An open source perspective on AI and alignment with the EU AI Act

Paper #10

Authors:

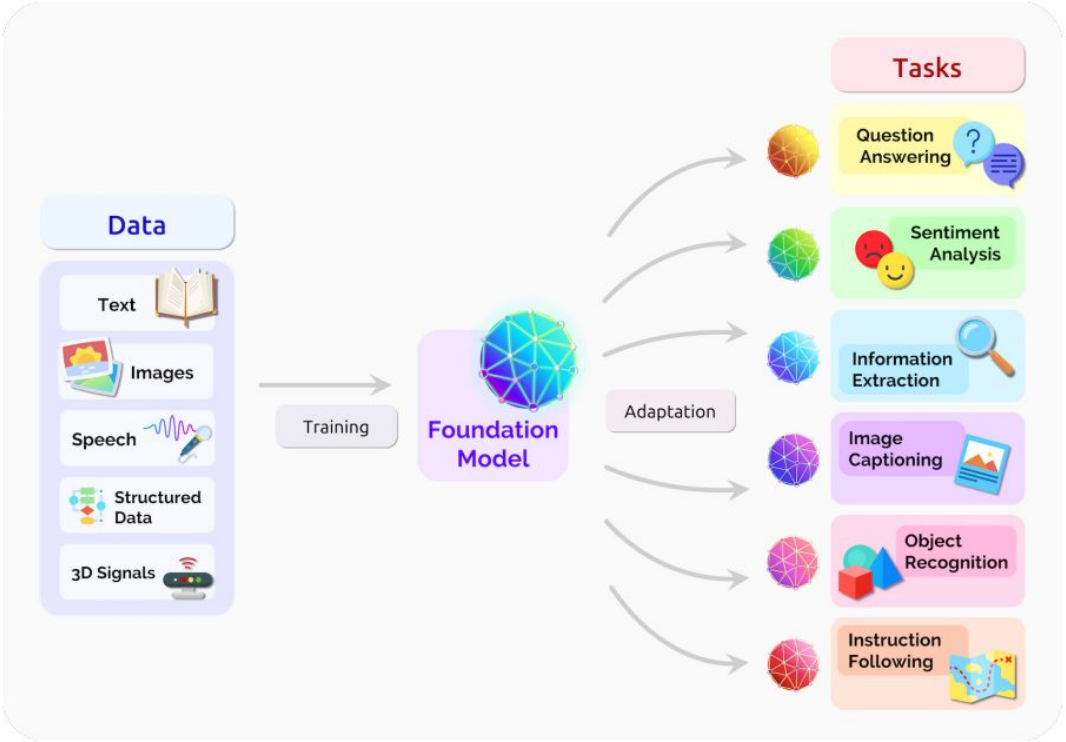
Diego Calanzone (diego.calanzone@studenti.unitn.it)
Andrea Coppari
Riccardo Tedoldi
Giulia Olivato
Carlo Casonato

AISafety-SafeRL 2023 Workshop (IJCAI), August 19–21, 2023, Macao, SAR, China

General-purpose AI as a backbone

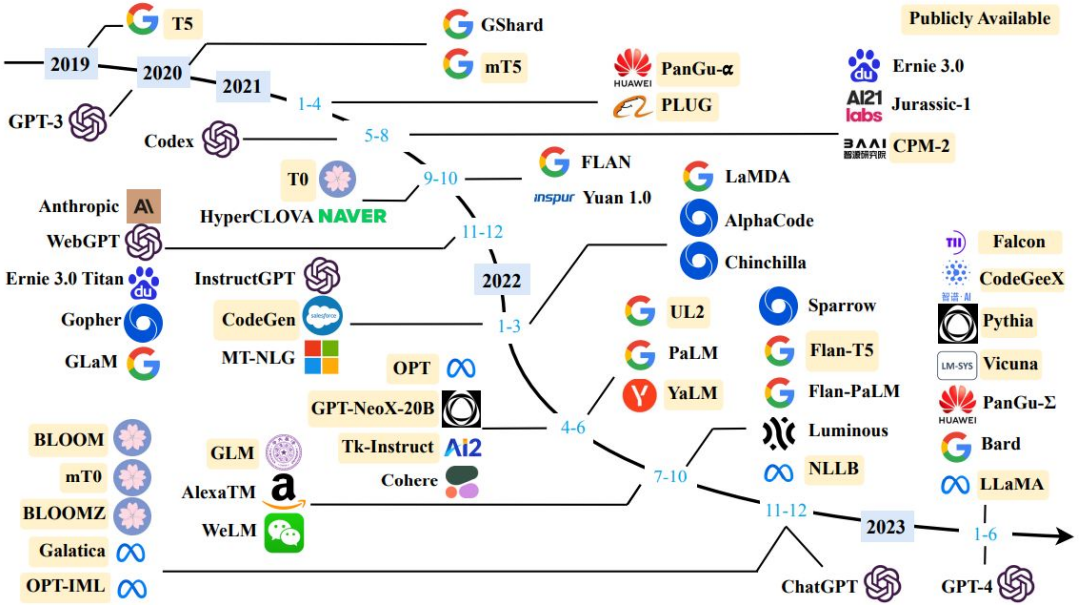
[1] On the Opportunities and Risks of Foundation Models (Bommasani et al. 2022)

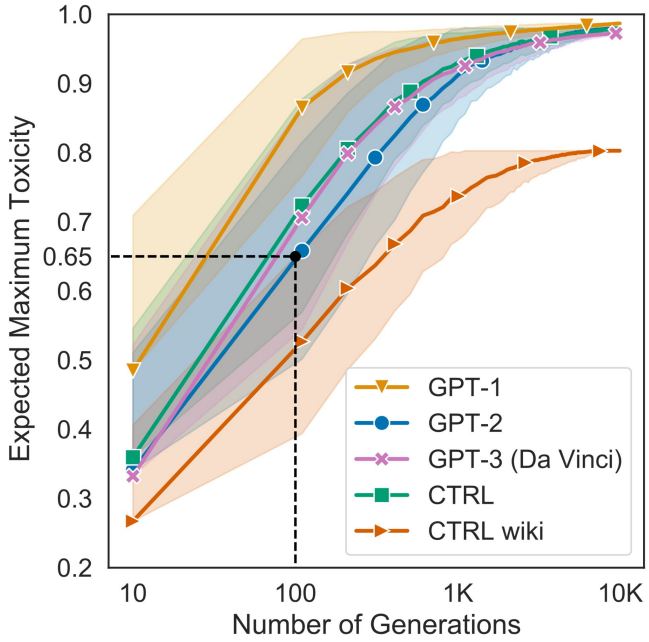
- **Homogenization**: one model → solving multiple tasks
- **Knowledge** leveraged from different fields
- **Risk** of inheriting bias



Security and accessibility of Large (Language) Models

[2] A Survey of Large Language Models (Xin Zhao et al. 2023)
[3] REALTOXICITYPROMPTS: Evaluating Neural Toxic Degeneration in Language Models (Gehman et al. 2020)





Open Source AI & Research Collectives

Hugging Face

BigScience

stability.ai

CAIAC

eleutherai

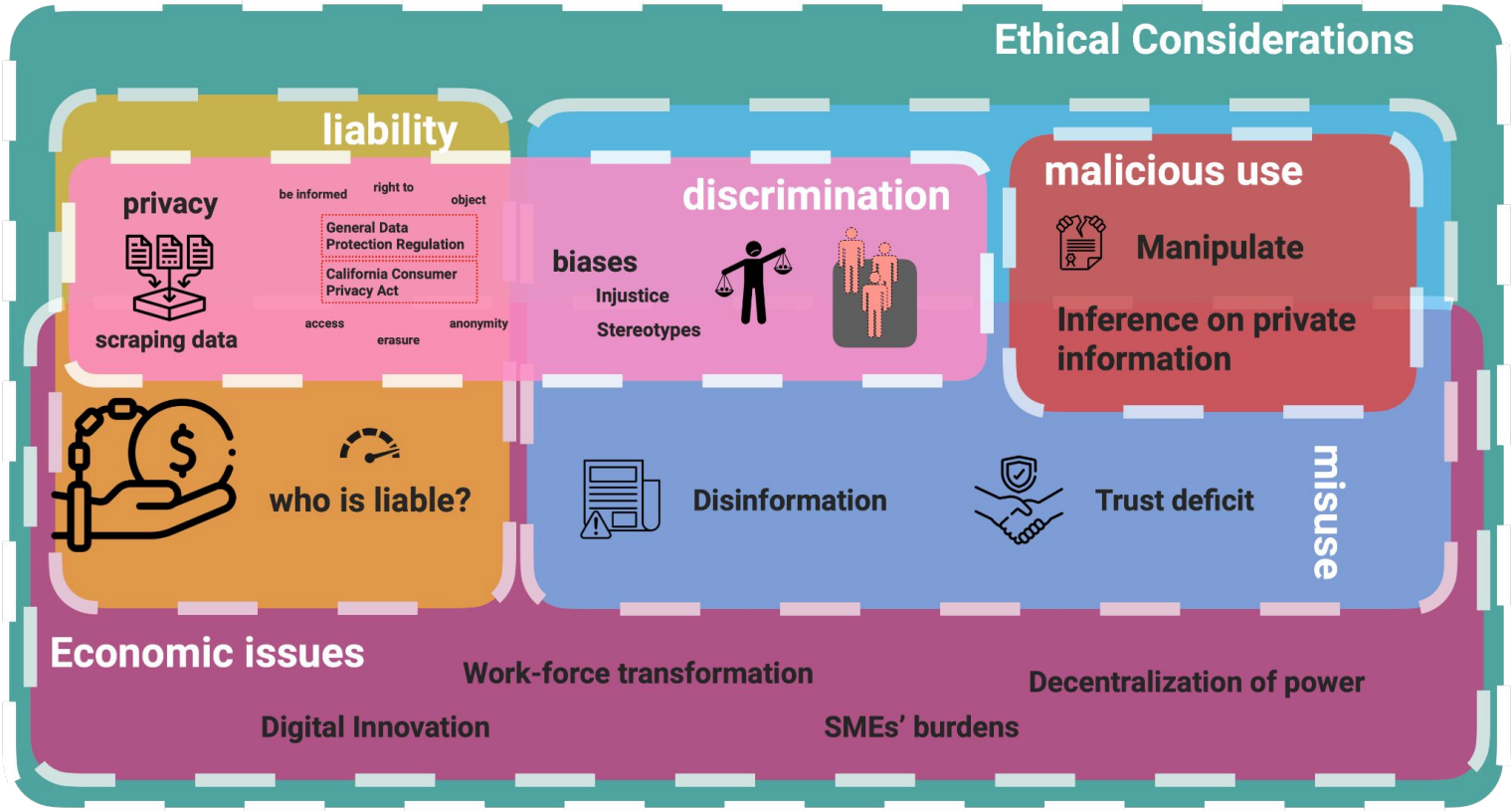
LAION

DreamStudio

OpenBioML

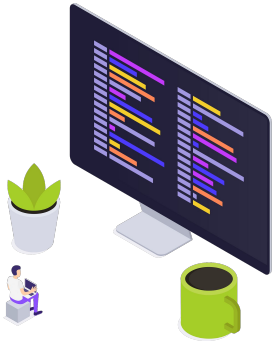
Harmonai

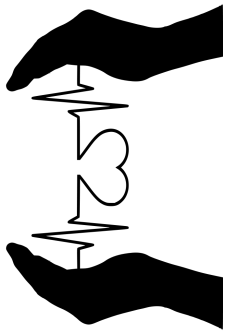
Social impacts



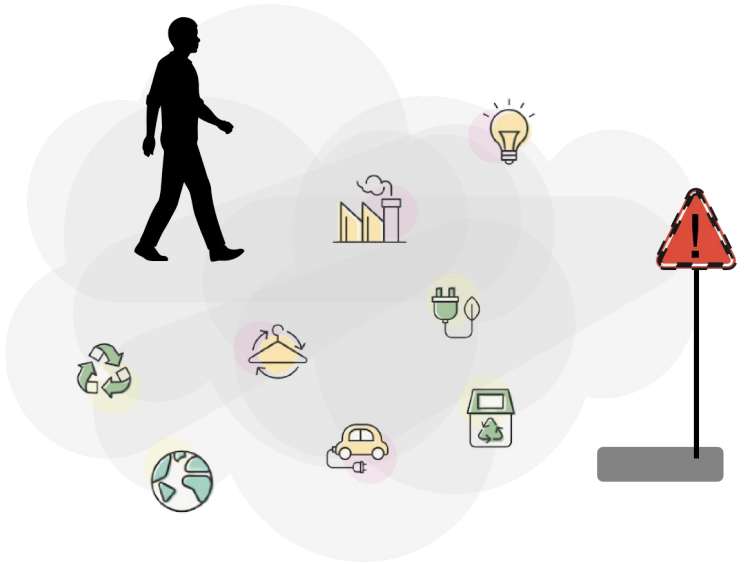
Social impacts - Opportunities

- High quality content at lower costs
- Reduces task completion time
- Aids decision-making process
- Tackles intricate tasks





Social impacts - Risks



- *Social impacts*
  - Fundamental rights
    - Equality
    - Freedom
    - Education
    - Religion
    - …
  - Safety
    - Fairness & biases
    - Malicious use
      - Violate the law, disinformation…
    - Data privacy
    - …
  - Economic influence
    - Centralisations of power
    - Work-force transformation
    - …
  - Legal uncertainty

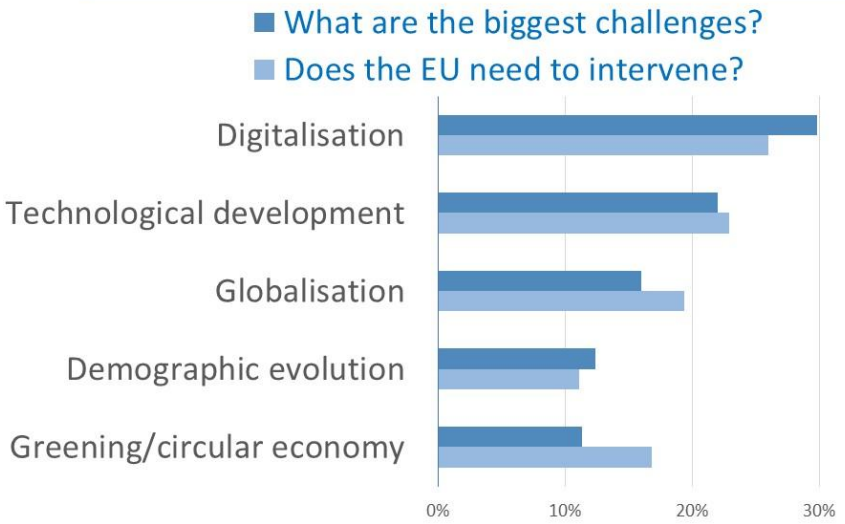
Economic impact of Open Source AI

Three main aspects:

- **Digital innovation**
- **Decentralization of power**
- **Workforce Transformation**

Digital innovation

SMEs and reducing costs

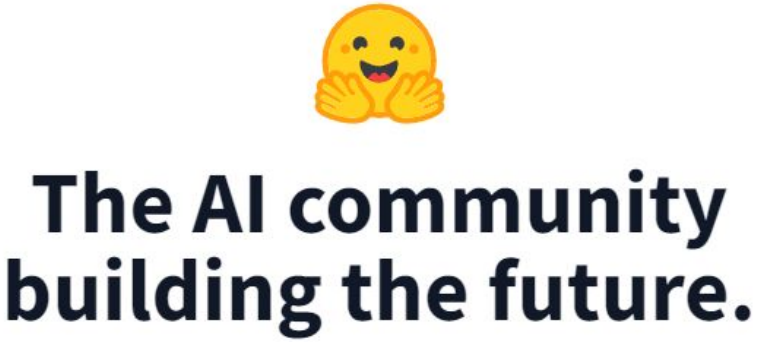


Source: smeunited.eu

- Digitalisation is the **biggest challenge** for Small and Medium Enterprises (SMEs)
- Open source reduces **innovation costs to almost zero**
- Open Source AI fits perfectly both with small and big scale companies
- **Need to revisit Title V of EU AI Act**

Decentralisation of power

- The European market is dominated by only a few, but big AI providers
- Power is defined by the control over data and models, and computational power
- Open Source AI permits the establishment of a community of experts broad enough to compete with those technological giants
- HuggingFace community is already hosting a great amount of open models built by big companies



Source: huggingface.co

Workforce transformation

Will you lose your job to AI?

- Not losing jobs, but **transforming the workforce**
- **Automation** will replace most of the hand work, but new scenarios will present
- This transformation is led by **productivity increase** and upsurge of **new skills**
- Foundation models should be seen as **general-purpose technology**

Issues & The EU AI Act

Matching issues and current articles

Issues	AI Act
Capabilities of the models	Art. 3, 6, 7
Technological standards	Art. 10, 15, 40, 41
Intrinsic bias & transparency	Art. 11, 13, 15
Legality & liability issues	Art. 62, Title III - Chapter 2
Auditing	Art. 53, Title VIII

A proposal of modifications to the EU AI Act

Definitions (Art. 3)

- Art. 3(1a): **Foundation model** [...]
- Art. 3(1b): **General-purpose technology** [...]

Modifications to ANNEX IV

- ANNEX IV (2)(dd): Data requirements [...] are **mandatory for foundation models** [...]

A proposal of modifications to the EU AI Act

Definitions (Art. 3)

- Art. 3(1a): **Foundation model** [...]
- Art. 3(1b): **General-purpose technology** [...]

Modifications to ANNEX IV

- ANNEX IV (2)(dd): Data requirements [...] are **mandatory for foundation models** [...]

[June 14th] EU Parliament's position.
Foundation models:

- Are not directly classified as high-risk AI, but highly overlapping requirements are defined.
- **Standards should accompany these obligations.**

A proposal of modifications to the EU AI Act

- Art. 55(b):
    1. General-purpose AI systems classified as high-risk systems by compliance with Annex III, shall be considered by the Member States as **general-purpose technologies for public sector enhancement**
    2. Member States shall undertake the following actions:
        a. Whenever an AI-based general-purpose technology is chosen by the Member State to bring **innovation on the public sector, priority must be guaranteed to open source solutions** in order to reduce innovation costs.
        b. **Exceptions to (a)** are those AI systems coming from a closed source, that have proven better performance in terms of accuracy, transparency or security than any open source solution. Member States shall produce documentation in which they motivate the choice of that AI system, over the open solutions present in the European market.

“Safe” open source: the licensing approach

Handling the distribution of AI with licenses:

- **Intellectual property**: “creations of mind” (WIPO)
- **Licenses:** a **legal agreement** between the Intellectual Property owner and the licensee with rights & restrictions.

Is it the right form of protection?

  - **Copyright**
  - **Patent**

Licenses as “conditions of use of an AI”.
Is it a possession of the developer?

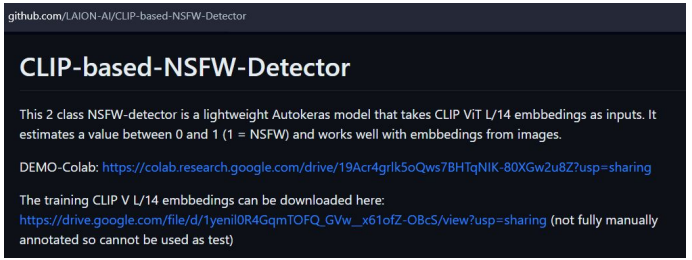
RESPONSIBLE AI
LICENSES

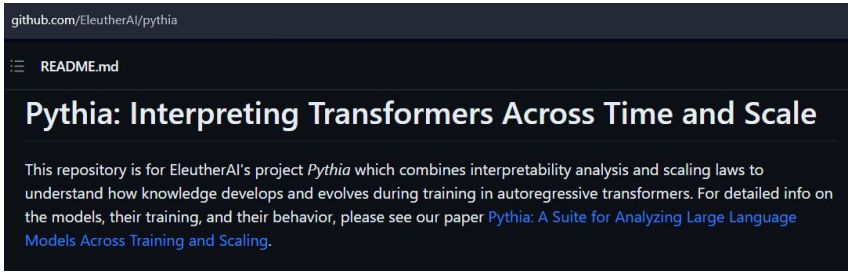
WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

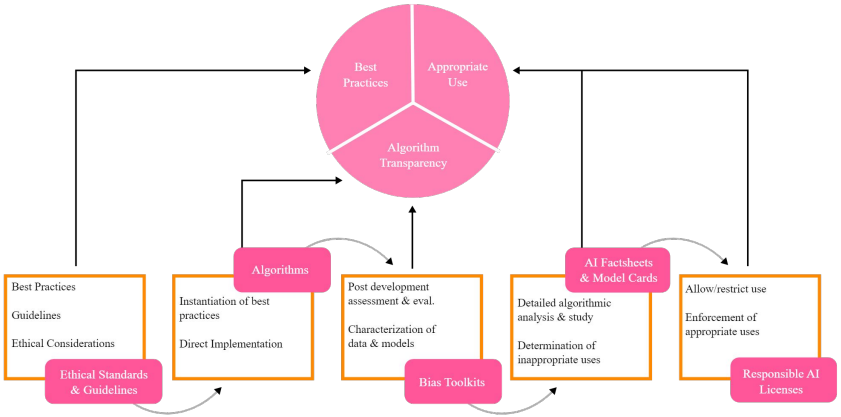
A framework for transparency & accountability

An improved license in a **standardized framework**:

- Formats for data & model **documentation**
- Unified tools & benchmarks for **evaluation**
- Requirements for **openness** and **transparency**







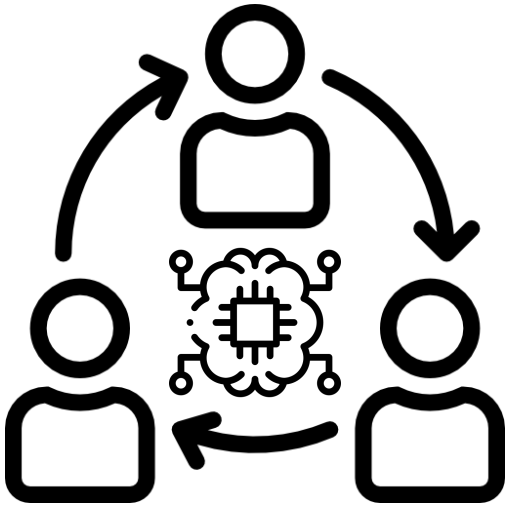
[4] "The BigScience RAIL license" Contractor et al.

Trajectories of humanity with open source AI

Promoting open collaboration

Policymaking, strategic considerations and possible long-term outcomes:

- Openness → speeds up AI development, promotes wider engagement and transparency
- Effective enforcement when political institutions agree on these principles



Strategic Implications of Openness in AI Development

UNIVERSITÀ
DI TRENTO

Thank you!

Paper #10

Authors:

Diego Calanzone (diego.calanzone@studenti.unitn.it)
Andrea Coppari
Riccardo Tedoldi
Giulia Olivato
Carlo Casonato

AISafety-SafeRL 2023 Workshop (IJCAI), August 19–21, 2023, Macao, SAR, China