
- 1970s** ● AI winter: Research and development of AI undergoes a major decrease in interest and investments, due to many false-starts and dead-ends.
- 1980s** ● A renewed boost in funding determines progress, e.g. in deep learning (now fully used for chatbots, facial recognition, personalized ads, etc.).
- 1990s** ● A decade of achievements for AI, e.g. Deep Blue, an IBM chess-playing computer that beats world champion Garry Kasparov, and Kismet, a robot able to recognize and display emotions.
- 2000s - today** ● AI gets more and more sophisticated, with industrial robots, drones and driverless cars being created. It also becomes mainstream through e-commerce and digital platforms.

AI and the future of work

New technologies have mostly proven to affect tasks, rather than jobs. This explains why digital technologies do not simply create and destroy jobs: they also change what people do on the job, and how they do it. AI is mainly taking over cheap routine-based labour opportunities, including those recently outsourced to developing countries, e.g. in customer service support, industry and manufacturing, and back-office administration. Yet AI advances do not necessarily imply job losses for middle-tier or lower-wage workers, as innovations also create jobs.⁶ For countries to keep up with new labour market needs resulting from technological progress it is key to invest in enabling policies focused on the innovative skills required to manage technology, as seen with past technological shifts.⁷

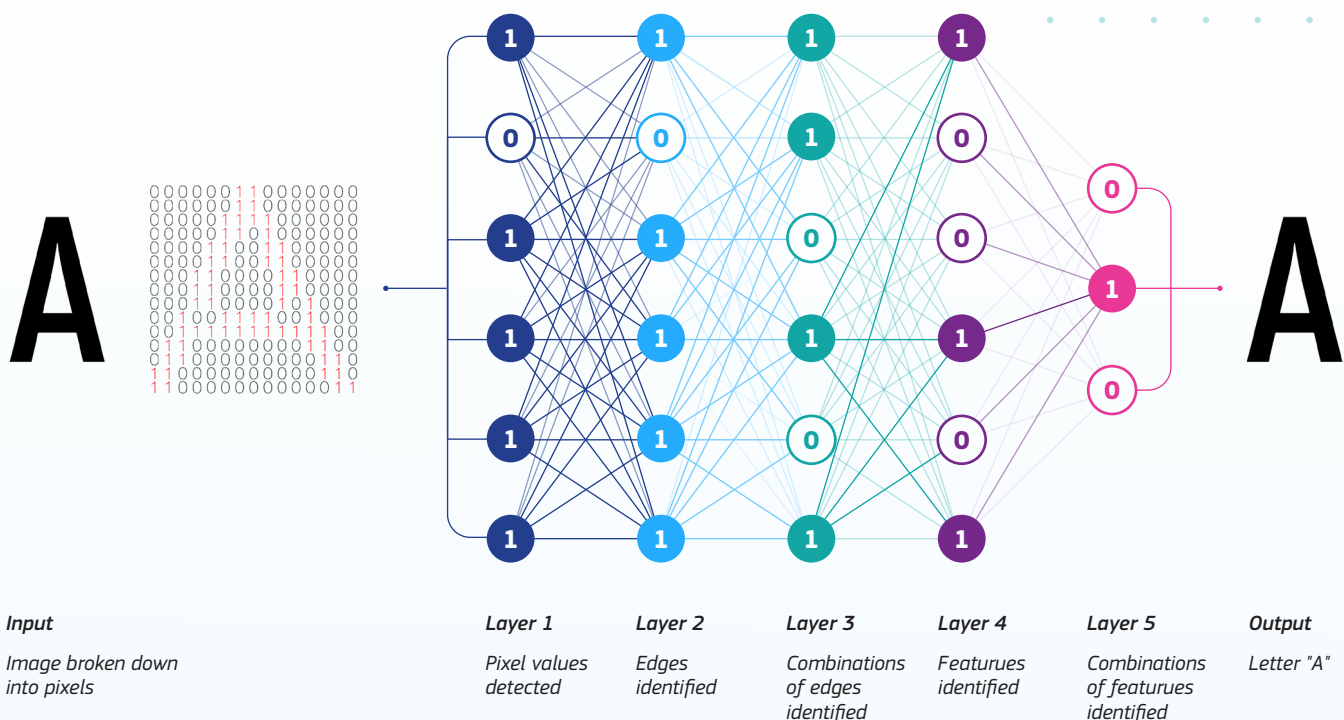


Entrepreneurship and Skills will be the main topics of two future infosheets in this series.



Policy and Regulation was the main topic of one of the previous infosheets in this series.

NEURAL NETWORK FOR OPTICAL CHARACTER RECOGNITION (OCR)



How does OCR work?

OCR first breaks down the letter “A” into pixels and the neural network uses these pixels as the input in layer 1. Pixels are then passed through the different layers of artificial digital neurons. Different features of the input are revealed in each layer and each artificial neuron transmits signals to certain neurons in the following layer to eventually achieve the output “Letter-A”.

⁶ European Commission, [The Changing Nature of Work and Skills in the Digital Age](#)

⁷ MIT Work of the Future, [The Work of the Future Report 2020](#)



CASE STUDY

Brief examples of how Big Data and AI are being used for international development

- [Springster](#) has developed a chatbot, called [Big Sis](#), that gives expert advice to girls' questions about sexual health. The platform operates through Facebook Messenger/Whatsapp and offers a safe space for girls to access information confidentially, in contexts where it is often impossible to access information about sexual health and relationships, due to social stigma and where lack of awareness often means girls are given unhelpful or incorrect information. The initial phase was conducted in the Philippines, South Africa, Tanzania and Nigeria.
- **The University of Chicago** created the [Million Neighborhoods Map](#), which visualizes building infrastructure and street networks – or the lack thereof. The objective is to provide local authorities and community residents with a visualization tool to help inform and prioritize infrastructure projects in underserved neighbourhoods, including informal urban settlements. The Million Neighborhoods Map has been created by the use of research developed algorithms applied to OpenStreetMap, an open-source GIS database freely available and based on crowdsourcing.
- **The Open Algorithms Project (OPAL)** is a socio-technological innovation that leverages big data from the private sector for social good purposes. It functions primarily by “sending the code to the data” (as opposed to the other way around) in a privacy-preserving, predictable, participatory, scalable, and sustainable manner. By unlocking the power of big data held by private entities, OPAL aims to contribute to building local capacities and connections and help shape the future technological, political, ethical and legal frameworks that will help govern the local collection, control and use of big data.
- **Google Earth Engine**, one of the most well-known platforms for analyzing Big Data, is a platform for the scientific analysis and visualization of geospatial datasets that include historical images of the Earth going back more than forty years. [Global Forest Watch](#) is a prime example of how this platform can be utilized, showcasing how vast amounts of satellite imagery can be analyzed to identify where and when tree cover has changed over time and on a global scale. More importantly, it has helped

to highlight the increasing rates of tropical forest loss and spurred the implementation of national forest monitoring programmes.

- The EU's [Copernicus](#) Earth Observation system provides a mass of data which can be processed by AI (Machine Learning) for public and private sector purposes.



Copernicus was the main topic of one of the previous infosheets of this series.

CASE STUDY

Emergency Response and Earthquake Forecasting

In recent years, many countries located in highly seismic zones (a zone where there is a high number of earthquakes) like Nepal are increasingly turning to AI to build resilience against future earthquakes. AI is already using predictive models to analyze massive amounts of seismic data, aiding understanding, and providing faster and reliable early warning signals before earthquakes hit, which reduces the loss of life and damage to property during the earthquakes. Furthermore, AI is also being used in the aftermath of earthquakes to help save lives and speed up the rescue mission. For example, a predictive model based on machine learning techniques was developed to mitigate the aftermath of a major earthquake hitting Nepal in 2015, with a magnitude of 8.1Ms and a maximum Mercalli Intensity of VIII (Severe), affecting more than 8 million people. During the aftermath of this earthquake, drones were deployed and used to map and assess the destruction, rapidly providing valuable information on the humanitarian needs of survivors. Three years later, Fusemachines (a company developing products that leverage AI) and Geospatial Information Systems (GIS) partnered with Sankhu city officials and again used drones and artificial intelligence to automatically estimate the allocation of the reconstruction needs. More specifically, the data accumulated from drone-powered aerial mapping of the region was processed using advanced machine learning algorithms. Combining drone imagery, digital mapping, and machine learning, the team configured regional modelling and infrastructure development with higher accuracy. Given the number of complex variables that need to be considered when forecasting earthquakes, such as the position of the tectonic plates or the type of ground involved, researchers concluded that machine learning is best suited to analyze these complicated seismic signals. By using AI and machine learning, resilience can be built against future earthquakes on a global scale, protecting lives, properties, and ultimately improving the livelihood of communities in the developing world.

Further reading

The International Telecommunication Union's (ITU) [Global AI repository](#) contains information on projects, research initiatives, think-tanks and organizations that are leveraging AI to accelerate progress towards the SDGs.

The 2018 [United Nations Activities on AI](#) report lists all AI-related initiatives implemented by UN agencies.

[IDIA AI & Development Working Group - Landscape Mapping](#) – A mapping of international actors involved in AI and development cooperation by the International Development Innovation Alliance (IDIA).

Multimedia material

AI

- ▶ ITU: "AI for SDGs - How can Artificial Intelligence address humanity's greatest challenges?"
- ▶ UNESCO series on artificial intelligence: "Nnenna Nwakanma: Artificial Intelligence, Youth, and the Digital Divide in Africa"
- ▶ Microsoft's AI for Good initiative
- ▶ Google AI Impact Challenge: "Using Technology to Change the World (Google I/O'19)"

Big Data

- ▶ UN Global Pulse: "PulseSatellite: A collaboration tool using human-AI interaction to analyse satellite imagery"
- ▶ Orange for Development: "How can Big Data encourage societal development and well being?"
- ▶ GIZ (German Corporation for International Cooperation): "Dark Days or a Brighter Tomorrow? How Big Data, Open Algorithms, and AI May Affect Human Development"

References

European Commission, (2019a). [A definition of AI: Main capabilities and scientific disciplines](#). High-Level Expert Group on Artificial Intelligence.

European Commission, (2019b). [Ethics guidelines for trustworthy AI](#). High-Level Expert Group on Artificial Intelligence.

Craglia M. (Ed.), Annoni A., Benczur P., Bertoldi P., Delipetrev P., De Prato G., Feijoo C., Fernandez Macias E., Gomez E., Iglesias M., Junklewitz H., López Cobo M., Martens B., Nascimento S., Nativi S., Polvora A., Sanchez I., Tolan S., Tuomi I., Vesnic Alujevic L., (2018). *Artificial Intelligence – A European Perspective*, EUR 29425 EN, Publications Office, Luxembourg, ISBN 978-92-79-97219-5, doi: 10.2760/936974, JRC113826.

European Commission, (2020). [A European strategy for data](#). Communication from the Commission to the European Parliament, the European Economic and Social Committee and the Committee of the regions. COM(2020) 66 final. Brussels.

European Commission, (2020). [On Artificial Intelligence - A European Approach to Excellence and Trust, White Paper](#). COM(2020) 65 final.

Gholami, S., (2018). [Spatio-Temporal Model for Wildlife Poaching Prediction Evaluated Through a Controlled Field Test in Uganda](#). Association for the Advancement of Artificial Intelligence.

Kshirsagar, V., Wiecek, J., Ramanathan, S. & Wells, R., (2017). [Household poverty classification in data-scarce environments: a machine learning approach](#). 31st Conference on Neural Information Processing Systems, Long Beach, CA. USA.

Chambers, M., Doig, C. & Stokes-Rees, I., (2017). [Breaking Data Science Open- How Open Data Science is Eating the World](#). O'Reilly Media, Inc. Gravenstein Highway North, Sebastopol, CA 95472.

Databricks, (2017). [The Democratization of Artificial Intelligence and Deep Learning](#). Apache Software Foundation.

Loukides, M., (2010). [What Is Data Science?](#) O'Reilly Media, Inc., Gravenstein Highway North, Sebastopol, CA 95472.

Maheshwari, A.K., (2016). *Big Data: Made Accessible: 2020 Edition*, 333 pp, ISBN-10: B01HPFZRBV

Parkash, O., (no date). [Scope of Artificial Intelligence](#), Lesson No. 01, Paper Code: MCA 402.

SAS Institute, (2021). [Artificial Intelligence – What it is and why it matters](#).

European Commission, (2021). [Big data - Shaping Europe's digital future](#).

European Commission, (no date). [Policies on Big Data – Shaping Europe's Digital Future](#).

European Commission, (no date). [Building a European Data Economy](#).

Elite Data Science, (2020). [Data Cleaning \(Chapter 3\)](#), Data Science Primer

Sharma, G., (2016). [Armed with Drones, Aid Workers Seek Faster Response to Earthquakes, Floods](#). Thomson Reuters Foundation.

Microsoft News Centre India, (2017). [Digital Agriculture: Farmers in India are using AI to increase crop yields](#).

Yao, S., Zhu, Q., & Siclait, P., (2018). [Categorizing Listing Photos at Airbnb](#).

Global Partnership for Sustainable Development Data, (no date). OPAL Case Study: [Unlocking Private Sector Data](#).

Melnichuk, A., (2020). [How Big Data and AI Work Together](#). Ncube.

Casey, K., (2019). [How Big Data and AI Work Together](#). The Enterprisers Project.

Vuleta, B., (2020). [How Much Big Data is Created Every Day?](#) Seed Scientific.

Sittón-Candanedo, I. & Corchado Rodríguez, J., (2019). [An Edge Computing Tutorial](#). Oriental journal of computer science and technology. 12. 34-38. 10.13005/ojcs12.02.02.

Gurcan, F. & Berigel, M., (2018). *Real-Time Processing of Big Data Streams: Lifecycle, Tools, Tasks, and Challenges*. 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMIT). DOI: 10.1109/ISMIT.2018.8567061.

European Commission, (no date). [Artificial Intelligence, Case Study Summary. Internal Market](#). Industry, Entrepreneurship and SMEs

European Commission, (no date). [Cloud Computing, Shaping Europe's Digital Future](#).

Mills, T., (2019). [Five Benefits of Big Data Analytics and How Companies Can Get Started](#). Forbes Technology Council.

Yeung, J. (2020). [What is Big Data and What Artificial Intelligence Can Do? Towards Data Science](#).

Hansen, M., Potapov, P., Moore, R., & Hancher, M., (2013). [The First Detailed Maps of Global Forest Change](#). Google AI Blog.

Page, V., (2020). [What is Amazon Web Services and Why is it so Successful?](#) Investopedia.

OPAL, (2017). [What is the Open Algorithms \(OPAL\) Project?](#) Paris 21.

Das, A., (2019). [AI for Earthquake Damage Modelling. Towards Data Science](#).

Shiwakoty, S., (2019). [Tools for Better Seismic Detection](#). The Katmandu Post

World Economic Forum (no date). [Fourth Industrial Revolution for the Earth](#).