



# Data Science & Machine Learning Workshop

Manuel GONZALEZ-BERGES (CERN)  
Gianluca VALENTINO (University of Malta)

7<sup>th</sup> October 2023

# *Discussion Session*

# Tools & Infrastructure key components

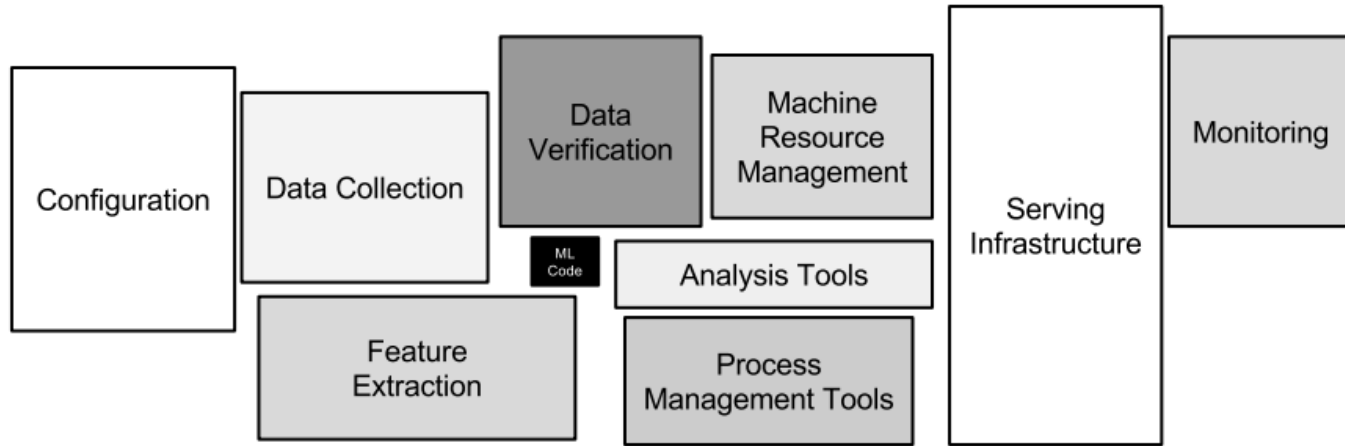


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

Hidden Technical Debt in Machine Learning Systems (Google paper 2015)

# Exercise



ChatGPT, like its predecessor GPT-3, is primarily based on a deep learning technique called the Transformer architecture. The Transformer architecture is a neural network architecture that was introduced in the paper "Attention is All You Need" by Vaswani et al. in 2017. It has since become a foundational building block for many natural language processing (NLP) tasks, including chatbot and language generation tasks.



Here are some of the key machine learning techniques and components used in ChatGPT:

1. **Transformer Architecture:** The core of ChatGPT is the Transformer architecture, which relies

M.

M.

Could you simplify the previous answer to 2 main types of machine learning, not specific techniques?



Certainly! The machine learning techniques used in ChatGPT can be simplified into two main types:



1. **Supervised Learning:** ChatGPT is initially pretrained using a massive dataset of text from the internet. During this phase, it learns patterns and language understanding through a form of supervised learning where it predicts the next word in a sentence. Later, it is fine-tuned for specific tasks, such as chatbot interactions, using supervised learning with labeled data.
2. **Reinforcement Learning:** In some cases, ChatGPT is further refined using reinforcement learning techniques. This involves providing the model with rewards or feedback based on the quality of its responses during interactions with users. Reinforcement learning helps improve the model's conversational abilities and generate more desirable responses.

# Discussion Topics

- Use of Assistants
  - Programming (e.g. GitHub Copilot, ChatGPT, Amazon CodeWhisperer)
  - Design of systems (e.g. electronics circuit, particle accelerator)
- Your proposals?