Master Degree in Computer Science

Natural Language Processing

UNIVERSITÀ DEGLI STUDI DI MILANO

Overview of real LLMs

LA STATALE

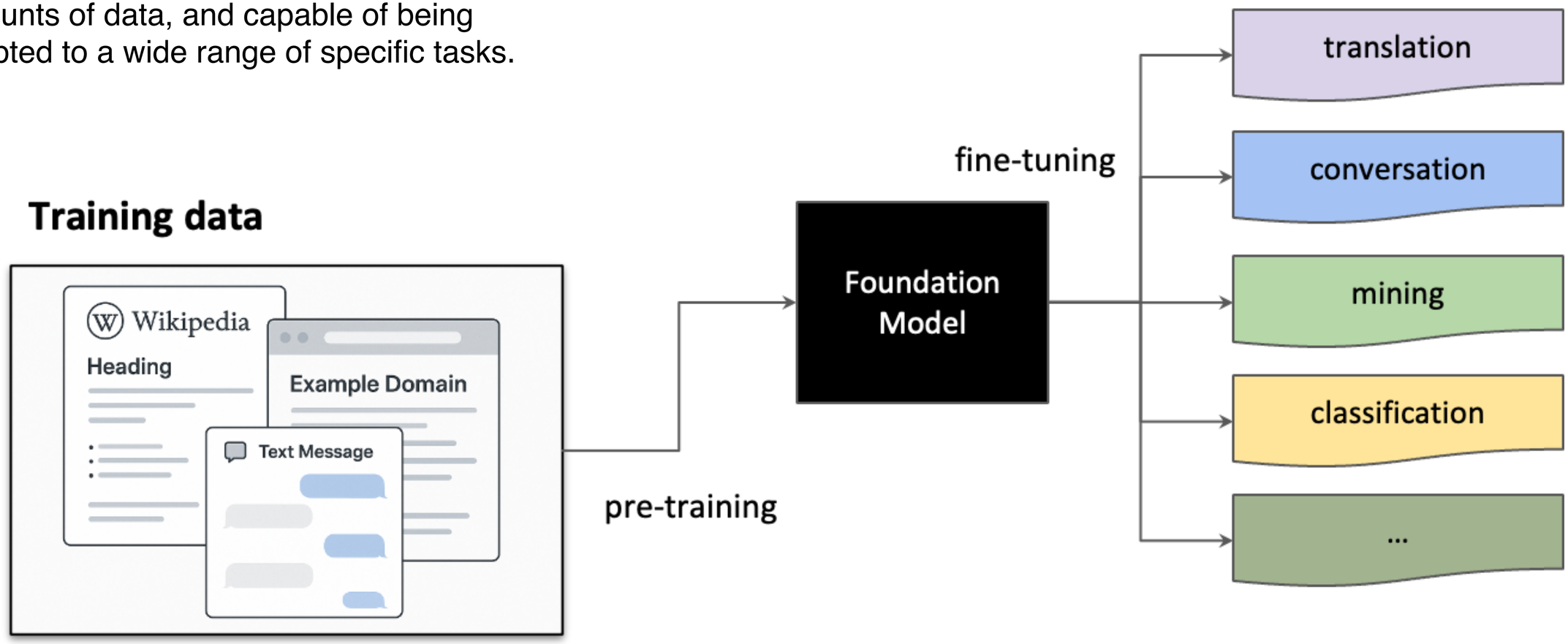
Prof. Alfio Ferrara

Department of Computer Science, Università degli Studi di Milano
Room 7012 via Celoria 18, 20133 Milano, Italia alfio.ferrara@unimi.it

sed noli modo

Foundation Models

Foundation models are large artificial intelligence models, trained on enormous amounts of data, and capable of being adapted to a wide range of specific tasks.



Examples of generative models

Proprietary

ChatGPT

Gemini

Claude

Open Weights

LLaMA

PHI-4

MISTRAL AI_

Apps based on foundation models

Microsoft Copilot

grammarly

NotebookLM

perplexity

Canva AI TOOLS

Main differences (with notable exceptions)

Proprietary

- **Pros:**
  - Higher performance
  - Ease of use
- **Cons:**
  - Subscription costs
  - Not customizable or minimally customizable
  - Dependence on external services

- **Open**
- **Pros:**
  - Adaptability to specific needs
  - Transparency in operation
- **Cons:**
  - Hardware costs
  - Requirement of technical skills
  - Lower performance

UNIVERSITÀ
DEGLI STUDI
DI MILANO

LA STATALE

What we learned about LLMs and some topics for the next classes

No true comprehension

The model does not truly "understand" what it writes. It generates every word based on statistical probabilities learned from enormous amounts of text, without any awareness, intention, or semantic understanding. Even when the result seems to make sense, it is driven by patterns, not by comprehension.

Extraordinary versatility

Even without real comprehension, the generation process is extremely adaptable: the same mechanism can produce stories, answers, code, poems, or academic texts. By simply modifying the prompt or adjusting parameters like "temperature," the model can imitate a wide variety of styles and linguistic purposes.

Limited but possible explainability

We can partially explain the process by observing the probability distribution over candidate words. However, the deep "why" of a choice is incorporated into the millions of parameters of the network, which makes it difficult to translate decisions into human terms. Explainability tools exist, but they remain approximate and constantly evolving.

Open topics

Hallucinations and errors
Hallucinations in generative models are plausible but false responses (truth vs. verisimilitude). Errors, on the other hand, arise from limits in the data or prompts: absence of semantic control and fact-checking.

BIAS, stereotypes, and censoring behaviors
Biases and stereotypes reflect the prejudices present in the training data: the model absorbs them and can reproduce them, amplifying them. Censoring behaviors, however, arise from the filters imposed to limit these risks, but can introduce new forms of partiality, preventing legitimate or nuanced responses.

Opaque behavior
The opaque behavior of generative models derives from their internal complexity: billions of parameters interact in ways that are difficult to interpret. This makes it arduous to understand how and why a certain response is produced, limiting transparency and the possibility of human control.