A red die is the central focus, resting on a board game. The board features green fields with yellow sheep and circular markers containing numbers like 8, 9, and 10. In the background, there are blue and yellow game pieces. The scene is brightly lit, suggesting an indoor setting.

Is generative A.I. a game-changer for the economics of data?

Frédéric Marty

CNRS – GREDEG, Université Côte d'Azur

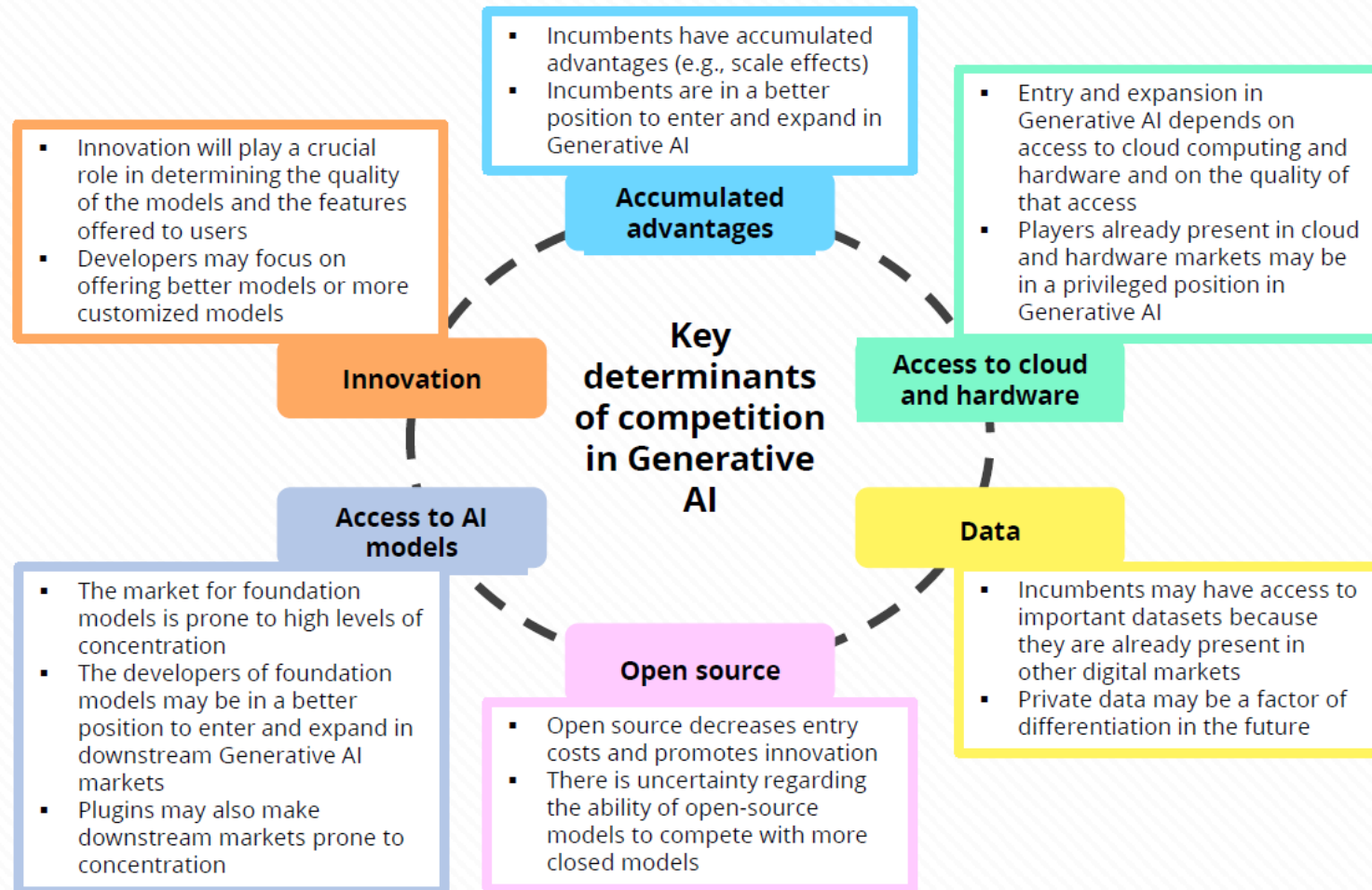
FIDES - Saint Sauveur Workshop, 15 March 2024

Can the development of generative AI reshuffle the cards of competition in the digital sector?

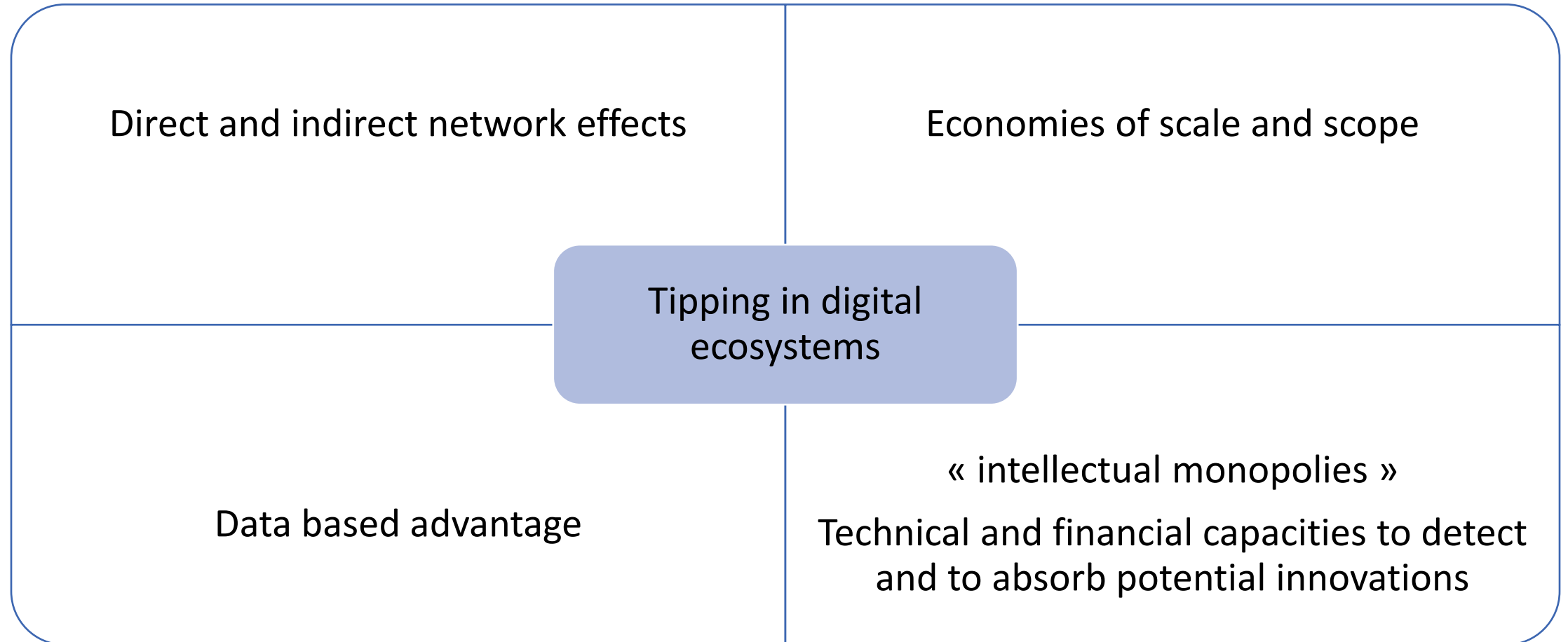
- A disruptive technology
- New players apparently distinct from the dominant firms of the Web 2.0
 - OpenAI, Midjourney, Anthropic
- A more 'precautionary' regulatory framework
- Several competitive bottlenecks
 - Data
 - GPUs
 - ...
- Incentives and capabilities to leverage existing dominant positions
 - Bundling and tying strategies
 - Self-preferencing
 - Abuses of economic dependence

Key determinants of competition

Source: Portuguese competition authority, November 2023



Data as essential facilities?





Data as essential facilities – conventional wisdom

- US Complaint against Google (2020,2023)
 - « Google intentionally exploited its massive trove of user data to further entrench its monopoly across the digital advertising market »
 - Data-based advantage:
 - A tool of consolidation in horizontal terms
 - A positive feedback loop
 - A tool of leveraging in vertical ones
 - Strategic use of data extracted in its origin market to extend dominance to adjacent ones
 - Early detection of risks and opportunities
 - Better knowledge / prediction of consumers preferences, willingness to pay,...
 - Even worst in the field of AI?
 - A monotonic relationship between data and predictive capacities?
-

FTC – comment submitted to US Copyright Office (30 October 2023)



FEDERAL TRADE COMMISSION
PROTECTING AMERICA'S CONSUMERS

[Enforcement](#) ▾ [Policy](#) ▾ [Advice and Guidance](#) ▾ [News and Events](#) ▾ [About the FTC](#) ▾ [Q](#)

[Home](#) / [News and Events](#) / [News](#) / [Press Releases](#)

For Release

In Comment Submitted to U.S. Copyright Office, FTC Raises AI-related Competition and Consumer Protection Issues, Stressing That It Will Use Its Authority to Protect Competition and Consumers in AI Markets

AI has the potential to 'turbocharge' schemes and deceptive practices, FTC contends

November 7, 2023



FTC – comment submitted to US Copyright Office

“The rapid development and deployment of AI also poses potential risks to competition. The rising importance of AI to the economy may further lock in the market dominance of large incumbent technology firms. These powerful, vertically integrated incumbents control many of the inputs necessary for the effective development and deployment of AI tools, including cloud-based or local computing power and access to large stores of training data”.

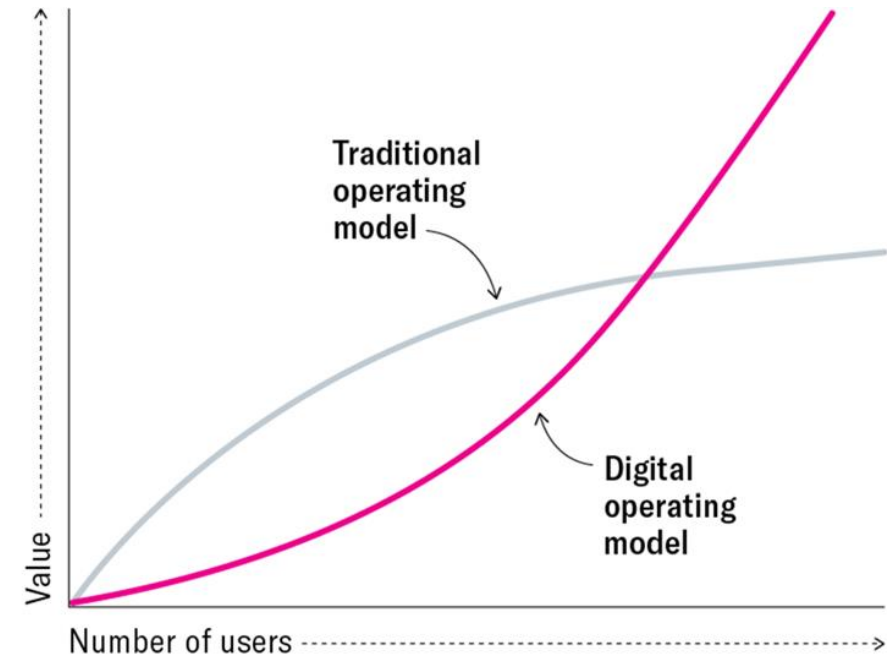
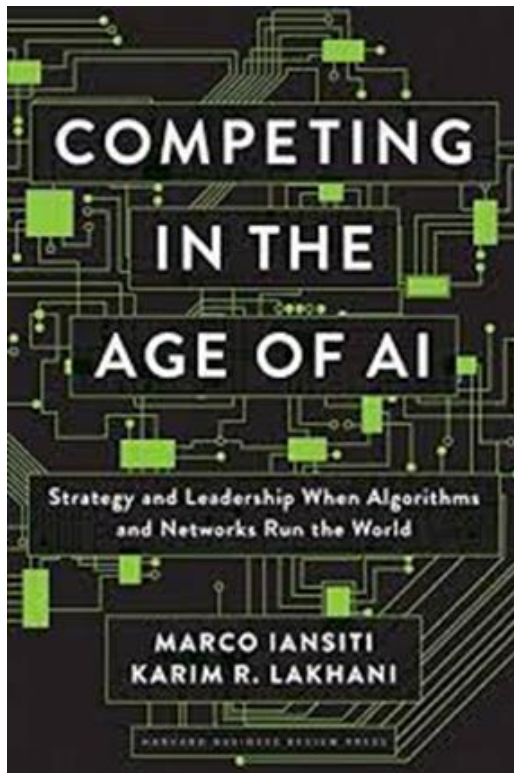


Data-based advantage and incumbency

- Furman Report (2019)
 - The innovators of yesterday are likely to remain the dominant firms of tomorrow?
 - Data “may confer a form of unmatched advantage on the incumbent business, making successful rivalry less likely”
 - *Unlocking Digital Competition*, §34
 - Rikap and Lundvall (2020)
 - “Data take the form of a new strategic resource and together with machine learning they introduce a new kind of endogenous permanent or dynamic innovation. New algorithms can be seen both as product and process innovations”.
-

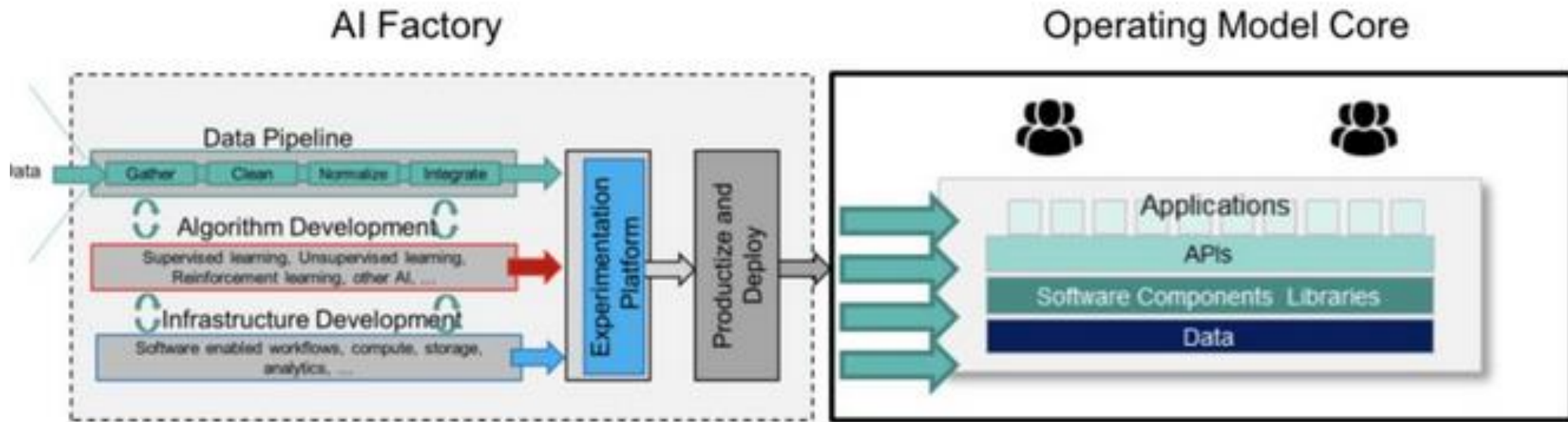
Data-based advantage and the collision model

- Gaining a competitive advantage by becoming an AI Factory (value of Data infrastructure)



From: "Competing in the Age of AI," by Marco Iansiti and Karim R. Lakhani, January–February 2020

Iansiti & Lakhmani (2020), Competing in the Age of AI



Key capacities

Financial capacities

Consolidating acquisitions, reverse killer acquisitions

Capacities to absorb new knowledge

(Rikap & Lundvall, 2021)

Infrastructural power (Jacobides et al., 2021)

Scalability related capacities

Capacity to imitate

Access to data (4V – volume, variety, velocity, veracity)

Data as essential facilities?

Data-oriented competition policies?

- Data-based theories of harm
 - Abusive extraction of data
 - Commission Notice on the definition of the relevant market for the purposes of Union competition law – 8 February 2024
 - From the SSNIP test to the SSNDQ one (see Google Android, 2018)
 - Small but significant non-transitory decrease of quality
 - Bundeskartellamt / Facebook (2020)
 - Discriminatory access to data
 - Self-preferencing cases
 - EU Commission Amazon case (December 2022)
- Data and Merger control
 - EU Commission Google / FitBit merger (December 2020)
 - FTC Complaint Meta/Within (2022)
 - FTC and DoJ merger guidelines (December 2023)



Competition Law
remedies and
regulations

Data portability

Data silos (avoiding self-preferencing) or
data lakes (mandatory data sharing)

Interoperability requirements

Control of “arbitrary and artificial” egress
fees

To the essential facilities doctrine to the
convenient facilities one?

Guideline 9 – When a merger involves a multi-sided platform...



Mergers that involve firms that provide other important inputs to platform services can enable the platform operator to deny rivals the benefits of those inputs. For example, acquiring data that helps facilitate matching, sorting, or prediction services may enable the platform to weaken rival platforms by denying them that data.

Are big data
essential in
digital markets?

Multing-homing and absence of
exclusivity

Data-brokers (FTC, 2014)

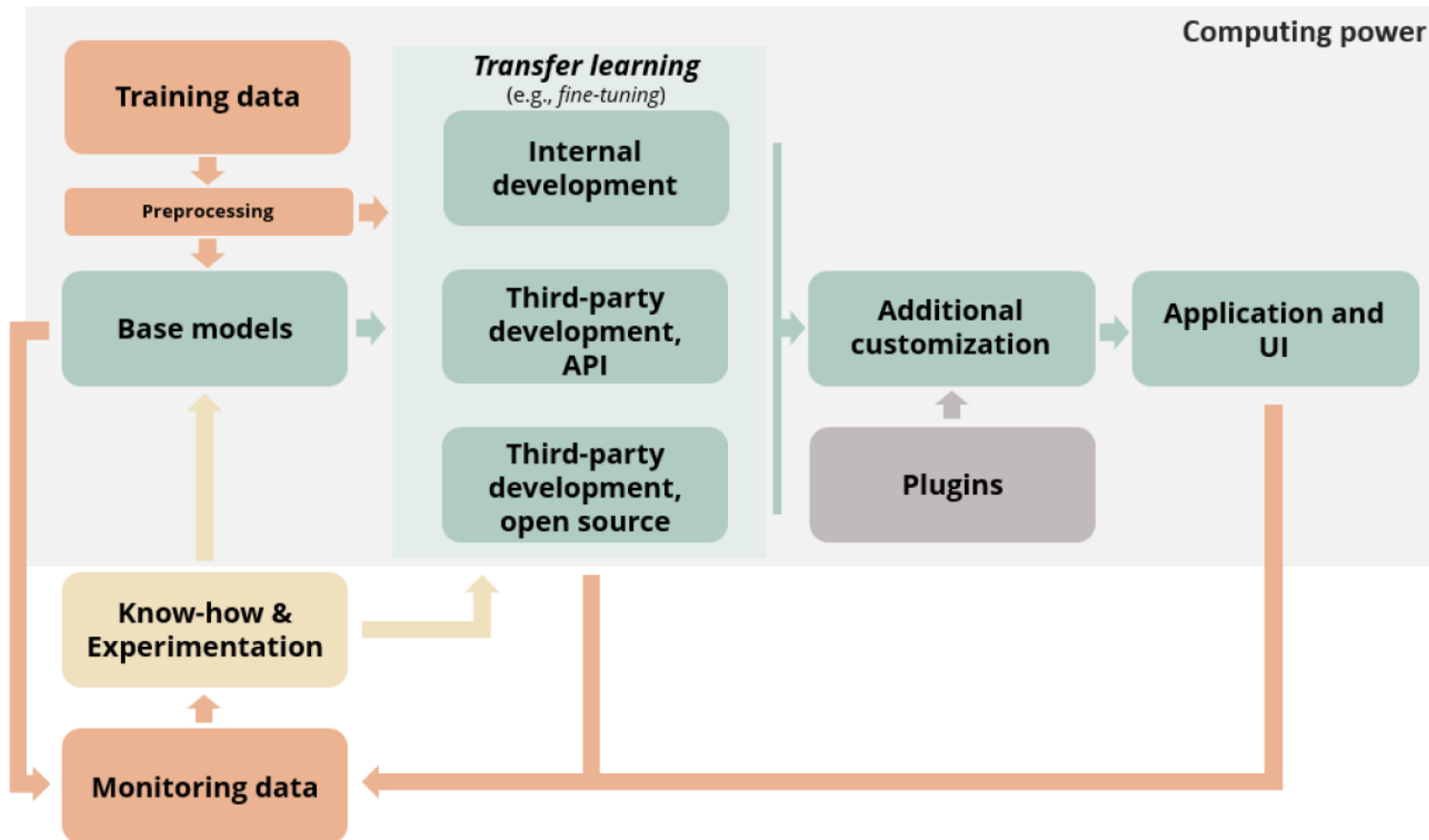
Successful entries possible
despite a data disadvantage

Data diminishing marginal
returns

Even so, data are still essential for generative AI?

- **Three cornerstones of competition** in GAI (Portuguese competition authority, 2023)
 1. Access to data
 2. Access to cloud computing or specialised hardware
 3. Access to foundation models
- *Bottlenecks of Business*
- **Three different markets**
 1. Horizontal competition on upstream markets (foundation models)
 2. Horizontal competition on downstream markets (fine-tuned and knowledge-distilled models; customized models, Plugins, apps...)
 3. Vertical relationships between foundation models developers and customized ones, between infrastructure-based services providers and downstream players...

Figure 4 - The value chain of Generative AI



Source: AdC

Value chain
of
generative
AI

Data and generative AI models

- Significant data requirements for training foundation models
 - Diversified sources
 - High acquisition costs of private data (especially high-quality ones)
 - Accessibility of public data (X vs OpenAI) – issues in terms of intellectual property, privacy, bias,...
- Not only training data but also monitoring data
 - Essential to ensure the quality and the necessary degree of personalisation
- Not only a matter of data
 - Computer power
 - Nevertheless, possible to run GAI models on less sophisticated hardware
 - Know-how and technical expertise


The case for the upstream market (foundation models)

Prolonging the dominance?

- Economies of scale and scope
- Feedbacks through the monitoring of mass end-users' deployment
- Compliance costs and relative advantage for big players
 - See the impact assessment of the regulatory costs induced by the AI Act proposal (Renda et al, 2022)
- Economic and technological dependence abuses
- Strategic use of IP rights (Vezzoso, 2024)
- Lock-in effects through partnerships
 - Open AI / MS
 - Anthropic / Google
 - Hugging Face / Amazon

Reshuffle the cards – the disruption scenario

- Open-source foundation models (direct access or APIs)
- From a “most data” model to an “enough data” one (Manne and Auer, 2024)
 - Diminishing returns (Tucker, 2019)
 - Better investing in data curation
 - Better improving training methods (Schrepel and Pentland, 2023)
- Relying on “synthetic data” (Gordon, 2023)
- Foundation models may be trained with public data, but high-profile fine-tuned models can be trained with start-up proprietary data (Garugati, 2024)
- Downstream players may win the game in specific markets



Manne and
Auer's views
regarding the
importance of
data in
Generative AI
(ICLE, 2024)

Diminishing returns (enough data)

Quality is not indexed to the number of
users or to the amount of data

Open-source databases

Algorithmic performance can compensate
a data-disadvantage

Better fitted data and better algorithm
may meet with consumers requirements



Remaining
competitive risks

Leveraging (bundling and tying)

Self-preferencing based strategies

Strategic acquisitions and partnerships

Excessive extraction of data, response
undue exploitation

Degraded access to data (discrimination in
terms of quality)

Symetric regulatory risks?

1

Damaging the competition for the GAI market by impeding Web 2.0 incumbents

Increasing compliance costs for small companies (Schrepel, 2024)

2

Preventing GAI « monopolization » by Web 2.0 incumbents

Avoiding the pitfalls of techno-conservatism e.g. the rhetoric of innovation (McLean, 2024)