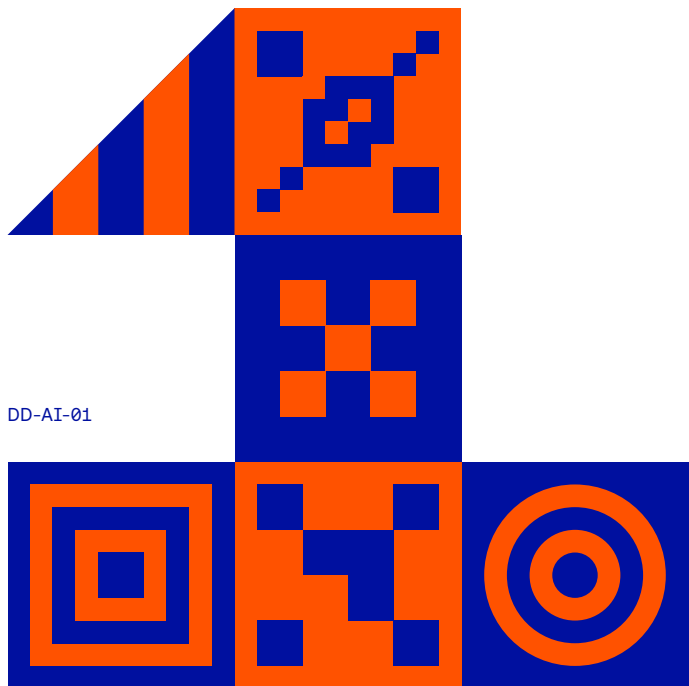


Some basic concepts



DD-AI-01

Artificial Intelligence and Human Rights

Some basic concepts

This publication was prepared by the Derechos Digitales team, coordinated by Jamila Venturini, Juan Carlos Lara and Patricio Velasco. This edition has been reviewed, updated and translated.

For more information on this project, visit <https://ia.derechosdigitales.org>

Text by Patricio Velasco and Juliana Guerra.

Editing and proofreading by Juan Carlos Lara and Ileana Silva.

Design and layout by Comunas Unidas.

Translation and adaptation by Urgas Traduc.toras.



Artificial Intelligence (AI)

System for analysis and decision-making using computerized and automated operations, based on a dataset. The term Artificial Intelligence is used for different technologies that share the feature of simulating human intelligence, applied to problems to offer solutions based on prediction or pattern recognition.

AI is considered "weak" or "narrow" when it executes a specific task, and "strong" or "general" if it contextualizes different specific problems and executes independent tasks based on that contextualization, similar to human intelligence.

Some AI systems, with differing levels of complexity and autonomy, are present in applications such as virtual assistants and voice recognition systems, or in specialized devices like drones or self-driving cars.

Algorithm

Formula or set of rules for articulating instructions, procedures or processes to solve a problem or perform a task. In the world of AI, algorithms are incorporated into a computer system's programming code to operate in an automated manner and thus orient the machines on how to search for answers to a question or solutions to a problem.

Problems associated with decision-making tend to be attributed to poorly designed algorithms or formulas that consider or weight data with erroneous, discriminatory or harmful results.



Data

Units of information, alphanumeric representations of singular facts or statistics that can be analyzed by different means. They are the essential ingredient for machine learning and the application of AI. For this reason, it is commonly thought that by using a larger volume of data, automated systems can be better trained to make more accurate decisions.

Personal Data

Unit of information that, on its own or in relation to other information, makes it possible to identify a physical person. Personal data include the following: name; legal identification number; telephone number; age; home or work address; etc. The anonymization of personal data is a process that is often needed for appropriate handling.

Open Data

The practice of keeping a dataset or database available for any person to freely use, reuse or share, with no restrictions related to confidentiality, copyright or patents. Data are considered usable when they are in a common format that can be read and processed by machines.

Dataset / Database

Groups of elements that are interrelated and store data such as numbers, dates or words that can be processed to produce information.



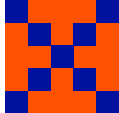
The quality of a database is directly proportional to the capacities available for building it, i.e., for gathering the data in an inclusive and rigorous manner. Therefore, any incompleteness (e.g., regarding people who live in extreme areas), inaccuracy, obsolescence or collection bias will be reproduced in the dataset.

Big Data

Management of databases that are too big or complex for traditional processing or storage methods, due to which their handling requires a different approach. Data with many cases (rows) offer greater statistical power, while data with greater complexity (more attributes or columns) can lead to a higher error rate if they are not processed with suitable tools. To address possible risks, it is considered relevant to address dimensions such as the volume of data, velocity of generation, variety of origin, the importance of the veracity of data and value assessment. This is the model known as the 5V of Big Data.

Data Mining

The process of collecting or extracting data from different sources, usually using automated systems, with the goal of building databases or identifying patterns in the information. Although it is a form of information processing, it can also serve for making inferences or identifying people against their will.



Data Science

Scientific field that includes disciplines like mathematics, statistics, probability studies, computing and data visualization to extract knowledge from a heterogeneous dataset (images, sounds, texts, physical measurements, genomic data, social media links, etc.). The methods and tools stemming from the application of AI are part of this field.

Open Source Code

A development model that consists of guaranteeing that people can freely access the source code of a program or application, facilitating its auditability, integration with other systems, reuse and adaptation in contexts that differ from the original. It is contrasted with closed source code, which for intellectual property or industrial secret reasons is kept confidential or with no possibility of modification by people who are not expressly authorized, which limits its auditability and scrutiny.

Cloud Computing

Set of technologies that facilitate remote access via internet to computing resources such as programs and applications, file storage and data processing. In this way, resources are stored in large data processing centers and therefore do not require local servers. Installing programs on end computers is also unnecessary, enabling remote service provision of "software as a service." It is also used in developing the Internet of Things (IoT), machine learning systems and Big Data.



Application Programming Interface

Mechanism that enables programming of additional components of a computer system, or making it interact with other programs, based on a set of operational definitions and communication protocols.

Bot / Chatbot

Computer program with a different level of complexity that is able to interact with individuals through an audio or text conversation, normally used to facilitate a "dialog" between a person and an institution or system.

Machine Learning

Branch of AI dedicated to the development of techniques and algorithms that enable a system to improve its performance ("learn") in resolving a proposed question or problem. This performance improvement is achieved based on inferences and in an automated manner, through repeated application and the accumulation of information.

Unsupervised Learning

A form of machine learning in which the system does not receive information on what the resulting data should look like, it just integrates them into an unstructured dataset. The algorithm must identify possible patterns in the data and relationships between them, and as a result, propose a structure, without requiring human validation.



Supervised Learning

A form of machine learning in which people provide the input and output data and how these should be used, so that the machine can understand how to relate them and propose improvements that will be validated by another person.

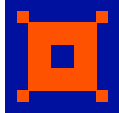
Deep Learning

A branch of machine learning that seeks to reproduce the functioning of the human nervous system. It consists of the analysis of large, unstructured datasets to find patterns with no need for prior training. To do this, it operates with layers of processing that are connected, for example, like neural networks: the result of the first layer feeds the second layer and so on, producing increasingly abstract results.

These algorithms have numerous applications such as automatic sound recognition, the development of computer vision systems, natural language processing, medical diagnoses, self-driving cars, etc. It is important that all risks of the environments in which these algorithms are deployed be considered.

Predictive Analytics

Method based on the observation and study of currently available data to predict the likelihood that events will happen in the future. It has been implemented in public policies that seek to prevent the violation of rights or to allocate social benefits. However, its results are debatable due to possible biases in



available data and opacity in processing such data. For this reason, it must be subject to broad scrutiny.

Model

An abstract representation of what a machine learning system has learned in training based on a dataset. For example, a climate data analysis model can be programmed to establish probabilities that inform daily forecasts.

Decision Tree

A type of predictive modeling for conducting a decision analysis where, through supervised learning, the data are iteratively divided based on predefined parameters or criteria. The goal is to teach the machines to make decisions and solve categorization or regression problems to obtain a definitive model.

Logistics Regression

A statistical process implemented in machine learning to predict the result of a dependent variable based on analysis of prior data. Through a supervised learning process, it is used for binary classification, i.e., one result or its opposite. One practical application is spam detection in an email service.

Neural Networks

The operational foundation of deep learning that consists of imitating the human brain's own neural functions. The neural



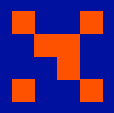
network works with layers of processing, where each neuron is a simple mathematical equation that is connected to another, and this is repeated throughout the layer. Thus, in each layer, unnecessary information is eliminated, and an increasingly simple, precise representation of the data is preserved. This kind of algorithm requires large datasets and facilitates the simultaneous modeling of multiple results. These networks can be used in tasks such as the automated detection of human faces ("facial recognition") or character reading.

Natural Language Processing (NLP)

Branch of AI applied to systems with which people interact, such as voice command assistants and those with verbal responses. It is focused on the development and optimization of programs that process human language with the goal of inferring its content. NLP includes Natural Language Understanding, which consists of transforming human language into a machine-readable format, and Natural Language Generation, which involves the production of spoken or written narratives based on a dataset.

Computer Vision System

An interdisciplinary scientific area dedicated to the exploration, recognition and analysis of images and videos. The tasks addressed include object recognition, event and movement detection, and object following, among others. Computer vision techniques are used in applications like the object detection



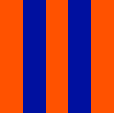
of self-driving vehicles and automated facial recognition, due to which its poor development can have catastrophic consequences, and great improvement can facilitate the work of massive or discriminatory surveillance.

Data Brokers

People and companies dedicated to marketing personal data obtained from the use of internet-based services or other sources (including other data brokers). They are also geared to creating profiles (e.g., for guiding marketing campaigns), for which they gather and analyze user information obtained from public and private sources. Although they concentrate a large volume of personal information, data brokers tend to operate without being detected by technology users, who have no control over the quality of these data or over how many people or companies have them.

Algorithmic Bias

The offset value from the origin of a formula, a defect that leads to an incorrect or unjust result. In machine learning systems, given the social effects that bias can have today, it must be understood in the context of the ethics of Artificial Intelligence, with at least two meanings: inductive bias, i.e., the stereotypes or prejudices of the person designing a data collection or processing system, which are incorporated in the system's operation; and confirmation bias, which is a systematic error introduced by a sampling procedure or the

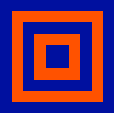


available information, i.e., they are biases that may exist in the conformation of the database. To reduce these biases, it is possible to use a supervised learning system to prevent the generalization of an algorithm's biased results. However, a system must be regularly audited to prevent new biases from appearing.

Algorithmic Discrimination

The inappropriate, incorrect, irregular or unjust result produced following the operation of an automated prediction, classification or decision-making system, which tends to harm certain individuals or groups. Algorithms often have biases, since in both the definition of variables and the development of coding, the values of the people responsible for their design are imprinted. There can also be biases in the dataset with which the algorithms are trained, depending on the criteria with which those data were collected and processed. This is especially sensitive in machine learning systems for decision-making and predictive models, including natural language processing and computer vision systems, when they are implemented in public administration or social welfare; the selection and hiring of personnel; allocation of loans or social benefits; and law enforcement, among other tasks.

Algorithmic predictions can reinforce discrimination toward traditionally vulnerable groups, based on socioeconomic status, income, race, ethnicity or gender, among others. These processes have been characterized, up to now, by their opacity.



For this reason, mechanisms like explainable AI have recently been proposed for the verification, transparency and regulation of AI systems, in order to reduce the disproportionate or unjust effects of these systems.

Algorithmic Transparency

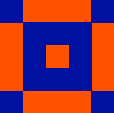
Principle targeting the factors that determine decisions made by algorithms operating in a transparent manner that is understandable to those who use, regulate and are affected by the data processing performed. It involves a fundamental principle for detecting operating problems and facilitating accountability in the operation of automated systems.

Personal Data Protection

Rules that regulate people's right to their own information, in particular regarding the collection, handling and processing of these data to know the details of that use, to correct errors or to request their elimination. The specificity and scope of these rules vary depending on each country's legal framework. Their effectiveness is a crucial element for control over a fundamental element of automated decision-making: information related to individuals.

Legality

Fundamental principle of public rights, orienting the operation of public administration strictly under the framework of legal regulations defined by competent agencies. It implies that all



public decisions must respect formal competency rules, as well as substantive rules on basic and legal rights.

Need and Proportionality

The principle of need is considered a principle of public administration which stipulates that the measure implemented be required for protecting a right or meeting a legitimate objective, after evaluating the possibility of other equally effective, but potentially harmful, decisions or solutions.

The analysis of need is joined by the analysis of proportionality. Proportionality is a principle of public administration that requires balanced weighting among measures adopted to obtain legitimate ends, such that the achievement of those ends does not have greater negative consequences. This entails analyzing the suitability of the measures for attainment of the ends.

In public decisions related to AI, the application of these principles involves an informed decision on the suitability of an automated system to assist in the attainment of public administration, the interests that will be affected by the system's deployment, the amount of information strictly necessary for its operation and the existence of

Ethics of Artificial Intelligence

Sets of principles that try to influence the development and deployment of AI systems to prevent their risks. Thus, to reduce the risks of algorithmic discrimination, different companies and



institutions around the world have ethical principles for the design, training and implementation of AI and machine learning systems. While dozens of initiatives on ethics and principles exist, in November 2021 UNESCO published a series of Ethical Recommendations, which have been adopted by 193 member States. Although they are not binding, up to now they are the most broadly accepted regulations at the global level.

Explainable AI (XAI)

A set of processes and methods that enable users to understand the results and products created by machine learning algorithms. XAI consists of overcoming the idea of a "black box" where AI systems are developed, and it is critical for generating trust in these systems by describing a model, its expected impact, its possible biases and reparation mechanisms in light of potential abuses. It facilitates the systems' auditability and also enables those who develop systems to fine-tune them into fairer models.

DD-AI-01



This work is available under Creative Commons Attribution
4.0 International license.

<https://creativecommons.org/licenses/by/4.0/deed.en>

