MIL GLOSSARY

Data, algorithms, artificial intelligence

With the availability of an ever-increasing mass of data and the widespread use of algorithms boosted with artificial intelligence, both the production and the consumption of information are changing rapidly. The challenge now is to take advantage of the new opportunities that are emerging for information and freedom of expression, while giving all citizens the means to fight against disinformation driven by algorithms and AI.

This critical glossary defines the terms before giving their interpretation in MIL (italics).

● Algo/AI- literacy
Algo-literacy is one of the components of Media and Information literacy. It enables citizens to better understand information or news shaped by algorithms and artificial intelligence. Algo-literacy covers the following skills: understanding algorithms, including those incorporating artificial intelligence; being able to analyse and criticise them; knowing how to use them wisely; making informed changes to behaviour and usage; and reacting to them in the mass and social media.

Since most algorithms now incorporate one or more intelligence systems, we also talk about Algo/AI-literacy. In this printable poster, you can view the 10 key points of algo-literacy according to Savoir Devenir.

● Algorithm
An algorithm is a finite, unambiguous sequence of operations or instructions. Algorithms are used to solve a class of problems or perform a task. For example, recommendation algorithms can be used to determine what people like so that personalised suggestions can be made to them.

In MIL, we are interested in algorithms that influence both the production of, and access to, information. And we are alerting people to algorithmic biases, due to poor quality databases, which can lead to sexist or racist representations of the world.

● Algorithm bias
An algorithmic bias is when the output from an algorithm treats situations or certain groups of people differently. It often occurs when the data used to train the machine learning system is biased . Sometimes an algorithm can be written that is intentionally biased.

In MIL, the most studied algorithmic biases are those that lead the media and social media to reinforce stereotypes, such as gender stereotypes, and promote discrimination.

● Artificial Intelligence (AI)
Strictly speaking, artificial intelligence is a field of research that studies the mechanisms of human intelligence and seeks to model them. Different AI systems attempt to imitate the human brain to help perform tasks or automate activities previously done by human s. These systems are in fact not at all intelligent by human standards and are most often based on machine learning algorithms.

In MIL, the emphasis is on understanding and mastering the AI systems that online media and social media use to control information flows and influence what users see and are invited to do on their platforms.

• Artificial media (or synthetic media)
Artificial media tend to use texts and/or images designed entirely by an AI developed for this purpose.

In May 2023, the first weather report 100% designed by AI and presented by an avatar was launched on a Swiss television channel. Artificial media are being added to the mass media and social media.

• Big D ata
The term Big Data refers to the collection, processing and storage of massive data, which is used to train algorithms. These techniques make it possible to analyse the statistical properties of very large databases. They are at the heart of our digital economy. Big data is characterised by the 5 Vs: velocity, volume, value, variety and veracity.

In MIL, we are alerted to the fact that many media services offer 'free' services in exchange for this precious data, without users being aware of all the consequences for their privacy.

• Black box
A black-box algorithm (for example, some of those used on Tik Tok) is an algorithm for which a user can only provide input data and watch the output, without knowing how the algorithm obtained the result.

In MIL, the issue is one of transparency. If even the creators of an algorithm no longer know how it works because it has evolved 'on its own', how can we expect them to explain to us how it works?

• Chatbot
A chatbot is a computer programme developed specifically to simulate and process a human conversation (written or spoken), enabling humans to interact with their machines as if they were communicating with a real person.

Equipped with high-performance artificial intelligence, a huge training database and a good interface, ChatGPT has popularised the use of this type of tool.

• Cookies
Cookies are small text files deposited on computers or mobile phones without the user's knowledge when visiting certain websites. They store information for subsequent visits to those websites .

Recommendation and prediction algorithms are fond of these cookies, which help them to 'profile' users. These algorithms use the data in the cookies to personalise or propose offers or information that are more likely to be successful.

• Data
Data is digital information (image, sound, text, video, signals) produced by an individual, community or institution. It can be qualitative, quantitative or technical. For example, a date, a place, a person's name or a photo are all data.

A legal distinction is made by the GDPR between personal and sensitive data, which is

protected, and all other data, which is not (such as data collected when surfing the Internet).

• Data Journalism
Data journalism dates to the late 1950s, when its original aim was to use statistical data in the interests of 'precision journalism'. In its broadest sense, the term now covers all journalistic practices that use data processing (fact-checking, information processing, data visualisation, robot journalism, audience research, etc.).

We need to look at the opportunities and risks associated with this new mode of journalism, and its repercussions on the profession of journalism.

• Datavisualisation
Datavisualisation - or dataviz - is a set of techniques for graphically summarising and presenting data. It is used to visualise important information and trends within a data set. In the media, dataviz is used to present large amounts of information in an attractive format that appeals to everyone.

IN MIL, we learn how to analyse these visuals without believing them... just because they're numbers and pretty.

• Deep learning
Deep learning is a sub-field of AI and a method of automatic learning based on the use of neural networks models with an extremely high number of parameters, requiring very large databases.

Deep learning is used today in machine translation, speech recognition and image search. *Deep learning is also used for visual deep fakes, misleading images or videos that are often very difficult to spot.*

• Attention Economy
In a world where we are saturated with information, according to the principle of supply and demand, our attention (and therefore the time we spend in front of an online document) is becoming a rare and therefore precious resource, particularly for selling advertising. Controlling this resource is at the heart of market competition.

The attention economy encourages the dissemination of all the sensationalist information that attracts attention. Ranking and recommendation algorithms are designed to bring us content we like... and just some of the techniques used to make us stay longer on the platforms.

• Filter Bubbles
The term refers to the mechanism used to filter the information received by Internet users based on their interests. It is the result of systems for personalising online content, based in particular on recommendation algorithms and the management of news feeds.

The impact of filter bubbles on information is currently the subject of debate: some believe that they reduce the diversity of information available to users and encourage misinformation. Others believe that the impact of these bubbles has been exaggerated. Either way, filter bubbles certainly don't encourage critical thinking!

• Generative Artificial Intelligence (GAI)
Generative AI refers to algorithms that can synthesise new content (text, images, music, code, etc.) from the big data bases on which they have been trained. Unlike most other AIs, which are used for analysis or decision support, generative AI aims to create realistic looking content autonomously, by imitating human creations. The best-known generative AI is the one used by Open AI in ChatGPT, which can create texts, images and videos from scratch based on requests

made by users of these services. There are many others!

In MIL, the challenges are to understand how these systems work, to know how to spot fake content and to know how to use them ethically and with integrity for information and creative purposes.

● Information overload
The term refers to the excess of information to which digital technologies subject us, and which we find difficult to process because there is too much of it. In the field of news, we also talk about information fatigue, a phenomenon that explains why many people stop keeping themselves informed.

Today, one of the key skills for staying well-informed is no longer 'finding information' but knowing how to sort and manage it.

● LLM (large language models)

LLMs are specific machine learning models of the neural network type used in generative AI. Their aim is to master the complexity of human language to be able to respond to all kinds of user requests, and also to be able to produce texts in natural language. LLMs work statistically by analysing masses of linguistic data, which they 'slice' into sequences called tokens to identify the most likely sequences in a given language or context.

LLMs could well radically change both information and creativity. Or not. It's a bit early to say.

● Metadata
Metadata is data that describes data. It is used to index, sort, analyse and facilitate information retrieval. For example, the name of the author of a book and its publication date; the format, definition and playback time of a video; the place and date the image was taken and the type of camera and lens used; the title of a web page, its type of coding, its author and the keywords associated with its publication, etc.

In MIL, knowing the nature of this metadata and how it is collected helps us to understand how the information presented to us is organised.

● Open data
Open data is digital data that users are free to access and use. It is private in origin, but usually publicly accessible, and can be re-used by anyone, without technical, legal or financial restrictions. Open data is widely used by journalists to conduct investigations.

Open data is a reminder of the diversity of data systems and the need to contribute to the 'information commons'.

● Prediction algorithms

These mathematical models are equipped with artificial intelligence that can learn from the past to make predictions about the future. For example, by studying the past performances of certain teams and their players against other teams, certain algorithms are used to try to predict the outcome of football matches.

While these types of algorithms are currently used mainly in the fields of health, science, fraud detection, economics and finance, sports betting and sales, in MIL we are also monitoring their use in terms of information.

● **Prompt**
A prompt is a short text (a question, a request, an instruction) that instructs an AI what is expected of it. The AI interprets the prompt and then generates a result in the form of text, an image, sound or video, depending on the type of AI.

One of the new essential skills in MIL is the ability to write effective prompts, the quality of which considerably improves the quality of the AI's response.

● **Ranking algorithms**
Used mainly by search engines, these algorithms select and list the information they deem most relevant in response to queries.

In MIL, we believe that these algorithms are not neutral and partly shape our representation of the world. We therefore encourage people to multiply and diversify the sources of their information rather than always asking Google for everything, for example.

● **Recommendation algorithms**
Recommendation algorithms are mathematical models that enable search results on search engines or social networks to be personalised by suggesting content that is supposed to correspond to users' tastes. They consider various elements such as your past browsing habits, your current location, the interactions you have had with your community, and the adverts you click on. It's the principle of 'if you liked this, then you'll like this too...'.

Information recommendation algorithms are under scrutiny insofar as understanding them and developing strategies to use them and not suffer harms from them is crucial, particularly on social media.

● **User Profile**
A user profile is a set of data and metadata that can be used to create categories of people or groups. IT systems process them in different ways, depending on their needs. For example, a media outlet will offer more 'right-wing' content to people who have been categorised as such. Similarly, a commercial site will promote certain products to certain profiles whose characteristics suggest that they will be interested.

It is important to know how to choose the type of data you want to be associated with your profile, and to make an informed decision as to whether you want to 'benefit' from personalised information feeds based on your profile. Or not.

● ***Token / Tokenisation***
In computer science , the term token can have various meanings depending on its function. In text analysis, it represents the lexical unit on which the model is based. For example, the LLMs at the heart of generative AI work by dividing texts into tokens. Depending on the model, these tokens can be words, sets of characters, combinations of words and punctuation.

No, GAIs do not understand us. They analyse our queries and produce texts by recombining, on a statistical basis, the lexical tokens stored and tagged in their databases, following... secret rules.