

# Data as the New Oil: Parallels, Challenges, and Regulatory Implications

Chiara Farronato  
(Harvard Business School, CEPR, NBER)

Making AI Work: What Firms and Workers Need  
World Bank  
January 26, 2026

# A Metaphor Dating Back to 2006

- Clive Humby (British mathematician) emphasized that data, like crude oil, is only valuable in combination with technologies that refine it into useful outputs.
- How far does this analogy go?
  - Complex regulatory and societal issues.



The Economist, 2017

# Innovation First, Regulation Later

- A recurring patterns:
  - Innovators exploit regulatory gray areas:
    - E-commerce and sales taxes
    - Ride-sharing and taxi regulation
    - Short-term rentals and lodging laws
- Regulation lags because:
  - Uncertainty of value / use of innovation
  - Fragmented jurisdictions
  - Incumbent lobbying
  - Long regulatory process

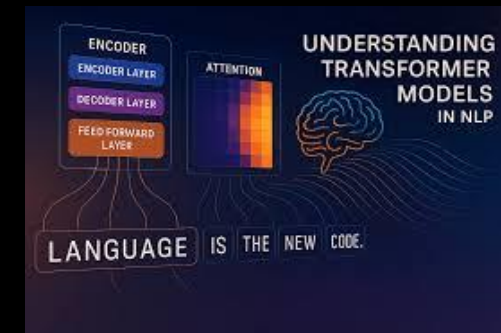


# Why the Metaphor Works

- Oil and data as dormant resources:
  - Value is unlocked through refining technologies developed over the course of many years.
  - Oil: drilling (1850-60s), refining (1860-1910s), gasoline engines (1890s-1920s).



- Data: large scale collection and storage (1900-2000s), deep learning (2010s), foundation models and generative AI (2020s).



# Where the Metaphor Breaks Down

- Oil is:

- Rivalrous
- Finite
- Tangible

- Data is:

- Non-rivalrous
- Reproducible
- Endogenously created



# Types of Data and Valuation Challenges

- Static vs. dynamic/behavioral data:
  - Static: date of birth, address → ~oil, easier to value
  - Dynamic/behavioral: searches, clicks, purchases → time-sensitive, co-created, hard to price.
- GenAI complication:
  - Recombinant nature of learning from many inputs makes it difficult to trace marginal contributions.

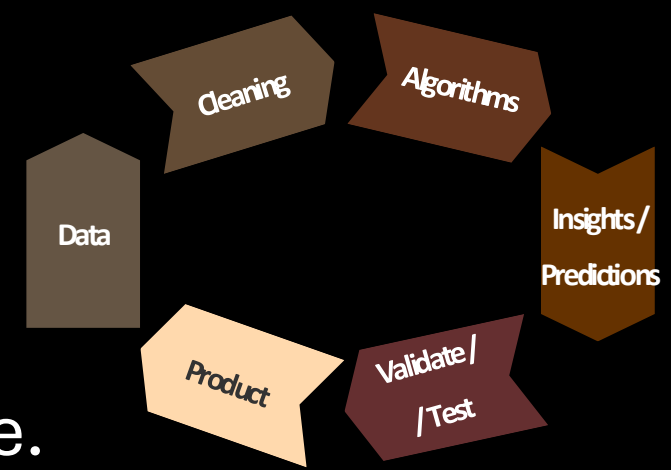


# Data and Property Rights

- GenAI meets copyright:
  - Training on text, images, music → is it fair use?
  - Two opposing effects (Waldfoegel, 2012):
    - Decrease in returns to creators (substitution)
    - Decrease in costs of creation (productivity)
  - Unclear net effect → unclear policy direction.
- Lessons from news aggregators (Athey et al., 2021):
  - Google News vs. publishers;
  - Regulatory outcome at least a decade later:
    - EU: limited snippets are allowed;
    - Australia / Canada: link taxes.



# Data and Antitrust



- Oil: market power from control of a scarce resource.
- Data: market power from scale economies in data aggregation, learning by doing in algorithms, data network effects.
- Classic antitrust tools (e.g., breakup of Standard Oil in 1911) may not translate:
  - Focus less on size and more on conduct, access, interoperability, and contestability.

# Data and Externalities

- Oil: pollution, climate change:
  - Clearly measurable and yet difficult to solve.
- Data: privacy erosion, surveillance, deepfakes, bias.

# Data and Externalities

- Oil: pollution, climate change:
  - Clearly measurable and yet difficult to solve.
- Data: privacy erosion, surveillance, deepfakes, bias.
- Bottom line:
  - The analogy is helpful for understanding the need for complementary technologies in extracting value from data;
  - But data also requires fundamentally different regulatory approaches.

Thank you.

[cfarronato@hbs.edu](mailto:cfarronato@hbs.edu)