Associazione Italiana Editori DAL 1869

Piero Attanasio

Areas of collaboration between publishers and libraries within the AI disruption

Fiesole
07 April 2025

Summary

Libraries and publishers in the pre-Gen AI era

Our common ground: recognizing the sources of information

Towards a trusted AI?

Is the answer to the machine in the machine again?

Conclusions

Before the generative AI

Big for (sm)all

Credits to: Paola Mazzucchi

At the IFLA «Conference Bibliographic Control in the Digital Ecosystem» (Florence, 2021) the AI topic was about **metadata + users' personal data**

P. Attanasio, 2022, New Challenges in Metadata Management between Publishers and Libraries, *JLIS*, Vol. 13 (1):116-22. DOI: 10.4403/jlis.it-12777

The **disruption** was in the use by big-techs of combination between work metadata and personal data to better profile their customers…

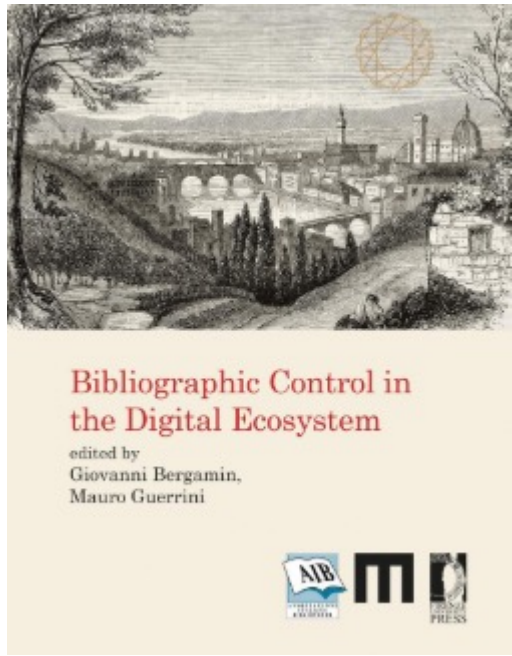
… and create dominant positions in the market

The issue was (and still is): which policy can allow access to big data by (sm)all…

… while preserving shared value such as:

respect personal data

protect freedom to publish and pluralism



The explosion of the generative AI

Is the generative AI a new form of disruption?

LLMs need **unprecedented amount of data and energy** to process data

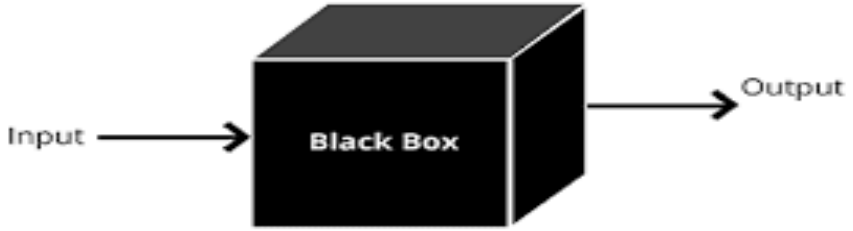
- Barriers to entry are more on the scale than on the (social) network effects

Generative AI is **general purpose**

- The emerging market seems B2B (corporate licences to third businesses)
  - rather than B2C (based on personal data and ads, like in the social media)
- Broader impact on the society (e.g. labour market, deep fake…)

LLMs **black boxes**

- European legislation on transparency and US debate about “disclosure”



Cultural concern

A society that does not recognize the value of the data source

From: «TV said…», to «Google said…», to «ChatGPT said…»

We have to invent a new *literacy* for a new age (*)

The value of recognizing data sources is a **common ground** between scholars, publishers and librarians

Is this a risk also for the young generation of scientists?

No one is immune.

For instance, in disciplines such as statistics or economics, some argue that the attention to the quality of input data decreases, as the use of AI increases.

(*) paraphrase of a quote by John Maynard Keynes

A concerning signal

In 2019, the European debate on the TDM exception focused on the possibility to store the datasets mined and used. Result: Art. 3, par. 2 of the CDSM Directive:

> 2. **Copies of works** or other subject matter made in compliance with paragraph 1 shall be stored with an appropriate level of security and **may be retained for the purposes of scientific research, including for the verification of research results**.

In 2024, the discussion about the AI-Act was about the minimum level of transparency that the AI operators are obliged to provide. Result:

> All providers of GPAI models must (…) publish a **sufficiently detailed summary** about the content used for training the GPAI model
>
> Difficult implementation of **a rule that seems like an oxymoron**
>
> **Tough ongoing discussion** on the definition «sufficiently detailed summary» in the forthcoming «Code of practice» for GDPI operators

Trusted AI?

The role of trust in the nascent GPAI market

Credits to: Francesco Ubertini, president of CINECA

If the market will be mainly at B2B level

Transparency is a key to build trust

Trust is also a key element **for research / scientific applications of GPAI**

We both, publishers and librarians, must care of transparency since we will survive only if our **customers / patrons continue trusting us**

We should play a role in the trust chain

GPAI provider → third party application → end user

Credits to: Maria Chiara Carrozza, Presidente of CNR

Expectation from technological evolution

Is the answer to the machine in the machine again?

Charles Clark, In *The Future of Copyright in the Digital Environment* (ed. by P. Bernt Hugenholtz), **1996**

If barriers to entry depend on the scale of LLMs (1st L = Large)

The development of **Medium-LM and Small-ML may reduce the oligopolistic trend**

Tough Notice and take down from a LLM is not conceivable

a vast literature on ***LLMs' unlearning*** is emerging

If transparency is difficult to enforce

The development of technologies able to understand what has been used in input analysing the output

A.V. Duarte, Xuandong Zhao, A.L. Oliveira, Lei Li, 2025, DE-COP: Detecting Copyrighted Content in Language Models Training Data, arXiv:2402.09910. DOI: 10.48550/arXiv.2402.09910

André Duarte

Conclusions

Do we have a role in this game?

The game is not over

Policy, research, culture still have a lot to say

It is worth playing

Thank you

Piero Attanasio
piero.attanasio@aie.it
www.aie.it

AIE Associazione Italiana Editori FROM 1869

We have to invent a new wisdom for a new age. And in the meantime we must, if we are to do any good, appear unorthodox, troublesome, dangerous, disobedient to them that begat us.

J.M. Keynes, *Am I a Liberal?*,

The Nation & Athenaeum, 1925, Part I (August 8, pp. 563-4) and Part II (August 15, pp. 587-8)