# LLM Apps

## Evaluation of an LLM App: Misaligned Behavior

# Misaligned behavior

- Lack of alignment between the behavior of the LLM App and the values of the company.

- Foundational LLM models have been created from content produced by humans. If these original contents were misaligned, the LLM model will generate misaligned content.

# Misaligned behavior: where it can occur

At any stage of the app lifecycle:

- Data Acquisition
  - personal data?
  - insecure data?
  - permission issues?
- Data Preparation
  - should some data be omitted?
  - should some data be anonymized?
  - is data encryption necessary?
- Data Modelling
  - is the model misaligned?
  - proper use of randomization?
  - does the model represent the data?
- Data Interpretation
  - are they misaligned?
  - are they consistent?
  - what implications do they have?