Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Home Affairs FDHA

Federal Statistical Office FSO
Data science, AI and statistical methods
Competence Network for Artificial Intelligence (CNAI)

Terminology

Competence Network for Artificial Intelligence (CNAI)

Author	CNAI unit
Version	2.1
Date	21.12.2023

CNAI unit
Federal Statistical Office FSO
Espace de l'Europe 10
CH - 2010 Neuchâtel

CNAI.swiss
CNAI@BFS.admin.ch

TABLE OF CONTENTS

1. AIM AND SCOPE OF THIS DOCUMENT 3

2. TERMINOLOGY RELATED TO DATA (SOURCES) 4

3. TERMINOLOGY RELATED TO 'LEARNING FROM DATA (SOURCES)' 6

4. ADDITIONAL TERMINOLOGY 9

5. CORE PRINCIPLES FOR HUMAN-CENTRED AND TRUSTWORTHY DATA SCIENCE (AND AI) 11

1. Aim and scope of this document

The creation of a 'Competence Network for Artificial Intelligence' (CNAI) is intended to promote the use of artificial intelligence and other new technologies within the Federal Administration (and beyond) in the long term.

One of the aims of the competence network is to introduce uniform terminology to ensure a common understanding of the terms relevant to artificial intelligence and new technologies within the Federal Administration.

A common language facilitates the active exchange of experience and knowledge both within the competence network and beyond. It also simplifies the communication of ideas, plans and services in this area.

Terminology management by the CNAI unit thus represents a central corner-stone for the functions of the CNAI in general and in particular for its unit.

This terminology management is based on the documents already available in the FSO and its 'Data Science Competence Center' (DSCC) and also on the documents that were drawn up within the framework of the Federal Council's decision to set-up the CNAI network, including the unit.

Additions to the terms already explained should only be made if they are necessary for the delineation of further new technologies. Together with CNAI's network nodes the glossary will be further developed in this area to illustrate the CNAI's future terminology in a targeted way.

The CNAI unit is the owner of the document.

This document is revised periodically and the current version (version number and date for reference) is valid.

Writing conventions

If a term has not been explicitly defined, it is placed in inverted commas, e.g. 'data science'.

2. Terminology related to data (sources)

In general, there are **three different types of data sources**:

- internal administrative (e.g. 'statistical data'[1], administrative data, official data, federal geodata, 'Open Government Data' - OGD[2]) and
- external administrative data sources (e.g. open data) and
- cyber-physical systems (such as data from networked sensors in the context of the 'Internet of Things' - IoT), which can be both an internal and external administrative data source.

The integrated use of internal and external administrative data sources, together with data from cyber-physical systems, opens enormous potential for a data (r)evolution for policy makers. It is with a view to using this data potential by applying 'data science methods' that a vast application potential for data science can be tapped into. For example, data sources can be merged to meet previously unmet user needs or to better meet user needs that are only basically covered. This opens new possibilities for generating relevant information on various problems as promptly as possible and making it available to policy makers.

In general, there are also **three different types of data**:

- structured data that have been formatted and converted into a well-defined data model. The raw data are mapped into predefined fields, which can then be easily extracted and read, e.g. via SQL. Relational SQL databases, consisting of tables with rows and columns, and classic tables with rows and columns are examples of structured data; and
- semi-structured (or partially structured) data, which have some consistent and unique characteristics without being limited to a rigid structure, such as is required for relational databases. Pictures are one example of this: for instance, when an image is taken by a smartphone, it has some structured attributes such as geotags, device IDs and timestamps. After they have been saved, images can also be assigned tags such as 'pet' or 'cat' to provide a structure; and

[1] Statistical data are all data that have been collected or passed on for statistical purposes, notably based on the Federal Statistics Act, and are thus subject to statistical secrecy.

[2] Open data are data that can be freely used, processed, analysed and transferred without any significant legal, financial or technical limitations. From a legal point of view, the use and processing of data must be free of charge; and from a technical point of view, it must be possible to process the data by machine. When data are published as open data, provisions governing data protection, information protection and copyright must be observed.

- unstructured data are data that are available in absolute raw form. These data are difficult to process due to their complex arrangement and formatting. Unstructured data storage can include data in many forms, including social media posts, chats, aerial and satellite imagery[3], IoT sensor data, emails and presentations.

[3] Aerial and satellite images can be in structured form (measurement matrix perspective) as well as in unstructured form (uncalibrated images).

3. Terminology related to 'learning from data (sources)'

Term	Definition
Data science	Data science is the interdisciplinary **science of learning from data** (data understanding), with the aim of gaining insights from the data to facilitate data-based decision-making. Data science covers, like statistics, the entire process of problem formulation, data collection, data selection, data preparation, data analysis, evaluation, interpretation, communication and provision of the findings. In contrast with traditional (and advanced) statistics, however, data science uses an **inductive approach**, starting with the data ('data first'). At the core of data science is thus a **problem-solving process** and a process of continuous improvement that aims to solve complex, unstructured and data-rich problems through the application of innovative data science methods (e.g. methods from 'machine learning' and the field of 'artificial intelligence'), techniques and practices. Data science is also referred to as the '**child of statistics and computer science**'. This child metaphor appropriately infers that data science inherits (ideally the best) from both parents, but eventually grows into its own entity. Its focus separates it from its parents.
Artificial Intelligence (AI)	'Artificial Intelligence' **(AI)** now sometimes called 'machine intelligence' is defined as '**building or programming a computer to do things that normally require human** or biological **skills** ('intelligence'), e.g. visual perception (image recognition), speech recognition, language translation, visual translation and playing games (with concrete rules). AI is about 'intelligent' machines ('smart machines') that can perform tasks which are normally performed by humans ('learning machines'), i.e. making machines 'intelligent'.

Term	Definition
AI system	An AI system is a machine-based system that, for explicit or implicit objectives, infers from the input it receives how it can generate output, such as predictions, content, recommendations or decisions, which can influence physical or virtual environments. AI systems can be equipped with varying degrees of autonomy.
AI decisions	AI decisions are conclusions from AI systems with real-world implications. These are dependent on human decisions at the level of the design of the system, the strategic level (decision on the use of the system) and the tactical level (design of interaction with the person using the system).
AI technology	AI technology refers to individual functions that can be implemented in computers to achieve AI (e.g. 'machine learning'). An AI system thus refers to a structured, context-bound combination of AI technologies for the purpose of achieving AI.
Natural Language Processing (NLP)	'**Natural Language Processing**'**(NLP**) is a branch of AI that deals with the analysis, understanding and generation of written and spoken words and sentences (natural language). Most NLP techniques and methods are based on machine learning, extracting meaning and context from human language. Areas of application include text recognition, speech recognition, bots, chatbots and digital assistants.
Generative AI	'Generative AI' is a broad term that refers to AI systems that are trained on large amounts of data from the physical and virtual world in order to generate data themselves (e.g. texts, imagery, sound recordings, videos, simulations, and codes). They are often multimodal, e.g. with input and/or output in one or several modalities (e.g. text, image or video). Different model architectures, including diffusion models and transformer-based models, can be used for generative tasks.

Term	Definition
Machine Learning (ML)	'Machine Learning' (ML) is another area of AI which 'gives computers the ability to learn'. ML explores the construction of algorithms that analyse data through the use of computers, learning, adapting and improving automatically (based on specific rules given by humans). The resulting statistical model enables, for example, predictions and classifications of (not yet seen) data, which can be used to support decision-making. In AI, ML is the most important subfield that deals with the problem of inductive reasoning. ML thus takes an inductive approach that starts with data ('data first'). ML can be roughly divided into three different sub-areas: supervised, unsupervised and reinforced learning. This separation is useful to understand the basic ML building blocks. However, current research is investigating how to combine the different types of ML.

ML can be roughly divided into three different sub-areas: supervised, unsupervised and reinforced learning. This separation is useful to understand the basic ML building blocks. However, current research is investigating how to combine the different types of ML.

Term	Definition
Supervised learning	In supervised learning, the ML algorithm is trained with data, and a human observer or benchmark data monitor the success of the learning process. Specifically, ML algorithms are trained in an iterative process using a training data set that contains a large number of 'features' ('inputs') and a 'target variable' ('output' such as 'correct' or 'incorrect'). Here, the algorithm automatically learns (complex) patterns between the 'features' and the target variable without explicit (human) instructions. With the help of the learned patterns, predictions or classifications can be made for (not yet seen) data. For example, '**deep learning**' is a sub-division of supervised learning and uses a specialised form of ML algorithms from the group of 'artificial neural networks'.

Term	Definition
Unsupervised learning	Unsupervised learning aims to find correlations and dependencies in data and, if necessary, to further process them into new features. In unsupervised learning, the training data consists of 'inputs' without any indication of the expected system performance, i.e. there is no target variable (e.g. 'correct' or 'incorrect'). In contrast with supervised learning, the ML algorithm is thus only trained with 'features'. Based on the 'features', the algorithm independently discovers patterns in the data. This makes it possible to learn from invisible structures in the data (e.g. using 'clustering'). Unsupervised learning is similar to human learning; through observations, experiences and analogies.
Reinforcement learning	Reinforcement learning stands for a set of ML methods in which an 'agent' independently learns a 'strategy' to maximise rewards received. In this process, the 'agent' is not shown which action is best in which situation, but receives a reward at certain points in time, which can also be negative. Based on these rewards, it approximates an 'utility function' that describes what value a certain state or action has. Reinforcement learning finds its main application in planning problems such as navigation (e.g. the agent has to reach a goal and the reward can be linked to time or distance) and video games (e.g. the agent has to play and its reward is linked to the game's final outcome).

4. Additional terminology

Term	Definition
Blockchain	A blockchain allows a collective of actors to maintain a trustworthy shared 'ledger' even without a trustworthy central authority. Blockchain technology is also known as 'Distributed Ledger Technology' (DLT).
Chatbots	Chatbots are dialogue systems with natural language capabilities (text or audio). They are used, often in combination with static or animated avatars, on websites or in instant messaging systems, where they guide prospects and customers to relevant products or services or deal with their concerns and answer questions.
Cloud computing / Cloud services	'Cloud computing' is the provision of computing resources (e.g. servers, storage, databases, network components, software, analytical and intelligence functionalities) over the internet. As a rule, only cloud services *('*services'*)* that are actually used are paid for ('pay as you go' principle). As a rule, highly standardised services are offered as 'cloud services'. In return, these services are available extremely quickly (often within minutes).
Internet of Things (IoT)	The internet of things is a collective term for the technologies of a global infrastructure of information societies that enables physical and virtual objects to be networked and to work together through information and communication technologies.

5. Core principles for human-centred and trustworthy data science (and AI)[4]

Core principle	Definition and explanation
Data and information protection	The purpose of **data protection** ('privacy') is to protect the privacy of each and every individual. Data protection thus protects against the misuse of data ('informational self-determination') and answers the question of whether the collection and processing of certain data should be allowed. It defines the authorised use of this data. At the heart of data protection law is personal data, which is information relating to an identified or identifiable person. Data protection is the 'abstract' protection of personal data from misuse (c.f. 'practical' protection under 'Data security'). **Information protection** defines the 'abstract' levels of information confidentiality ('unclassified', 'internal', 'confidential', 'secret') in order to protect the interests of a country or an organisation.
Information security	**Information security** safeguards the integrity, availability and confidentiality of information and information-processing systems, regardless of how the information is presented, stored or used. This protects information from being lost, falsified or inappropriately disclosed.
Data security	**Data security** concerns the protection of data flows and use of all types of data, i.e. not just personal data. Technical solutions or organisational, personnel and local physical measures ensure data integrity, availability, confidentiality and authenticity. This is the 'practical' protection of data of all kinds.

[4] «Verhaltenskodex des Bundes für menschenzentrierte und vertrauenswürdige Datenwissenschaft (und KI)» (in German): https://www.bfs.admin.ch/bfs/de/home/dscc/dscc.assetdetail.29325686.html & https://www.bfs.admin.ch/bfs/de/home/dscc/dscc.assetdetail.28405274.html

Core principle	Definition and explanation
Data governance	While 'data management' implements the technical administration of data, **data governance** sets out the internal guidelines and procedures for managing data within an organisation. Data governance thus defines roles and responsibilities and establishes processes for the handling of data within an organisation. The aim of data governance is to ensure the quality, integrity and availability of data in an (information-processing) system. Data governance also includes the harmonisation and standardisation of data. At the same time, data governance should ensure that these data are used in accordance with current data protection laws and ethical standards.
Non-discrimination	**Non-discrimination** prohibits the discrimination of persons or groups of persons on the basis of certain characteristics. This includes discrimination or preferential treatment on the grounds of disability, gender, parentage, language, ethnic or social origins, genetic characteristics, sexual orientation, religion or ideology, as well as a person's political or other views. It is therefore a matter of protection from discriminatory unequal treatment, which is prohibited by law. This includes both direct discrimination and indirect (i.e. implicit or hidden) discrimination.
Explainability	**Explainability** refers to the capacity of a data science problem-solving approach to explain its results and findings in a way that people can understand. Specific requirements for explainability differ depending on the target group and context. These may include, for example, the obligation to disclose the algorithms and methods used as well as the data sources and processes on which they are based ('completeness'), to provide explanations that are clear and understandable to people ('interpretability'), and to continue improving the explainability (insofar as permitted under data protection law and other guidelines). Overall, explainability aims to ensure that data science problem-solving approaches are fair, trustworthy and acceptable, and that they are used in a way that is in society's interests.

Core principle	Definition and explanation
Traceability	**Traceability** refers to the ability to track, verify and correct the data-driven decision supports that are developed based on the application of data science methods, techniques and practices. Such decision supports (e.g. in the form of recommendations) should be documented in a clear, traceable and understandable manner and their impact on the context in which they are applied should be understood. Specific requirements for traceability differ depending on the target group and context. These may include, for example, the obligation to document data science problem-solving approaches, the provision of clear and understandable explanations of data-driven decision supports or the provision of feedback and corrections to ensure that decision-making processes are continuously improved and that data-driven decision sup-ports become increasingly accurate and reliable. Traceability thus aims to ensure that data-driven decision supports resulting from the use of data science are consistent and reliable and that they fulfil the needs and requirements of the context in which they are used.
Transparency	**Transparency** refers to the obligation to openly present and communicate the use of data science methods, techniques and practices and the associated problem-solving steps and decisions. Specific requirements for transparency differ depending on the target group. For example, they may include the obligation to disclose data sources and processing, algorithms and methods or the obligation to explain decisions and processes (insofar as permitted by data protection law and other guidelines). The purpose of transparency is to give a clear picture of the entire data science problem-solving process and the individual steps of this process. This is to ensure that data science problem-solving approaches are fair and transparent and that the use of data science in the Federal Administration can gain the public's trust.

Core principle	Definition and explanation
Reproducibility	**Reproducibility** refers to the fact that others must be able to reproduce the results and findings obtained through the use of data science methods, techniques and practices. This includes an obligation to clearly document and communicate the data, data sources, algorithms and methods used, as well as the results and conclusions, so that others are able to independently verify and reproduce the results (insofar as permitted by data protection and other guidelines). It also includes an obligation to provide detailed information for reproducing the results in order to enable comparisons between different methods and approaches. Together with all the other core principles mentioned here, reproducibility therefore aims to ensure that data science problem-solving approaches are trustworthy und meaningful.
Neutrality	In general, **neutrality** means that a person or organisation acts impartially and without bias and has no prejudices or personal interests that could influence their decisions. This also applies to a person or organisation using data science. Neutrality is upheld, for example, by disclosing potential conflicts of interest.

Core principle	Definition and explanation
Objectivity	The aim of **objectivity** is to ensure that data science problem-solving approaches are unbiased and neutral and that decisions and results are based on objective and independent criteria. A person or organisation developing, using and providing data science methods, techniques and practices for problem-solving does so while respecting scientific independence and working in an objective, professional and transparent manner.
Ethical handling of data and results.	The **ethical handling of data and results** is intended to ensure that data and results are used in an ethical, responsible and sustainable manner and that the rights and dignity of individuals are respected while taking into account the impact on society and the environment. This includes, for example, the obligation to guarantee data protection and security, to encourage transparency and openness and to ensure that the use of data and results complies with ethical principles and standards that respect human dignity, autonomy and privacy. It also includes avoiding discrimination and prejudice as well as taking into account the impact of data and results of any data science problem-solving approach on stakeholders, society and the environment in general.