

To cite this article: Prod dr m s s el namaki (2025). The AI Data Syndrome. International Journal of Education, Business and Economics Research (IJEBER) 5 (2): 28-33

THE AI DATA SYNDROME!

Prod dr m s s el namaki

Retired Dean,
Maastricht School of Management, MSM, the Netherlands

<https://doi.org/10.59822/IJEBER.2025.5203>

ABSTRACT

Data lies at the heart of our artificial intelligence revolution. Massive volumes of data hold the key to the analytical processes that induce artificial intelligence outcomes. Data is drawn from a wide variety of sources. Yet data are neither homogeneous, generic nor mailable! Data are amorphous whole and carry with it a myriad of problems. This could and does induce "data malfunctions" or "data syndrome" and the derived issue of data reliability. The following article addresses this issue.

The article is qualitative in approach. It identified several triggers of data malignance going all the way from data deluge and data synthetics to data poisoning, data cannibalism and data multi science roots. It further identifies and analyses two prime sources of syndrome: data deluge and the multi science roots of data.

The article tries to place this observed data malignance within a framework. It identifies the sources and tries to develop a conceptual framework for the phenomenon.

The conclusion suggests a set of hypotheses that could be explored further.

KEYWORDS: - AI data, Data syndrome, data deluge, multi science data roots.

© The Authors 2025

Published Online: March 2025

Published by International Journal of Education, Business and Economics Research (IJEBER) (<https://ijeber.com/>) This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at: <http://creativecommons.org/licenses/by/4.0/legalcode>

I. AI DATA: THE ESSENCE

Data constitutes a prime input into AI systems.

Data is essentially "information in raw or unorganized form (such as alphabets, numbers, or symbols) that refers to, or represents, conditions, ideas, or objects.... Data is limitless and present everywhere in the universe. To computers, data include symbols or signals that are input, stored, and processed by a computer, for output as usable information. (Business dictionary, 2017).

Data segmentation follows different parameters. Most data can be categorized into 4 basic types from a machine learning perspective: numerical data, categorical data, time series data, and text data. "Data could also be segmented according to different criteria: useless, nominal, binary, ordinal, and count." (El Namaki, 2019) Time, interval, image, video, audio, and text (Jeff Hale, 2018).

Data could be logged or collected. Data logging, according to Technopedia (Technopedia,) is the process of collecting and storing data over a period to analyze specific trends or record data-based events/actions of a system, network, or IT environment. A data logger is an electronic device designed to measure and store data values, often independently of a PC.

Data collection consists of data acquisition, data labeling, and improvement of existing data or models. As AI technology advances it will develop an autonomous ability to seek, classify and validate data.

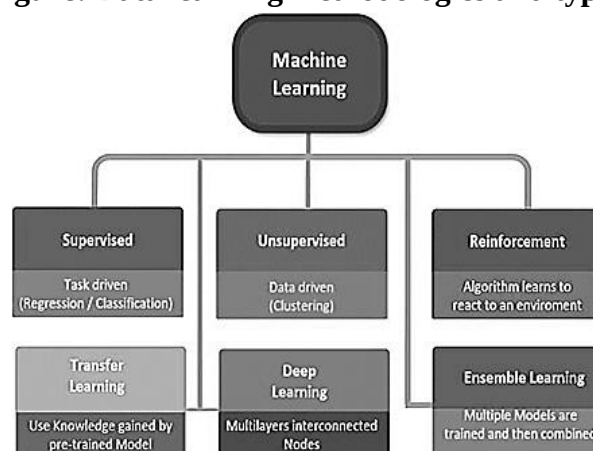
II. DATA ANALYSIS

Data holds the key to AI performance, outcomes, and system flows. Data analytics lead to key to those variables.

AI is either data centric or data driven.

Data-centric artificial intelligence (AI) is a methodology that prioritizes the management, quality, and relevance of data. The primary emphasis of a Data-Centric AI framework is to guarantee that the data utilized for training and modeling purposes is exhaustive, of superior quality, and in accordance with the goals of the AI system. Data-Driven AI is distinguished by its dependence on predictions, insights, and patterns that are extracted directly from data. Data-centric AI establishes the foundation for efficient modeling. Conversely, Data-Driven AI leverages this meticulously curated data to facilitate ongoing learning, adjustment, and enhancement.

Figure: Data learning methodologies and types



Source: Future directions, Journal of intelligent fuzzy systems, Jan 24

The problem here is that data centric and data driven AI rely on a wide variety of data sources. Those could suffer from many flaws and could induce many illnesses. The following is an analysis of five of those flaws: data deluge, data science roots, data synthetics, data poisoning and data cannibalism.

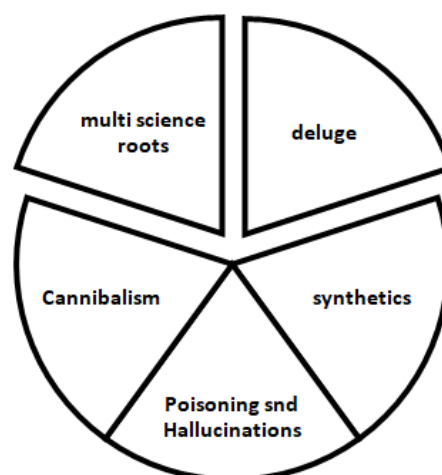
III. DATA SYNDROME

Data's far reaching and penetrative impact is derived from "data analytics," a process of deep analysis and exploration of, among other things, data content, formats, and properties.

Data analytics could suffer from structural flaws leading to malperformance and deviant outcomes. Five sources could trigger this data malaise: data deluge, data synthetics, multi science roots of data, data poisoning and data cannibalism (see figure). Two of those constitute, in the author's view, the prime triggers and the key damage inducers: data deluge and the multi science roots of the AI concept.

Those will provide the focus of the following analysis.

Figure: The multiple triggers of the data syndrome



- **The two prime sources of the syndrome:**

The following is a closer look at what the author considers the two-prime source of the syndrome: data deluge and the multi science roots of data.

1. DATA DELUGE

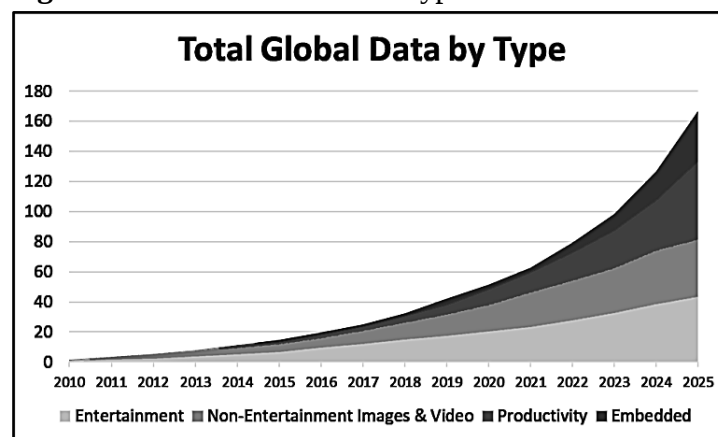
Data volumes constitute a prime threat to data use, data relevance, data insight generation and data application.

Data comes from a myriad of sources. It is a new energy source. It fuels innovation, drives decision-making, and empowers businesses but, also, creates the formidable challenge of managing a data deluge.

A data deluge is a situation where more data is generated than can be successfully and efficiently managed. The capacity to produce large volumes of data is growing faster than the infrastructure and the technologies required enabling data-driven performance. Data analysis is then undermined, and outcomes could also then lose productive applicability and proper focus.

Massive volumes of data describing patterns and shapes of decisions and behaviors are generated, daily, by “actors” and assembled by professional modes all around the globe. The challenge—and opportunity—for many industries and applications, is to access and analyze this mountain of data in a way that allows for parameters and key traits to be distilled and placed within a coherent context. This task of amassing all this data and examining it critically can be overwhelming. The myriads of sources and the myriads of shapes and formats that this data takes constitute a challenge to the processes of data aggregation, duration, ingestion, and standardization. Simply said, having access to data does not ensure that its value can be harnessed. For that to happen organizations need technology and skills to identify meaningful patterns and draw conducive conclusions (Spitzer, 2019).

Figure: Cumulation of diverse types of data 2010-2025



Source: Mark Spitzer, Data: A Twenty-First Century Arms Race, Amazon, January 2019.

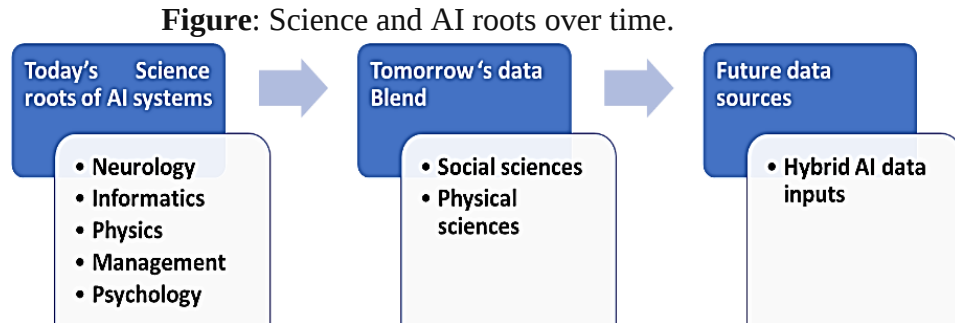
2. DATA MULTI-SCIENCE ROOTS.

It is the author’s contention that what we know today as artificial intelligence is in fact a blend of several sciences. Each science provides an input and each input blends with, yet another one drawn from another science. The synergy produces what we call today artificial intelligence. This hypothesis could be substantiated by an examination of the system’s approach to artificial intelligence (El namaki, 20). If one considers artificial intelligence as a system with inputs, transformation processes, output, and feedback, one can trace diverse science roots related to each of those system components. Inputs are made of biological and electronic components. Transformation is done by resorting to machine and deep learning analytical tools and their derived sub concepts or a blend of mathematics, statistics, and informatics. The outcome be it insights or sub technologies could also belong to psychology, strategy, or informatics.

Artificial intelligence /science interaction along the lines highlighted above could be looked at in terms of today's state of the art of both AI and the related sciences as well as today's configuration of each. The following matrix illustrates the relationship.

Today's relationship is straight forward. The initial stages of artificial intelligence, being rather basic, relate directly to elements of today's science that are equally basic and fundamental. The "reaction" and "correction" stages of artificial intelligence relate to basic elements of mathematics and statistics and derived software s. The advanced elements of artificial intelligence or those introducing mind-like analysis as well as awareness in the picture rely on advanced sciences and phases of sciences that are "not known yet." Mind-like analysis will relate to advances in neurology including inter neural communication and memory functions. It could also relate to new modes in physics allowing for faster data communication and novel data and memory interaction. Research is being conducted in this direction but the supportive stage to an artificial intelligence that creates a state of self-awareness of a state of biological storage and transmission of data is still a long way off.

The following figure illustrates AI science blend process overtime. Future sciences including physics, psychology and neurology will blend, over time, into a "socio-physical" body of knowledge that leads, again over time, to an advanced form of artificial intelligence. And a different blend of data inputs. Generative artificial intelligence that we are witnessing today could very well be the first phase in a long and complex phase.



- **Derived hypotheses**

The above analysis could lead to two hypotheses:

- Data volumes constitute a prime threat to data use, data relevance, insight generation and data application.
- Artificial intelligence is in fact a blend of several sciences. Each science provides an input and each input blends with, yet another one drawn from another science.

Each could provide a continuation for this phase of the issue.

IV. SUMMARY AND CONCLUSIONS

Data lies at the heart of our artificial intelligence revolution. Massive volumes of data hold the key to the analytical processes that induce artificial intelligence outcomes. Data is drawn from a wide variety of sources. Yet data are neither homogeneous, generic nor mailable! Data are amorphous whole and carry with it a myriad of problems. This could and does induce "data malfunctions". Sources of this data malfunction can go all the way from data deluge and data synthetics to the multi science roots of data, the poisoning of data and ultimate data suicide!

The article explores the broad framework of the syndrome, and it proceeds with a closer look at two prime sources: data deluge and the multi science roots of data.

Two hypotheses are suggested for further research.

REFERENCES

1. "What is data poisoning (AI poisoning) and how does it work?" ("What is data poisoning (AI poisoning) and how does it work? - TechTarget") (techtarget.com) Nihad Hassan 15 May 2024 Enterprise AI.com data poisoning (AI poisoning) By Nihad Hassan
2. Amartya Gupta, "Understanding Data Logging: Benefits and Best Practices," Motadata, March 10, 2022
3. April 2020
4. Business Dictionary, 2017.
5. Cohut T, "Empathy and the Historical Understanding of the Human Past,"
6. El Namaki M, "Systems Approach to the Artificial Intelligence Concept," 2024.
7. Future directions Journal of intelligent fuzzy systems, Jan 2024.
8. Hale J, "The Most in Demand Tech Skills for Data Scientists," KD nuggets, 2018
9. HBR White Paper | Navigating the Data Deluge: Turning Data into Insights across the Health Care and Life Sciences Industries.
10. Mark Spitzer "Data: A Twenty-First Century Arms Race," Amazon, January 2019.
11. What is data deluge, 2024 Hewlett Packard Enterprise Development LP