Data Science (and Artificial Intelligence) as a Problem-Solving and Continuous Improvement Process

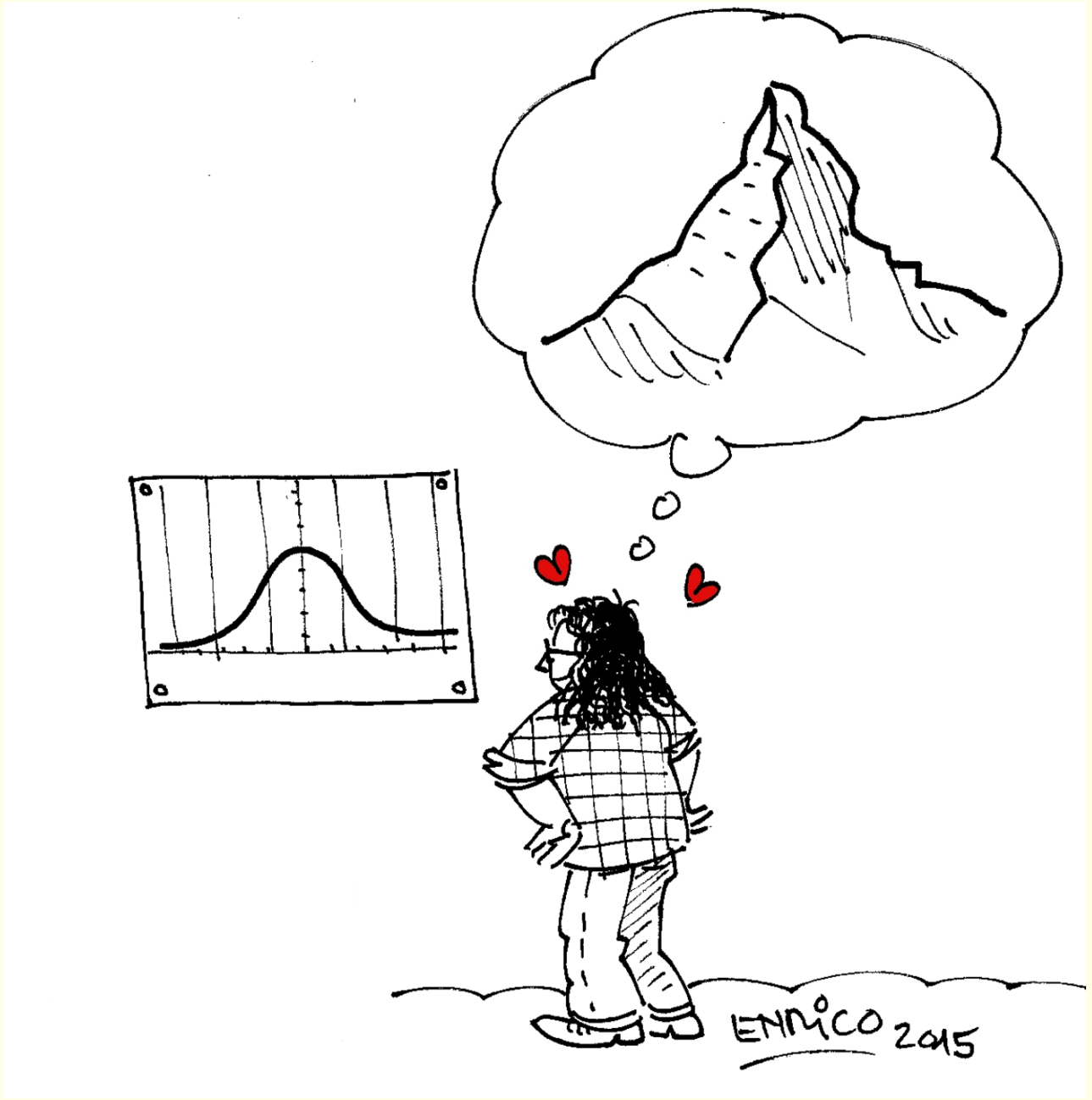
Demystification, Challenges, Opportunities, and Principles for Success

Prof. Dr. Diego Kuonen, CStat PStat, PhD EPFL

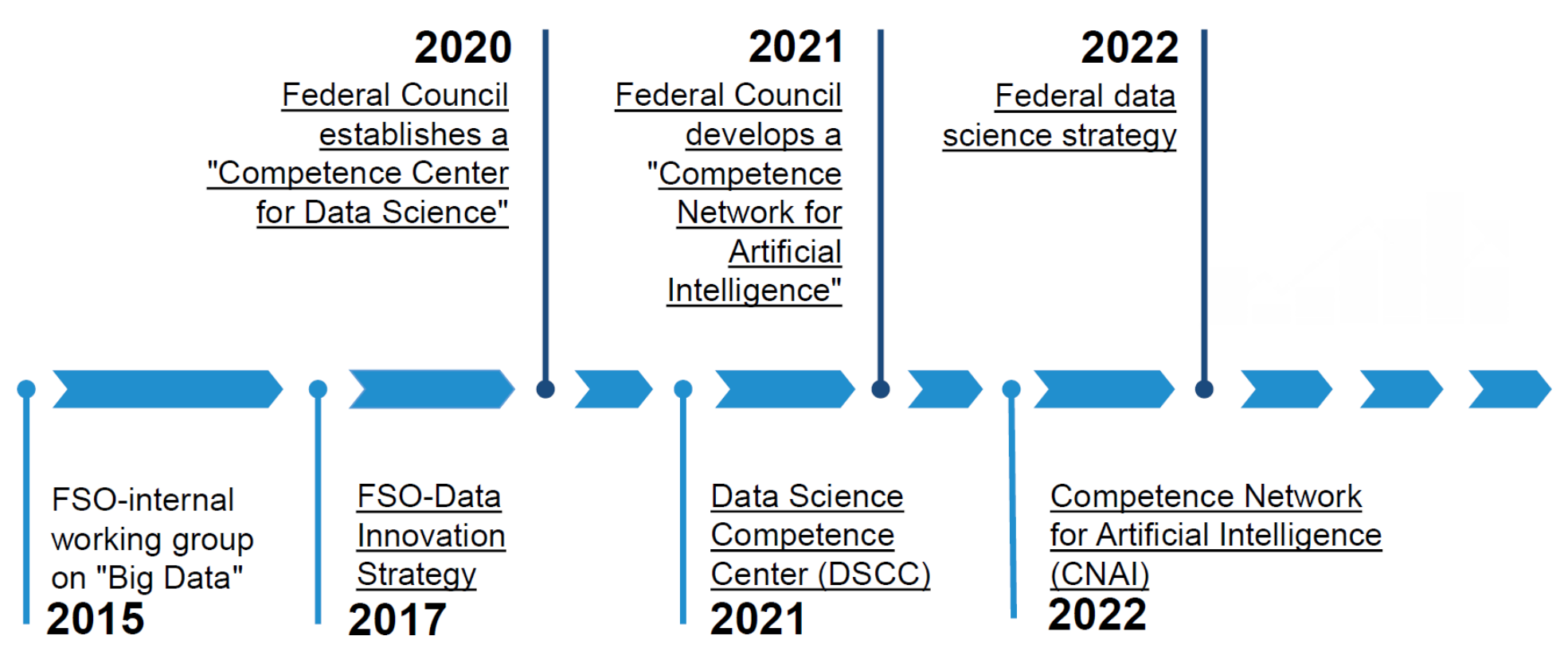
Statoo Consulting, Berne & GSEM, University of Geneva, Switzerland

@DiegoKuonen + kuonen@statoo.com & Diego.Kuonen@unige.ch

Keynote @ 'Geoinformationstag 2023', Berne, Switzerland — October 5, 2023



My advisory & consultancy journey @ 'Swiss Federal Statistical Office' (FSO) (since 2016)



‘Data are not taken for museum purposes; they are taken as a basis for doing something. ... The ultimate purpose of taking data is to provide a basis for action or a recommendation for action.’

W. Edwards Deming, 1942

$\rightsquigarrow$ **Data are the ‘fuel’ and ‘learning from data’ is the engine** of the digital transformation and the related data revolution!

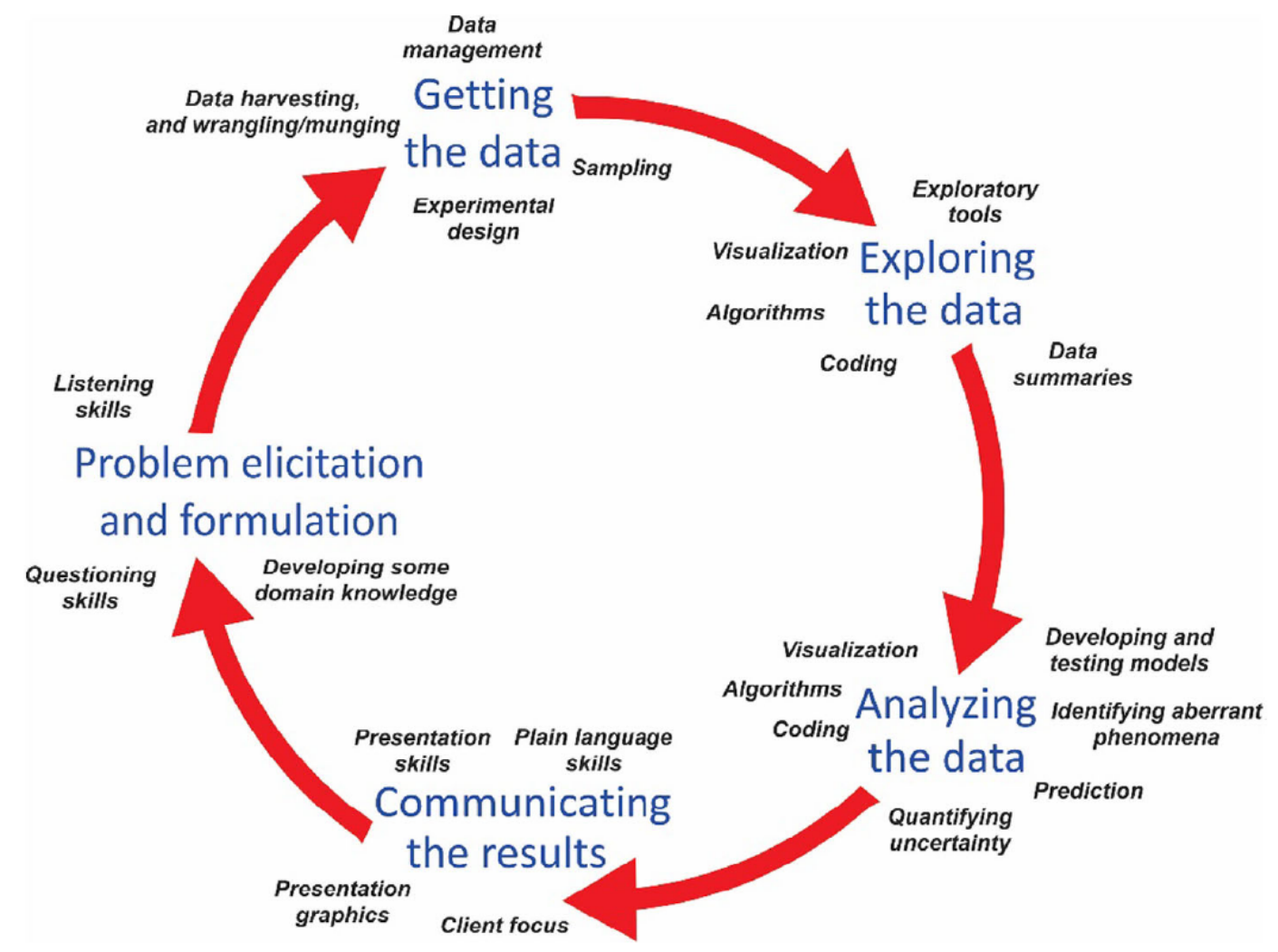
Data science is a process of data-driven problem solving!

◇ Data science, *i.e.* 'the science of learning from data' or 'the science of making sense out of data', is a whole iterative problem solving and continuous improvement process, aimed at solving large, complex, unstructured and data-rich problems sustainably.

'If you can not describe what you are doing as a process, you do not know what you are doing.'

W. Edwards Deming

The varying activities involved in (the science of) learning from data



Source: *'International Data Science in Schools Project' Frameworks v1.0*, September 2019 (idssp.org).

‘Coming together is a beginning. Keeping together is progress. Working together is success.’

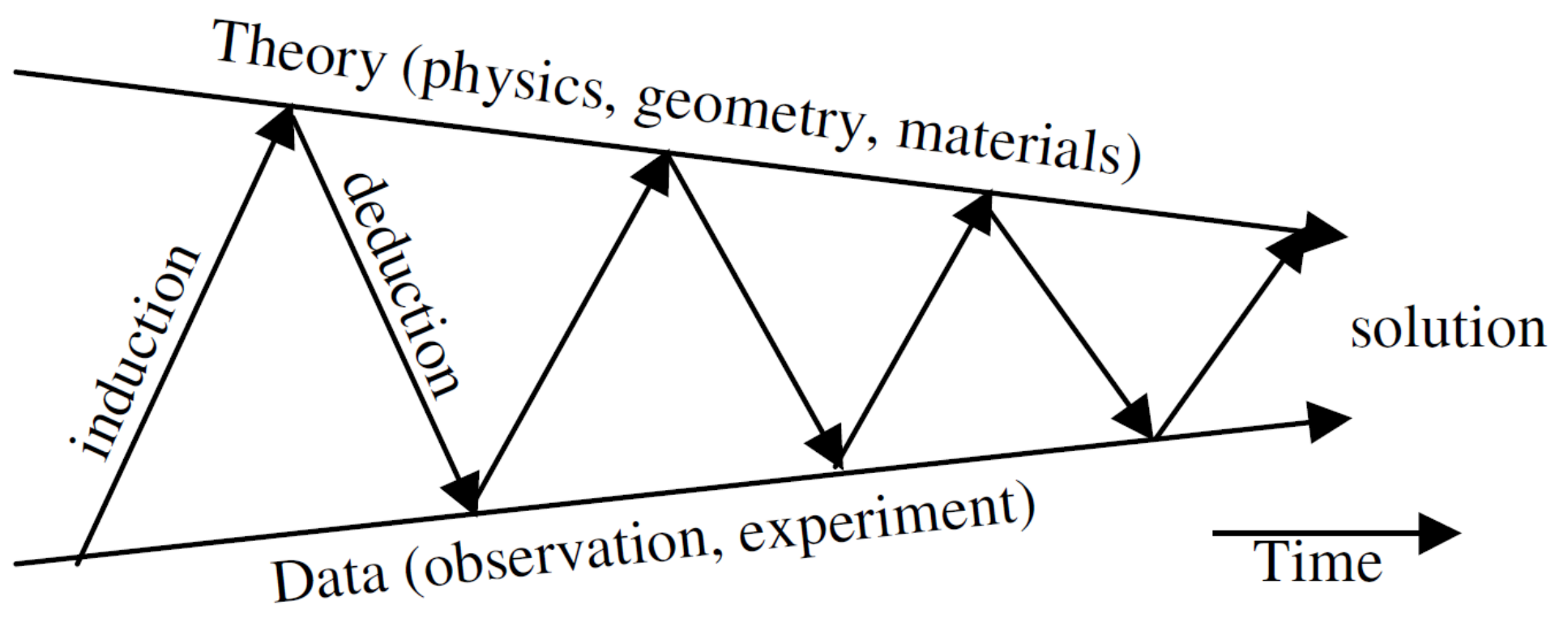
Henry Ford

$\rightsquigarrow$ Data science is a multidisciplinary, interdisciplinary and transdisciplinary **team sport**!

‘If you do not seek out helpers and allies, then you will be isolated and weak.’

Sun Tzu

$\rightsquigarrow$ **Learning through iteration between theory and data using induction** (*i.e.* 'data first' for 'idea generation' using data science) **and deduction** (*i.e.* 'idea first' for 'idea evaluation or testing' using traditional statistics)**:**

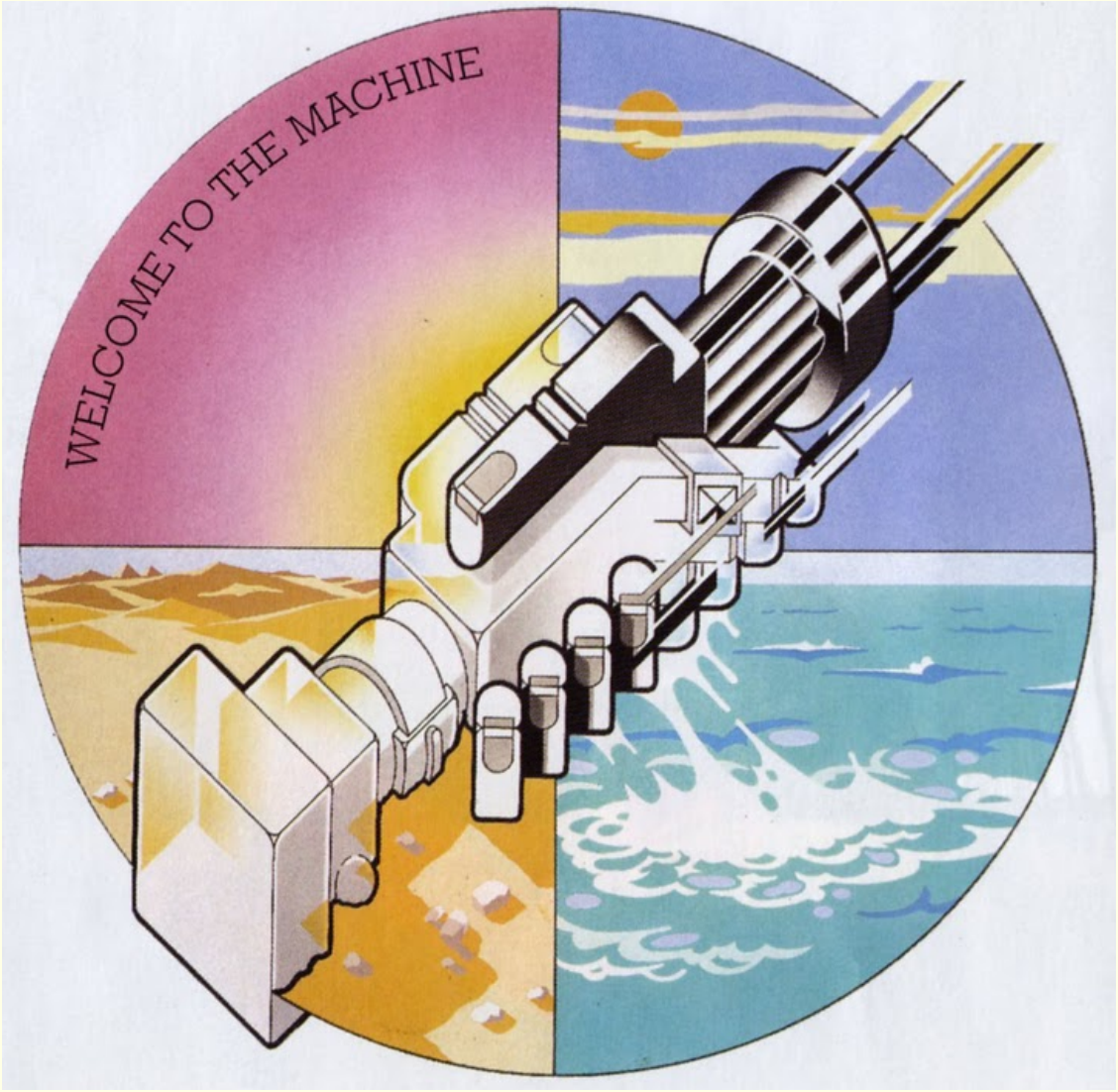


Source: Davis, T. P. (2006). Science, engineering, and statistics. *Applied Stochastic Models in Business and Industry*, 22, 401–430.

‘Neither exploratory nor confirmatory is sufficient alone. To try to replace either by the other is madness. We need them both.’

John W. Tukey, 1980

‘Welcome to the Machine’ (Pink Floyd, 1975)



Demystification of 'machine intelligence and learning'

◇ John McCarthy, one of the founders of 'Artificial Intelligence' (AI) (now sometimes referred to as 'machine intelligence') research, defined in 1956 the field of AI as

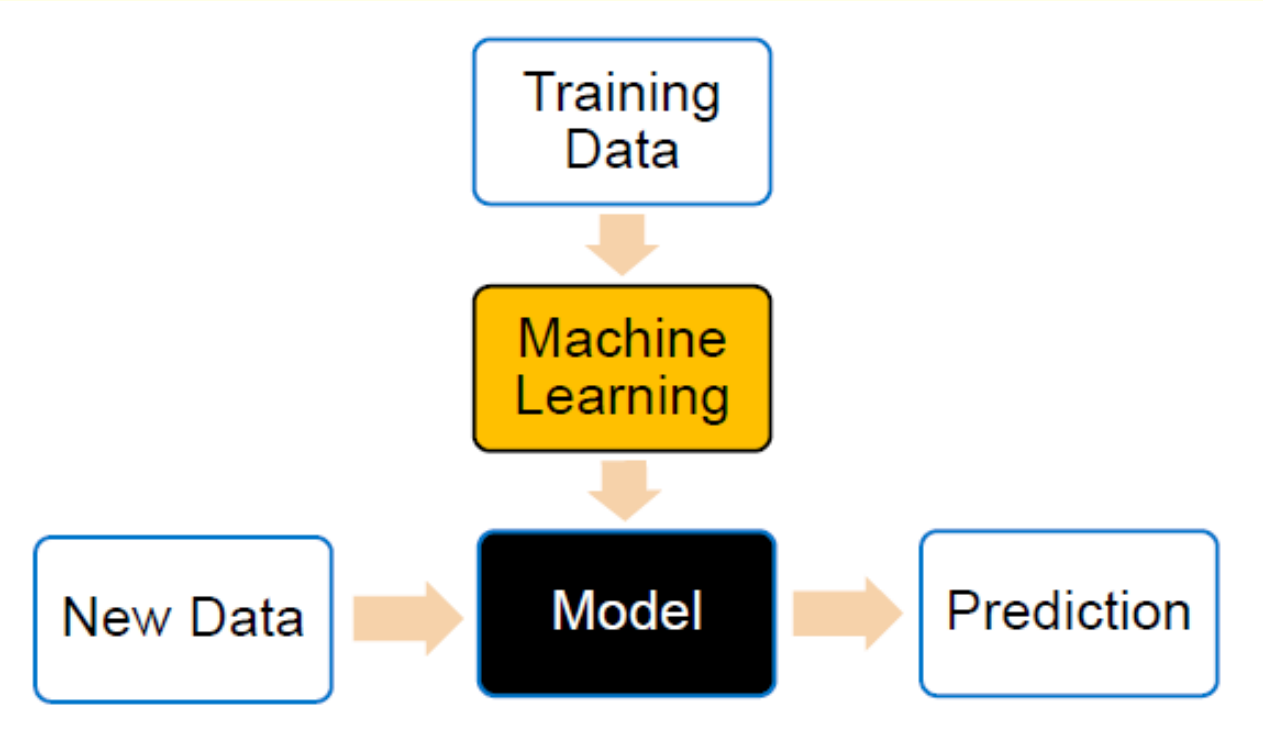
> 'getting a computer to do things which, when done by people, are said to involve intelligence',

e.g. visual perception, speech recognition, language translation, visual translation and playing games (with concrete rules).

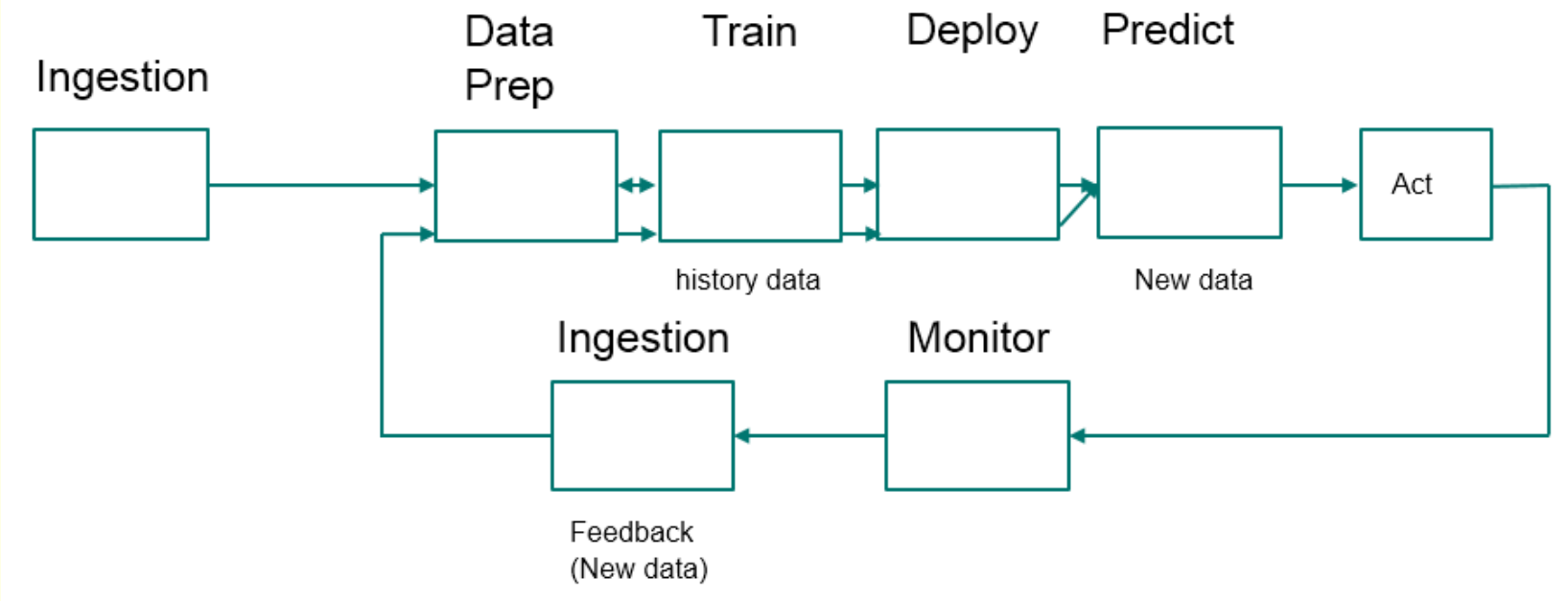
$\rightsquigarrow$ AI is about (smart) machines capable of performing tasks normally performed by humans ($\rightsquigarrow$ 'learning machines'), *i.e.* 'making machines smart'.

$\diamond$ In 1959, Arthur Samuel defined 'Machine Learning' (ML) as one part of a larger AI framework 'that gives computers the ability to learn'.

$\rightsquigarrow$ ML explores the study and construction of algorithms that can learn from and make predictions on (yet-to-be-seen) data, *i.e.* 'prediction making' through the use of computers, and help make decisions.



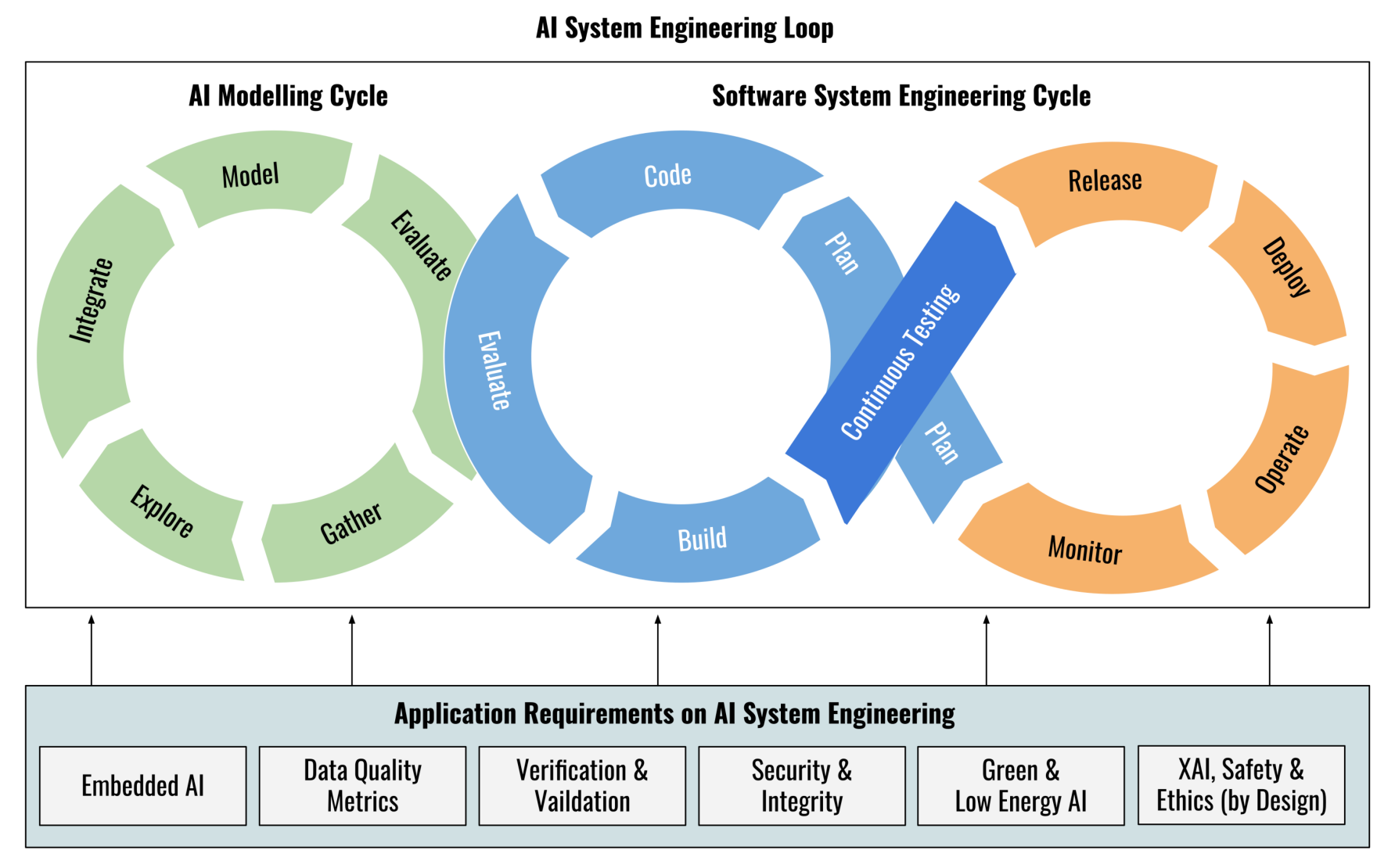
An example of a machine learning workflow



$\rightsquigarrow$ **Monitoring** and using the resulting **feedback** are at the core of machine learning.

$\rightsquigarrow$ Implementation requires the automation of the monitoring step and the feedback ingestion step. Assuming this is done, we have a 'learning machine'.

Source: Jean-Francois Puget, Chief Architect, IBM Analytics Solutions, 'Machine learning algorithm $\neq$ learning machine', April 27, 2016.



Source: Fischer, L., Ehrlinger, L., Geist, V., Ramler, R., Sobiezky, F., Zellinger, W., Brunner, D., Kumar, M. & Moser, B. (2021). AI system engineering — Key challenges and lessons learned. *Machine Learning and Knowledge Extraction*, 3, 56–83.

‘AI algorithms are not natively ‘intelligent’. They learn inductively by analyzing data.

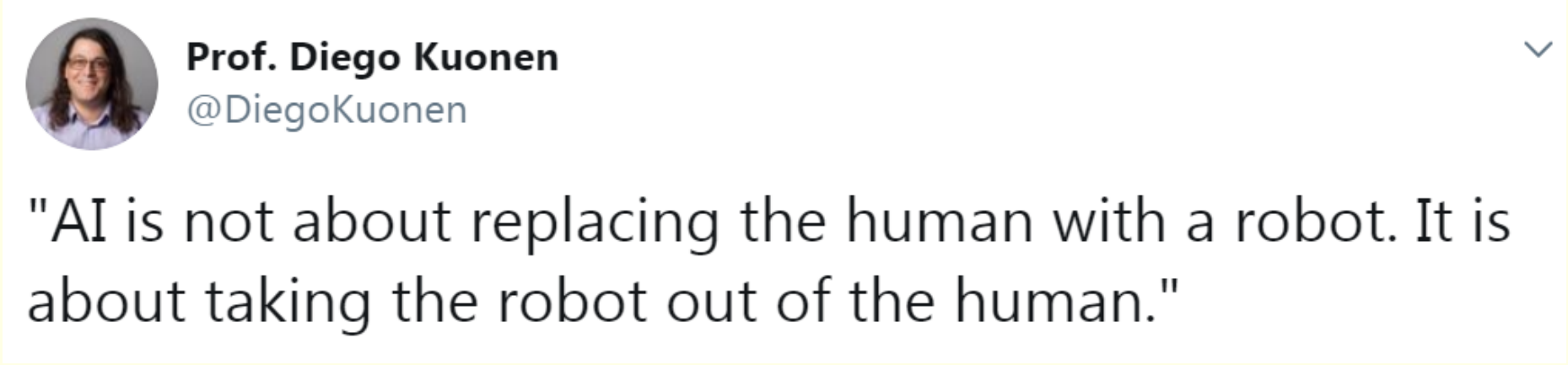
Sam Ransbotham, David Kiron, Philipp Gerbert and Martin Reeves, 2017

Source: Ransbotham, S., Kiron, D., Gerbert, P. & Reeves M. (2017). *Reshaping Business With Artificial Intelligence*. MIT Sloan Management Review & The Boston Consulting Group (goo.gl/wnGqr3).

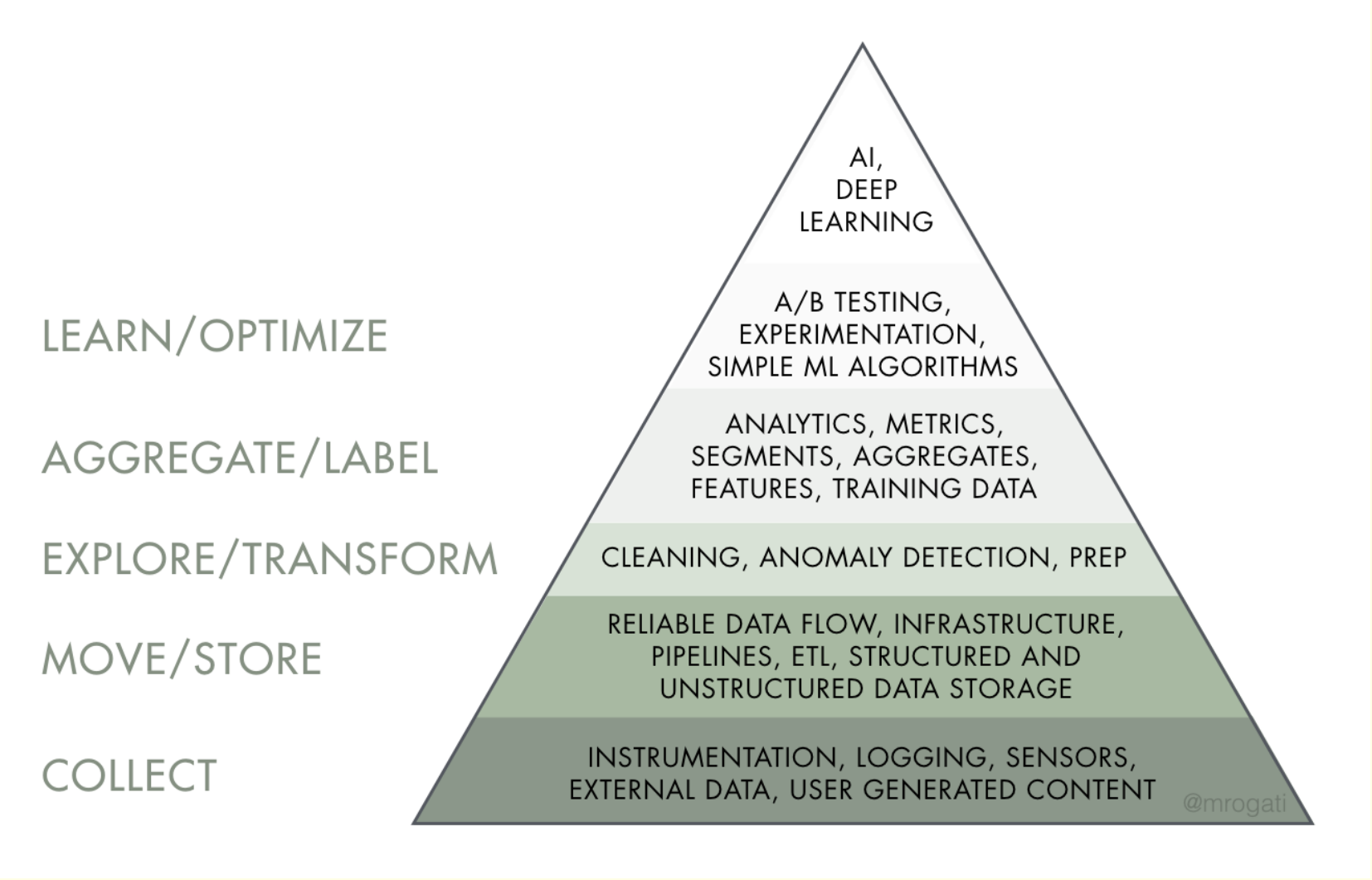
• However, without humans as a guide, current AI is no more capable than a computer without software!

• AI without trustworthy data is like a swimming pool without trustworthy water!

• There is nothing artificial about AI: it is inspired by humans, it is created by humans and impacts humans!



• The largest and **most basic 'need'** is a 'strong' data collection (Monica Rogati, 2017; goo.gl/F7hKH7):



⇝ A 'strong' **'data pedigree' is key**!

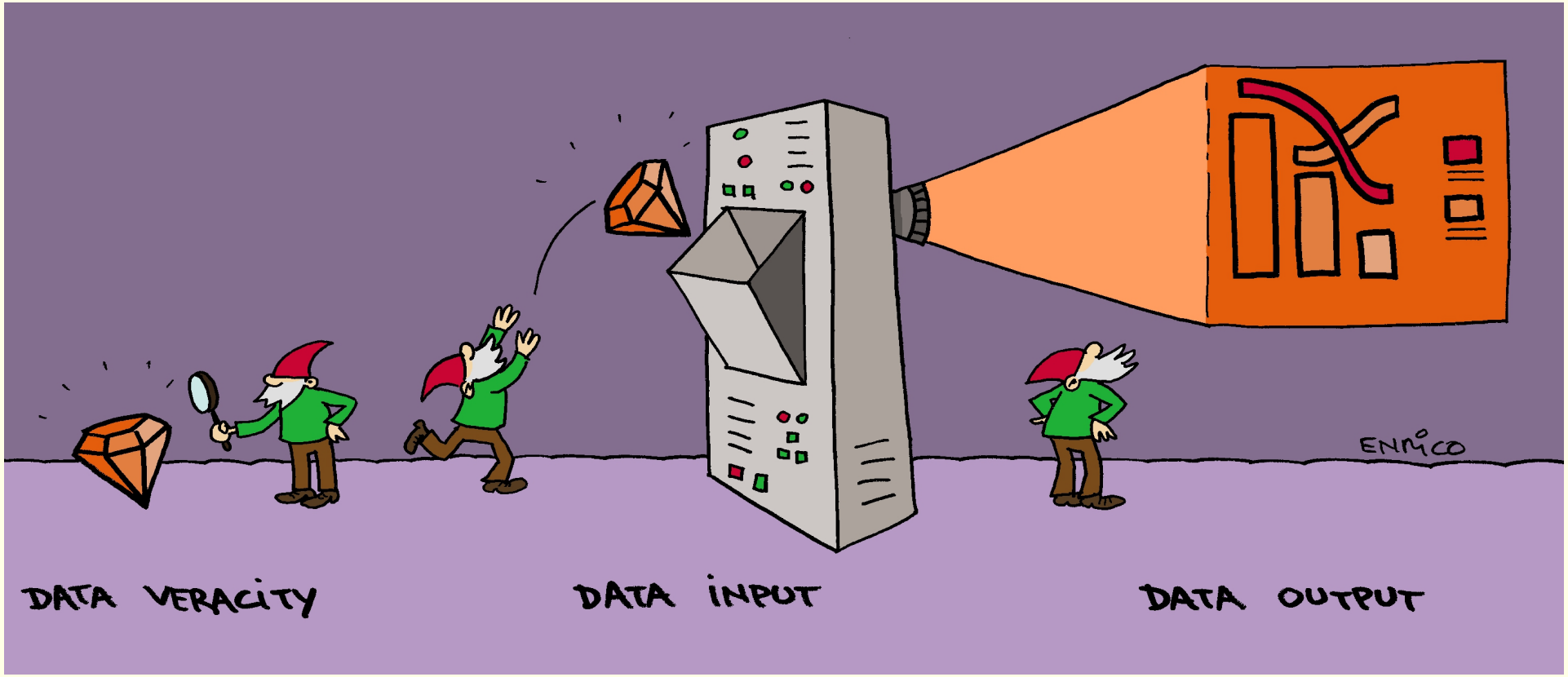
‘Data themselves are a central raw material of the knowledge society. However, this means that the data must be of high quality, accessible and trustworthy.’

Swiss Federal Council, September 5, 2018

Source: ‘Digital Switzerland Strategy’, adopted by the Federal Council on September 5, 2018 (goo.gl/b7K8aE).

Challenges, opportunities, and principles for success

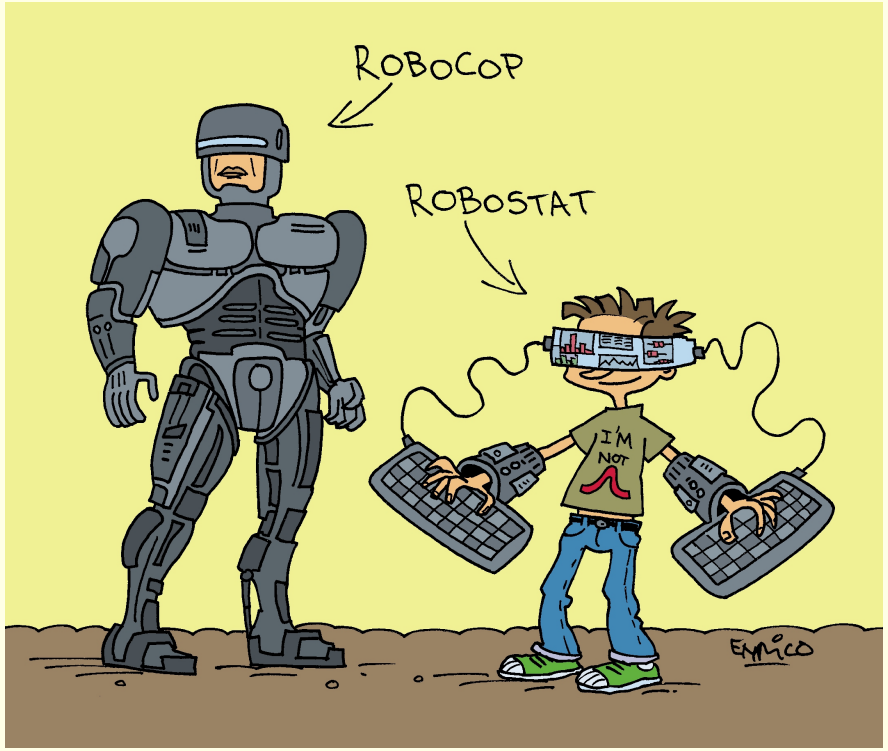
• The ‘data veracity’, *i.e.* the ‘trustworthiness’ of data, and the related data quality are more important than ever!



- Data science (including AI) is an aid to thinking and not a replacement for it!

- Data and data science (including AI) should be envisaged to complement and augment humans!

$\rightsquigarrow$ **Humans need to augment their strengths to become more 'powerful'**: by automating any routinisable work and by focusing on their core competences.



‘By ‘augmenting human intellect’ we mean increasing the capability of a man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems.’

Douglas C. Engelbart, 1962

Source: Engelbart, D. C. (1962). ‘Augmenting human intellect: a conceptual framework’ (1962paper.org).

My key principles for success

- **Do not neglect** the following four principles that ensure successful outcomes:
  - use of sequential approaches to problem solving and improvement, as studies are rarely completed with a single data set but typically require the sequential analysis of several data sets over time ($\rightsquigarrow$ 'continuous improvement');
  - having a strategy for the project and for the conduct of the data analysis; including thought about the 'business' objectives ($\rightsquigarrow$ 'strategic thinking');
  - carefully considering data quality and assessing the 'data pedigree' before, during and after the data analysis; and
  - applying sound subject matter knowledge ('domain knowledge' or 'business knowledge', *i.e.* knowing the 'business' context, process and problem to which data science will be applied), which should be used to help define the problem, to assess the 'data pedigree', to guide data analysis and to interpret the results.

‘Experiments may be conducted sequentially so that each set may be designed using the knowledge gained from the previous sets.’

George E. P. Box and K. B. Wilson, 1951

$\rightsquigarrow$ Scientific investigation is a sequential learning process!

‘All improvement takes place project by project and in no other way.’

Joseph M. Juran, 1989

First things first

Why do so many companies get this wrong?

- ☑ Business → Data → Technology
- ☒ Data → Technology → Business
- ☒ Technology → Business → Data
- ☒ Business → Technology → Data
- ☒ Data → Business → Technology
- ☒ Technology → Data → Business

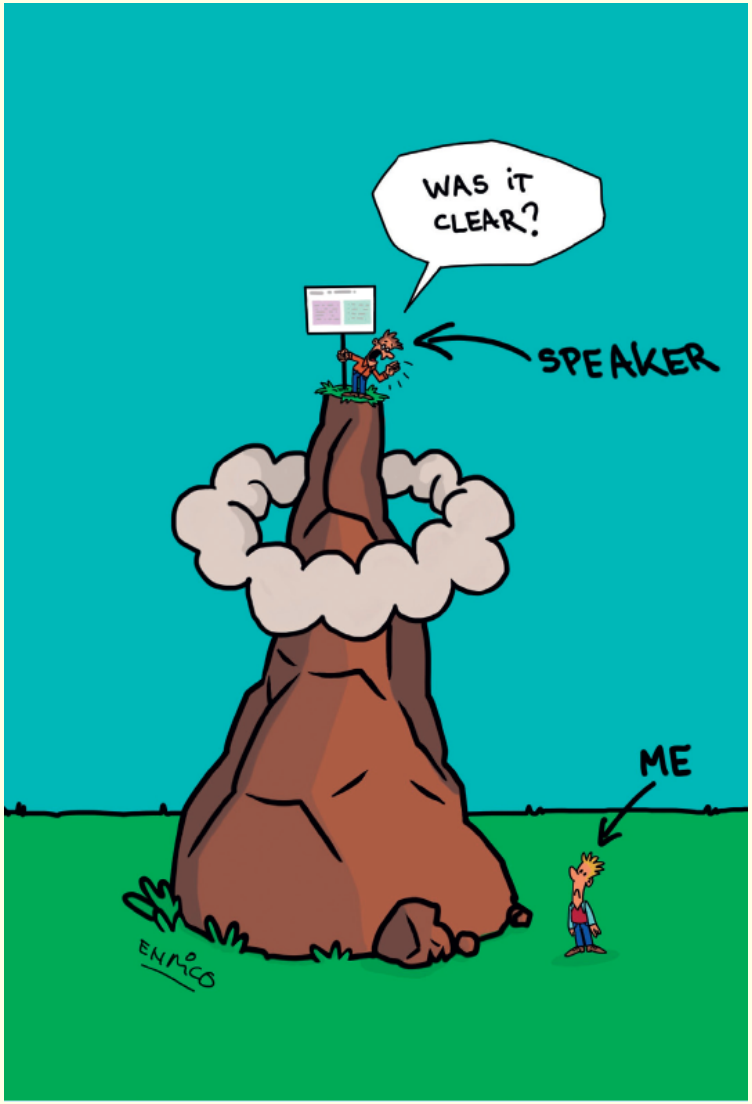
Successful Projects Proceed 'From Left to Right'

Source: Adapted from John Roberts' *The Modern Firm*

The idea is that a company (or any department or team) should start at Step 1, by sorting out what it wants to achieve — its strategy — and its business objectives. This makes sense: You have to know where you want to go before you start moving. Next, in Step 2, participants should sketch out the organizational capabilities they need to execute that strategy, including people, structure, culture, and management routines. For Step 3, leaders should define the processes and data they (and the project overall) will need to do the work. Finally, during Step 4, they should apply the technology necessary to increase scale and decrease unit cost.

Source: Redman, T. C. (2022). *The Trust Problem That Slows Digital Transformation*. MIT Sloan Management Review, July 2022 (`bit.ly/3Z8h2xA`).

A common language is key to success!



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Council
Federal Council
Il Consiglio federale
Il Cussegl federal

Press release

Date 25.08.2021

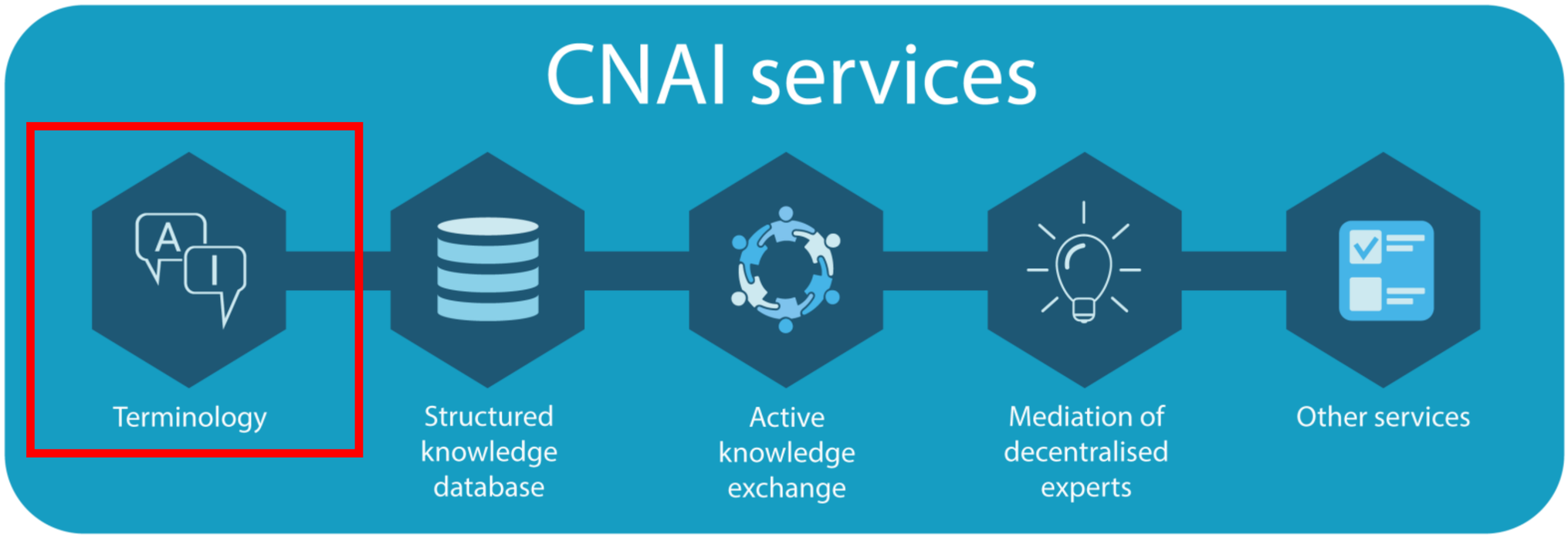
The Confederation develops a “Competence Network for Artificial Intelligence”

Artificial Intelligence (AI) is among the pioneering technologies of the digital future and has also a lot of potential to be used in the federal administration. At its meeting on 25.08.2021, the Federal Council decided to develop a “Competence Network for Artificial Intelligence” (CNW AI). The corresponding unit for this will be attached to the Federal Statistical Office (FSO).

CNAI

Competence Network for Artificial Intelligence
Kompetenznetzwerk für künstliche Intelligenz
Réseau de compétences en intelligence artificielle
Rete di competenze per l’intelligenza artificiale

CNAI.swiss



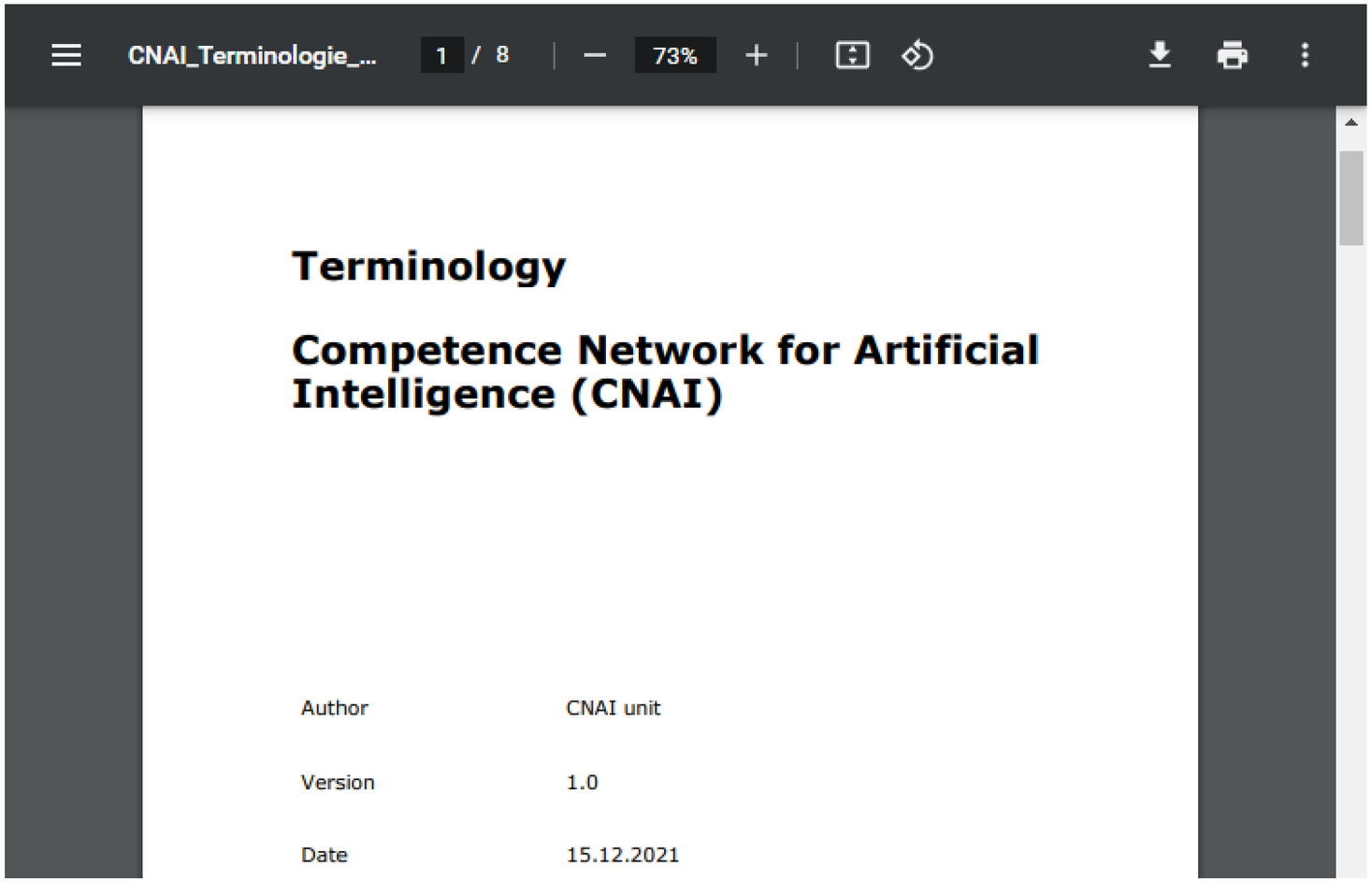
Competence Network for Artificial Intelligence
Kompetenznetzwerk für künstliche Intelligenz
Réseau de compétences en intelligence artificielle
Rete di competenze per l'intelligenza artificiale

CNAI.swiss

Terminology

What is artificial intelligence and what are new technologies? And how is AI different to data science or machine learning? Both these and other relevant terms key to a central understanding of AI are defined in the *Terminology*.

The introduction of a standardised terminology is an important foundation for the functions of the CNAI. A common language and a corresponding common understanding of terms at the level of the Federal Administration facilitates the active exchange of experience and knowledge both within and beyond the CNAI network. Furthermore, a common understanding simplifies the communication of ideas, plans and services in this area.



‘Data-driven decision support in the federal administration: increased effectiveness and efficiency in public policy through data science.’

Swiss Federal Council, December 2, 2022

Source: ambition of the ‘*Federal Data Science Strategy*’, adopted by the Federal Council on December 2, 2022.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

The Federal Council

Bern, 2 December 2022

Federal data science strategy

The federal administration's common bases, competences and objectives for the application of data science

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Press release

Date 02.12.2022

Confederation sets targets for the use of data science

The federal government wishes to use data science in a more targeted way in the future to support the government and the administration in their work. On 2 December 2022, the Federal Council adopted the federal data science strategy and issued various mandates.

www.bfs.admin.ch/news/en/2022-0738

Common language and understanding as a basis

3.1 Importance of the common language

Creating a common language can be seen as the heart of the federal data science strategy. A shared understanding of terms and a resulting common language enables the active exchange of experience and knowledge and simplifies communication regarding ideas, projects and services. This is because although considerable experience in data science is already available within the federal administration to some extent, the degree of maturity in the application of data science varies.

Especially in light of the ambitious data science development envisaged for the federal administration, it is essential to prevent misunderstandings right from the start by using a common language and clearly defined core terms. The following shows what data science and its related concepts mean for the federal administration.

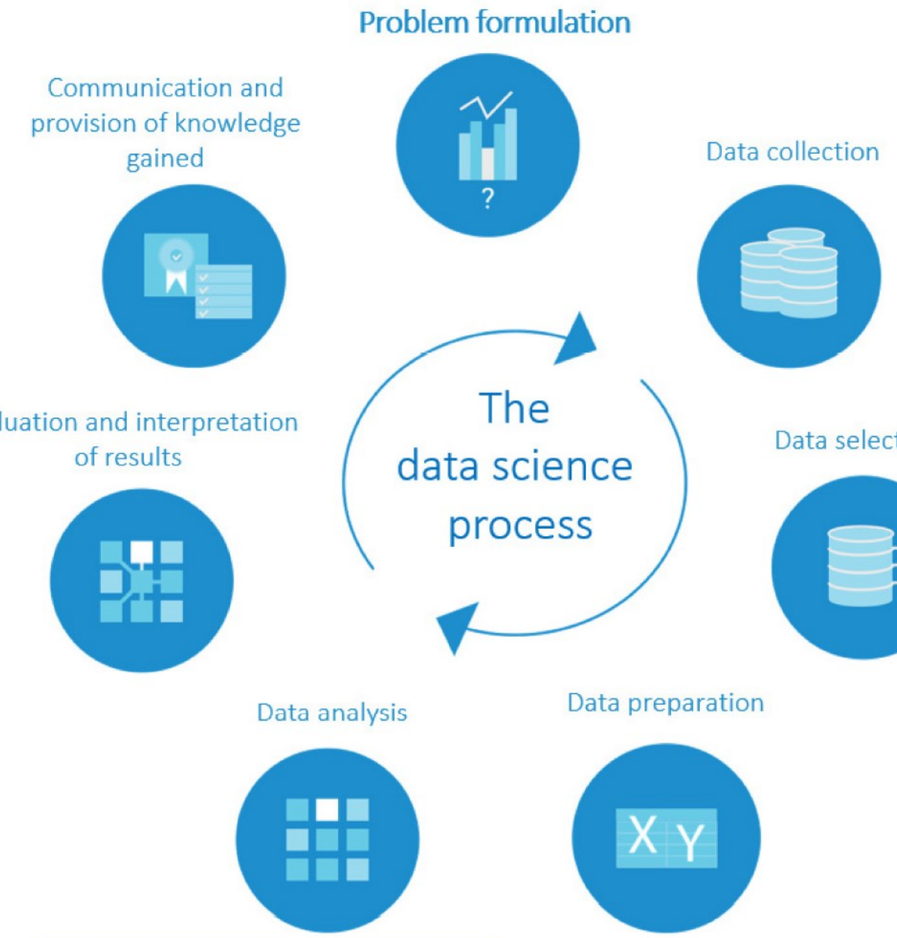
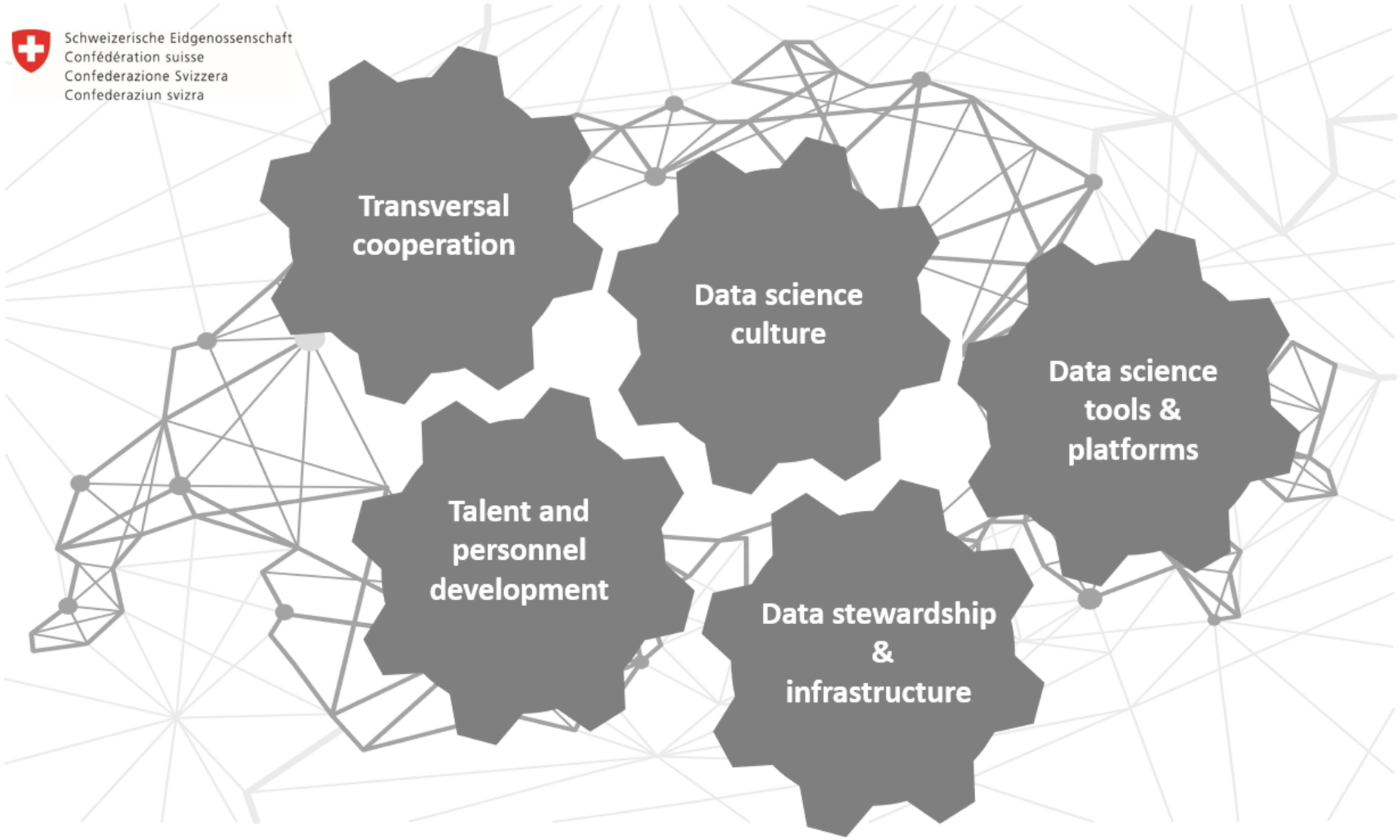


Figure 2: Interdisciplinary data science problem-solving process. *(Source: Own diagram based on the https://www.bfs.admin.ch/bfs/en/home/dscc/dscc.html).*

Data science is the interdisciplinary science of learning from data with the aim of gaining insights from the data in order to facilitate data-based decision-making. It is a problem-solving process based on continuous improvement that aims to solve complex, unstructured and data-rich problems through the application of data science methods (e.g., methods from advanced statistics, machine learning and the field of artificial intelligence), techniques and practices. Data science covers the entire process of problem

Target image: transversal “data science ecosystem”



Vision statement: "Human-centric and trustworthy data science supports the public good and public policy".

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Council
Federal Council
Il Consiglio federale
Il Cussegl federal

Press release

Date 02.12.2022

Various mandates

At the same time as adopting the strategy, the Federal Council issued various mandates. In this way, the Federal Statistical Office (FSO) is working together with other departments and the ETH Domain to create a report describing where data science can be used throughout the policy-making process. In addition, a code of practice for human-centric and trustworthy data science, and a concept on the application of data science to safeguard privacy are to be drawn up. Finally, the Federal Chancellery (Digital Transformation and ICT Governance Sector (DTI)) and the FDHA/FSO, together with internal federal IT service providers, are to create a concept on how a collaborative data science platform can be established in the federal administration.

www.bfs.admin.ch/news/en/2022-0738

⇝ Do not let culture eat strategy — have them feed each other!

⇝ Culture change is key in any transformation!

‘The only person who likes change is a wet baby.’

Mark Twain

‘The transformation can only be accomplished by man, not by hardware (computers, gadgets, automation, new machinery). A company can not buy its way into quality.’

W. Edwards Deming, 1982

‘It is getting better. . . A little better all the time.’

The Beatles, 1967

Have you been Statooed & GSEMed?

Prof. Dr. ès sc. Diego Kuonen, CStat PStat, PhD EPFL

Statoo Consulting
Morgenstrasse 129
3018 Berne
Switzerland

GSEM, University of Geneva
Bd du Pont-d'Arve 40
1211 Geneva 4

email kuonen@statoo.com
Diego.Kuonen@unige.ch

web www.statoo.info
gsem.unige.ch/rcs/kuonen


@DiegoKuonen

Presentation code: 'Geoinformationstag2023.Okt5.2023'. Typesetting: LaTeX, version 2ϵ. PDF producer: pdfTeX. Compilation date: 04.10.2023.