Thesis in submitted in partial requirement for the degree of DPhil in Information, Communication, and the Social Sciences

# Forces of Neural Production:

## The Infrastructural Geography of Artificial Intelligence

Fabian Ferrari
Green Templeton College

Word Count: 66,213

Supervisor: Professor Mark Graham

Oxford Internet Institute
University of Oxford
Trinity Term 2022

## Table of Contents

## Abstract

It is commonly argued that only a handful of technology firms control the infrastructure that underpins contemporary forms of artificial intelligence (AI). However, little is known about how this concentration of computational resources manifests in new production geographies. This thesis asks: What forces enact, and are enacted by, AI's geographies of production? In answering this question, it develops the original framework of neural production networks: geographically dispersed but computationally enveloped production arrangements powered by artificial neural networks. Due to the impacts of AI on economic, political, and cultural interactions, neural production networks already infiltrate infrastructural life.

Drawing on a paradigmatic case study design that is primarily informed by document analysis, the thesis applies this new framework to answer three research questions: First, how do technology giants act as lead firms within AI's geographies of production? Second, how does state action shape, and is shaped by, AI's geographies of production? Third, what do AI's geographies of production imply for platformised cultural production?

The thesis answers these questions by disassembling neural production networks into three of their mutually conditioning forces. The first force is *envelopment*: Acting as industry-dominating lead firms, Amazon, Google, and Microsoft decentralise the production of AI and centralise its provision, thereby enclosing users into their proprietary ecosystems. The second force is *sovereignty*: Although the EU's Digital Single Market Strategy attempts to push back against an infrastructural dependency on the computational resources of American lead firms, its agency to reconfigure power relations in neural production networks remains limited. The third force is *hyperproduction*: Underlying the penetration of cultural life with AI-generated media, new revenue streams emerge that purport to destabilise the boundary between reality and simulation. By opening out a conceptual space that probes how those forces relate to each other, the thesis expounds a new understanding of production geographies.

This conceptual space contributes to cross-disciplinary scholarship at the intersection of platform studies and economic geography. Additionally, the arguments and insights of this thesis are relevant to policymakers and regulators facing the task of governing the diffusion of neural production networks into ever-more societal spheres.

## Acknowledgements

Within a few paragraphs, I cannot do justice to everyone that has supported, inspired, and encouraged me while working on this thesis. What follows is an incomplete attempt to express my gratitude. Mark Graham has been an excellent supervisor. From the first day, he has given me confidence to develop my ideas. Because of him, I have never felt out of place in what sometimes felt like a foreign universe. For that, and for his kindness and reliability, I am deeply grateful. I could not have asked for a better supervisor. Thank you, Mark. I am also indebted to David Feeny, my college advisor. Every conversation with David has been thought-provoking. David put things into context when I simply was not able to see a context. Between 2021 and 2022, Fenwick McKelvey has kindly hosted me as a visiting researcher at Concordia University in Montréal. This year has had a profoundly positive impact on my life and on this thesis. I greatly appreciate Fenwick's enthusiasm and support.

There are so many more people that I need to extend my gratitude to. There is my wonderful DPhil cohort at the Oxford Internet Institute: Cindy Ma, Chris Blex, Josh Cowls, Julian Albert, Marie-Therese Png, Tomas Borsa, and Yung Au. I also wish to thank my engaged and inspiring colleagues at the Fairwork project: Adam Badger, Alessio Bertolini, Callum Cant, Daniel Arubayi, David Sutcliffe, Funda Ustek-Spilda, Kelle Howson, Kristin Thompson, Maren Borkert, Matthew Cole, Michelle Gardner, Mounika Neerukonda, Nancy Salem, Navneet Gidda, Oguz Alanyak, Pablo Aguera, Patrick Feuerstein, Pradyumna Taduri, Robbie Warin, Sanna Ojanperä, Srujana Katta, and Tatiana Lopez. I feel extremely fortunate to have been a member of this project and the OII's DPhil community. Inevitably, working on a thesis is a long-term endeavour of a solitary nature, but the camaraderie and support of my DPhil cohort and my colleagues has made this journey much more enjoyable.

Gillian Rose, Peaks Krafft, Derek McCormack, Jean-Christophe Plantin, and Jonathan Gray took the time to examine different versions of this thesis and pointed me in productive directions. Their thorough review of thesis chapters has helped me to refine and strengthen my arguments. I am not only grateful for their constructive criticism, but also for their open-mindedness regarding cross-disciplinary perspectives and approaches.

As part of collaborations, conferences, and workshops, I was lucky enough learn from many inspiring and dedicated colleagues, including but not limited to: Alex Wood, Andreas Hackl, Bart Simon, Barry Colfer, Blake Hallinan, Cailean Osborne, Eleonora Mazzoli, Felix Simon, Florian Ranft, Franziska Baum, Gemma Newlands, Guillaume Dandurand, Hannah Johnston, Hong Yu Liu, James N. Gilmore, Jane Bennett, Jamie Winders, Johann Laux, João Duarte Albuquerque, Jonathan Roberge, Julian Posada, Katharina Bohnenberger, Kai-Hsin Hung, Lukas Stevens, Margath Walker, Marek Blottiere, Maurice Jones, Meaghan Wester, Mihail Caradaica, Mira Wallis, Moritz Altenried, Nola Haddadian, Nick Gertler, Niels van Doorn, Rafael Grohmann, Rob Hunt, Ryan Burns, Sophie Toupin, Sumin Lee, Thomas Vogl, Valentin Niebler, Ville Aula, William Connolly, Yuan Stevens, and Zoë Glatt.

I would also like to extend my gratitude to the staff at Green Templeton College for creating such a warm, welcoming, and supportive environment throughout my time there. At the Oxford Internet Institute, I would like to thank Chrissy Bunyan, David Sutcliffe, Duncan Passey, Laura Maynard, Marten Krijgsman, and Ornella Sciuto for their invaluable support. Greg Taylor kindly acted as my departmental advisor and greatly supported me to navigate the specificities of Oxford's DPhil milestone procedures and processes.

This thesis would not exist without the generous financial support from the Economic and Social Research Council (grant number ES/P000649/1, studentship 2094254) as well as from the University of Oxford (Scatcherd European Scholarship).

Much more could be written. Thanks to my friends in Oxford, Montréal, Berlin, and elsewhere. Thanks to my family and friends in Kevelaer. Thanks to my parents, Claudia and Hermann. Thanks to Gabrielle, my best friend. I am so grateful for all of you.

## List of Tables

## List of Figures

# Chapter 1: The Infrastructural Geography of Artificial Intelligence

## 1.1 Prologue: Project Maven and the Gospel of Wealth

> We accept and welcome therefore, as conditions to which we must accommodate ourselves, great inequality of environment, the concentration of business, industrial and commercial, in the hands of a few, and the law of competition between these, as being not only beneficial, but essential for the future progress of the race.

In his 1889 essay, *The Gospel of Wealth*, steel magnate Andrew Carnegie stressed the responsibility of the rich to donate significant amounts of their fortune to charitable causes. By administering their surplus capital in such noble ways, wealthy people may alleviate problems caused by conflicts between the rich and the poor. However, his essay is also a fierce defence of patriarchal social hierarchy and the laws of economic competition. It is thoroughly sprinkled with Social Darwinist claims on the necessity of a concentration of power for the 'survival of the fittest in every department' (Carnegie 1889). Seen this way, Carnegie's 'urtext of modern philanthropy' (Soskis 2014) entails a double message. First, economic divisions are inevitable features of social evolution. Second, as the perseverance of factory owners generates social progress and wealth, they have an obligation to support the common good through charity.

Nearly 130 years later, in March 2018, it was reported that Google had signed a contract to work with the US Defence Department on a new $250m warfare initiative, Project Maven. Google's key task was to marshal deep learning, a subset of artificial intelligence that makes use of artificial neural networks[1], to assist drone analysts in

[1] Artificial neural networks are computing systems that mimic the electrical operations of the brain's neuronal connections. Most contemporary artificial neural networks apply 'deep learning' techniques in that their layered structures of individual neurons 'are not designed by human engineers: they are learned from data' (LeCun et al. 2015: 436).

interpreting satellite imagery for improved drone-strike battlefield accuracy (Conger & Cameron 2018). Later that year, 4,000 Google employees signed a letter to CEO Sundar Pichai urging him to stop providing military technology to the Pentagon. In the wake of this controversy, it became evident that Google commissioned an outsourcing company to manually annotate the astronomic amounts of satellite imagery – a necessary task for optimising algorithms to recognise patterns in the data. An executive of this outsourcing firm later admitted that their workers 'did not know that they were working for Google or for the military, which is not an unusual arrangement' (Fang 2019). This controversy erupted at a moment in which many companies and governments had intensified their efforts to seize the perceived benefits of artificial intelligence, exemplified by the United Nations' creation of an AI for Good platform in 2017.[2]

At first glance, a juxtaposition between Carnegie's *The Gospel of Wealth* and Project Maven might seem far-fetched. This is not a claim that the robber barons of the industrial revolution and modern-day technological behemoths are equivalent in terms of their empirical modus operandi. However, I argue that this contrast provides a useful exercise to carve out two historical parallels that help to set the stage for clarifying the central research problem of this thesis. Although it is paramount to elaborate on both parallels, the focus of this thesis is only on one of them.

The first historical parallel has to do with the symbolic eradication of collective labour from societal discourses about the driving forces of technological progress. It is

[2] Organised by the Geneva-based International Telecommunication Union, as part of the United Nations, AI for Good (2022) is a 'a year-round digital platform where AI innovators and problem owners learn, build and connect to help identify practical AI solutions to advance the United Nations Sustainable Development Goals.' This example shows that influential supranational bodies promote artificial intelligence as a general-purpose and nascent set of technologies that may herald developmental and socio-economic transformations.

remarkable that Google was not able to annotate the satellite footage in-house and had to commission an outsourcing company despite the project's sensitive nature. This case shows that the economic benefits of an outsourced workforce outbalanced the risks of data misuse – a trade-off that remains absent from Google's marketing materials touting the 'democratisation of artificial intelligence.' Historically, of course, such a discursive obfuscation of devalued labour is by no means a new phenomenon, and *The Gospel of Wealth* certainly provides one of the most striking examples of this. By positioning his managerial genius as the primary engine of wealth generation, Carnegie instigated a cultural campaign that 'objected to producerist moral claims, and starting in the 1890s, reached out with a new ideology that claimed, to the contrary, that capital, not labour, creates wealth and prosperity' (Durrenberger & Doukas 2008: 216). As the late David Graeber (2018: 231) argued, the ramifications of Carnegie's cultural counteroffensive last until today, so that 'if one speaks of "wealth producers," people will automatically assume one is referring not to workers but to capitalists.'

The second parallel presents itself in the infrastructural role of corporate power. The Pentagon, a staggering apparatus that is headquartered in the world's largest office building, had to rely on Google to provide a quintessential service that it simply could not develop on its own: the use of deep learning for pattern recognition in satellite data. The case of Project Maven illustrates an argument that has gained significant traction in recent years: that is, 'the infrastructures and forms of power that enable and are enabled by AI skew strongly toward the centralisation of control' (Crawford 2021: 223). Instead of 'democratising' artificial intelligence, corporate behemoths like Google centralise the means of producing it in such a dominant manner that even the Pentagon cannot avoid relying on their services. In this regard, deploying a broader historical lens points to the possibility that the repercussions of this centralisation of control may extend beyond economic considerations such as industry consolidation and market dominance. After all, Carnegie's financing of schools, museums, colleges, and libraries indicates that his infrastructural legacy is not confined to the nuts and bolts of his monopolistic mode of industrial organisation either. Although it is not possible to anticipate whether future observers and commentators will classify the long-term consequences of technology giants' infrastructural role as having an impact akin to Carnegie's, the significance to understand and theorise this role cannot be overstated. As Dyer-Whiteford et al. (2019: 56) note, like 'earlier conditions of production, […] if AI becomes generally available, it will still remain under the control of these capitalist providers.'

But while it is commonly argued that only a handful of powerful firms own and control the infrastructure to deploy contemporary forms of artificial intelligence, little is known about how this concentration of resources relates to reconfigured geographies of production. Hitherto, the majority of critical scholarship in this domain explores the first parallel between *The Gospel of Wealth* and Project Maven: the obfuscation of collective labour. Scholars have developed numerous frameworks to theorise the gaps between the metaphoric elevation of artificial intelligence systems as magical technologies and the sobering realities of their reliance on outsourced labour. Ekbia and Nardi (2014: 7), use the term heteromation to describe 'the extraction of economic value from low-cost or free labour in computer-mediated networks' as a 'new logic of capital accumulation.' Gray and Suri (2019: 1) label that phenomenon 'the paradox of automation's last mile.' They make a case that new and superfluous human-in-the-loop tasks, such as annotating satellite images to improve Google's artificial neural networks, are continuously created and dismantled as technological innovations gain sophistication.

By contrast, there is a lack of conceptual and empirical work that problematises the second parallel between *The Gospel of Wealth* and Project Maven, the concentration of infrastructural power, vis-à-vis changing geographies of production. Undoubtedly, the extraction and exploitation of outsourced labour is a key driving force of artificial intelligence in contemporary capitalism, but it is by no means the *only* issue that is worth exploring in this context. Beyond outsourced collective labour, what *other* forces enact, and are enacted by AI's geographies of production? At its heart, this thesis is motivated by the necessity to develop a new theoretical framework for answering this crucial question – a framework that analyses the geographies of artificial intelligence not just in terms of their specific patterns of industrial organisation, but also in terms of their regulatory dimensions and cultural ramifications.

Before providing a roadmap of how the thesis goes about unfolding this answer, it is worth clarifying the significance of this research problem. A helpful starting point is that both within public discourse and academic debates, the use of historical, often epochal, analogies is a common tool to describe the dominance of technology firms and the cross-sectoral implications of artificial intelligence. For example, as Plantin et al. (2018: 306-307) write, 'Google, Facebook, and a handful of other corporate giants have learned to exploit the power of platforms […] to gain footholds as the modern-day equivalents of the railroad, telephone, and electric utility monopolies of the late 19th and the 20th centuries.' In other words, the highly concentrated ownership and control of quintessential resources that are in high demand is anything but a new phenomenon. While infrastructural monopolies are no unchartered territory, their current exponents differ from historical predecessors, so the argument goes. Such historical analogies can also be found in the media. The Economist (2022) made the following observation in a piece on the nexus of artificial intelligence and creativity:

> In a sense, that computer is merely the descendant of the power looms and steam engines that hastened the Industrial Revolution. But it also belongs to a new class of machine, because it grasps the symbols in language, music and programming

> and uses them in ways that seem creative. A bit like a human. The "foundation models" that can do these things represent a breakthrough in artificial intelligence, or AI. They, too, promise a revolution, but this one will affect the high-status brainwork that the Industrial Revolution never touched. There are no guarantees about what lies ahead – after all, AI has stumbled in the past. But it is time to look at the promise and perils of the next big thing in machine intelligence.

Consequently, artificial intelligence may become, or may already be, one of the frontiers in which a handful of technology giants acquire an infrastructural role (Plantin et al. 2018; Dyer-Whiteford et al. 2019). As ever more is speculated about such societal impacts, there is a need to investigate AI's actually-existing geographies of production. The theoretical and empirical stakes of this endeavour are significant. On a theoretical level, as Walker and Winders (2021: 165) argue, insufficient 'attention has been given to AI's potentialities and ramifications in relation to place, space, and other foundational concepts in human geography.' On an empirical level, the evidence base of how current forms of artificial intelligence, such as Project Maven's deep learning tools, are brought into being in border-crossing production networks remains scarce. This is a serious omission, given that a more comprehensive understanding of *how* industry-dominating lead firms such as Google create technological dependencies would help to appraise the efficacy of regulatory frameworks designed to curb their power. However, according to Walker and Winders (2021: 164), 'AI's complex geographies […] may necessitate new theoretical frameworks, methodologies, and analytic approaches.'

Precisely herein lies the most important contribution of this thesis. I develop a novel framework to conceptualise the infrastructural geography of artificial intelligence. Inspired by Edwards (2003: 221), I understand infrastructure not as a 'rigid background of overpowering technologies, but a constantly changing social response to problems of material production, communication, information, and control.' That said, a focus on infrastructural geographies generally aims to uncover 'the material and organisational structures of social life in diverse settings, including the role of the state and a host of other mediating institutions' (Infrastructural Geographies 2022). To paraphrase this definition for the purposes of this thesis, I wish to theorise and analyse the material and organisational structures of artificial intelligence, including the role of the state and a host of other mediating institutions. Thus, this thesis does not position itself as a narrow economic-geographical study of the clients, suppliers, and processes that co-constitute AI's geographies of production. Instead, I argue in favour of an expanded conceptual and empirical focus – one that goes beyond economic interactions in the strict sense. I do so by presenting a new analytical heuristic, the idea of *neural production networks*, that is designed to bring together three interconnected areas of inquiry:

(1) the consolidation of market dominance by technology giants as lead firms in AI's geographies of production (e.g., Amazon, Microsoft, Google);
(2) the multifaceted role of state actors and state action vis-à-vis AI's geographies of production (e.g., regulatory frameworks and public procurement);
(3) the implications of AI's geographies of production for platformised cultural production (e.g., text-to-image models and synthetic media).

The crucial point here is not to prioritise one of these areas above one another, but to think across disciplinary boundaries in the pursuit of delineating a bigger picture of the forces that enact, and are enacted by, AI's production geographies. Uncovering what those forces are and how relate to each other is the research problem of this thesis. The term "forces" is fruitful in this context because it can be interpreted as both a 'strength or energy exerted or brought to bear' (Merriam-Webster 2022) and as a 'measure of the influence that changes movement' (Cambridge Dictionary 2022). The fact that the term has the potential to coalesce empirical objects of analysis (e.g., gravitational forces) and higher levels of theoretical abstraction (e.g., forces of nature) has long made it appealing to social theorists. This semantic elasticity is best exemplified by the Marxian concept of productive forces – an idea that has never been unambiguously defined by Marx or Engels themselves, who were 'more concerned with the relation between things, and their transformation, than with their inclusion in a static and fixed formula which would inevitably reduce one's comprehension of them' (Saba 1980).

But while such etymological issues and historical analogies help to *contextualise* the research problem of this thesis, their utility in *addressing* it remains limited. Unlike the steel, coal, and coke that were at the centre of Carnegie's industrial empire, artificial intelligence is not a tangible commodity that needs to be transported on ships or trains. Instead, firms like Google can capitalise on their large-scale computational resources to instantly transmit a range of tailored services for problems such as finding patterns in satellite imagery. How does this work? What do such practices imply for regulatory debates, and what are their cultural ramifications? The next section offers an overview of how the following chapters will address those questions.

### 1.2 Thesis Structure and Chapter Overview

The thesis is divided into seven chapters. While this introductory chapter has aimed to provide a brief synopsis of the central research problem, the purpose of Chapter 2, *Neural Production Networks: A Theoretical Framework*, is to acquaint readers with the original conceptual framework that builds the foundation for this thesis. After defining three key terms (artificial intelligence, infrastructure, platform), the chapter outlines why it is not only useful but also necessary to bring a new cross-disciplinary framework for theorising AI's geographies of production into being. To substantiate this argument, it relates a well-established idea in economic geography, the global production network framework, to more recent insights from platform and infrastructure studies. Chapter 2 demonstrates the utility of conducting new research on AI's geographies of production by situating this endeavour conceptually and empirically. The chapter defines neural production networks as *geographically dispersed but computationally enveloped production arrangements powered by artificial neural networks*. It concludes by introducing the interrelated research questions that divide the central research problem into three empirically relevant, feasible, and answerable questions.

Building on these arguments, Chapter 3 offers a comprehensive account of the methodological approaches and epistemological assumptions of this thesis. The chapter elaborates on the case study-based research strategy, which draws on the paradigmatic case selection approach. The primary data collection of this thesis is document analysis, supplemented by informal background conversations with industry and policy experts, and the attendance of industry events and workshops. Chapter 3 explains my criteria for selecting case studies and data-gathering techniques, outlining my rationale for resorting to 'paradigmatic cases' (Pavlich 2010). Throughout the chapter, I reflexively discuss the impacts of the COVID-19 pandemic, which necessitated a fundamental reconfiguration of the methodology and research strategy. Before the pandemic hit, my thesis proposal involved ethnographic fieldwork in India, which had to be fully cancelled. Ultimately, the chapter addresses ethical considerations and limitations.

Chapter 4, *The Rise of Neural Production Networks*, provides an analysis of how three technology giants (Amazon, Google, and Microsoft) operate as lead firms in AI's geographies of production. It demonstrates that those companies do not operate as lead firms in the conventional sense of how economic geographers conceptualise the role of lead firms in border-crossing production networks. This discontinuity presents itself in three interconnected ways. First, technology giants combine the extraction of rent by providing infrastructure offerings to a range of other economic actors with the extensive cross-subsidisation of strategic nodes within their proprietary ecosystems. Second, Amazon, Google, and Microsoft integrate these proprietary ecosystems horizontally and vertically, thereby operating at the conjuncture of multiple geographies and algorithmic production topologies. Third, technology giants do not consolidate their dominance by centralising the *production* of artificial intelligence, such as some producers of physical commodities. Instead, they centralise the *provision* of artificial intelligence, facilitating widespread use as long as they are the ones that provide the necessary tools. Chapter 4 ends by hinting at the implications of neural production networks as a distinct form of platformised industrial organisation – one in which lead firms took up an infrastructural role that goes beyond the realm of economic production.

The contribution of Chapter 5, *The Regulation of Neural Production Networks*, is to substantiate the theorisation of this new understanding of production geographies by linking it to current regulatory proposals put forward by the European Commission, a key geopolitical actor in the area of digital policy. The chapter demonstrates that the emergence of neural production networks does not only manifest itself in the industrial operations of lead firms but also in a reconfigured policy repertoire of state actors. The chapter begins by distinguishing between four overlapping roles of how state actors in global production networks: regulator, facilitator, buyer, and producer. Drawing on the analysis of the case of the EU's Digital Single Market Strategy as a paradigmatic case of state action in neural production networks, including the world's first comprehensive approach to AI regulation, the chapter carves out two interconnected sets of tensions between those roles. First, the EU's scope of interventions as a regulator is constrained by its role as a facilitator of markets. Second, the EU's shortcomings in acting as a producer exacerbate its dependency as a buyer. Those findings indicate that there are complex reciprocal links between the repertoire of state actors and AI's geographies of production. When analysing points of regulatory leverage, it is vital to take seriously the constraints of state actors: a finding with relevance beyond the case of the EU.

Chapter 6, *Neural Production Networks and Platformised Cultural Production*, investigates what neural production networks imply for platformised cultural production by focusing on the proliferation of AI-generated media. The chapter's starting point is that it has become easy and fast to create AI-generated media, coining a new term to grasp this development: hyperproduction. Comparing two seemingly opposed use cases of hyperproduction – that is, autonomous vehicles and virtual influencers – the chapter problematises what AI's geographies of production mean for the relationship between simulation and reality in the context of synthetic media. Although those use cases seem to be at odds with each other, they exemplify the emergence of new revenue streams due to the confluence of video game engines and deep generative models, a particular neural network architecture. However, far from ushering a Matrix-style simulation that escapes the grip of social theory, hyperproduction remains not only grounded in, but also bounded by, reality: the reality of rentier capitalism.

Finally, the aim of Chapter 7, *Conclusion: The Forces of Neural Production*, is to wrap up the thesis by synthesising and extrapolating the combined insights of the previous chapters in relation to the aforementioned research problem. The chapter does this by amalgamating the findings unearthed in Chapter 4, Chapter 5, and Chapter 6 into three interrelated forces that co-constitute AI's geographies of production: the force of envelopment, the force of sovereignty, and the force of hyperproduction. Those forces are not in a state of harmony with each other. Therefore, the chapter sketches out the tensions and contradictions between them. It concludes by stressing how the theoretical framework developed in this thesis can be deployed in other contexts, shedding light on empirical and theoretical avenues for future research.

# Chapter 2: Neural Production Networks: A Theoretical Framework

## 2.1 Introduction: Key Concepts and Objects of Study

To justify the need for developing a new theoretical framework for making sense of AI's geographies of production, this chapter proceeds in three steps. First, it clarifies the definitions of three central terms that are at the core of this thesis: artificial intelligence, infrastructure, and platform. A key insight of this definitional work is that it is pivotal to be specific about the differentiation between using these terms as *objects of study* and as *concepts*. The chapter situates its cross-disciplinary understanding of these terms at the intersection of media studies and political economy, using this combined lens to stress the blurring boundaries between platforms and infrastructures.

After laying out this definitional groundwork, the chapter introduces the notion of global production networks as a useful heuristic device emanating from relational economic geography. Given that both infrastructural thinking and the idea of global production networks represent deeply relational ways of seeing the world, a synthesis between those cross-disciplinary strands of thought is fruitful for the purposes of this thesis. Resorting to three analytical categories of how scholars have investigated global production networks – value, embeddedness, and power – the chapter addresses the state-of-the-art of research on AI's geographies of production, thereby setting the stage for the theoretical and empirical contributions of this thesis.

Ultimately, the chapter goes beyond this interdisciplinary literature review by sketching out the baselines of an original theoretical framework to analyse the forces that enact, and are enacted by, AI's geographies of production: the notion of *neural production networks*. A key emphasis of this section is on the comparison between global production networks and neural production networks. The chapter concludes by introducing the research questions that operationalise this framework and serve as the guiding structure for the following empirical thesis chapters.

### *2.1.1 Defining "Artificial Intelligence"*

It has become a cliché to say that AI is omnipresent. At times, the term appears as a promise. Occasionally, it is presented as a dystopian warning. Within academic and public discourse, AI is everywhere and nowhere, symbolizing a wide variety of nascent and emerging technologies that often remain undefined. In recent years, a range of scholars has productively leveraged the term's ambiguity to craft powerful arguments that cut through disciplinary boundaries. For example, in *Atlas of AI*, Kate Crawford (2021: 227) embraces the term's evasiveness, using it for bringing together 'issues of climate justice, labour rights, racial justice, data protection, and the overreach of police and military power.' For Crawford, the term AI does not function as a clearly delineated (or delineable) object of enquiry but as a registry of overlapping questions of power – a proxy that connects mineral extraction and military contracts, shipping containers and facial recognition, warehouse floors and digital sweatshops.

From this holistic perspective, AI can be seen as 'an idea, an infrastructure, an industry, a form of exercising power, and a way of seeing; it's also a manifestation of highly organized capital backed by vast systems of extraction and logistics' (Crawford 2021: 18). But while it might not be surprising that interdisciplinary scholars such as Crawford refrain from using a narrow definition that would restrict the scope of their analysis, it is remarkable that even amongst computer scientists, there is no consensus about a singular definition (Ford 2018). Drawing on a survey of AI researchers and the analysis of policy documents, Krafft et al. (2020: 77) propose three necessary criteria for a policy-facing definition: '(i) inclusivity of both currently deployed AI technologies and future applications, (ii) accessibility for non-expert audiences, and (iii) allowing for policy implementation of reporting and oversight procedures.' According to the authors, the OECD's (2022) definition of an AI system meets these criteria:

> An AI system is a machine-based system that is capable of influencing the environment by producing an output (predictions, recommendations or decisions) for a given set of objectives. It uses machine and/or human-based data and inputs to (i) perceive real and/or virtual environments; (ii) abstract these perceptions into models through analysis in an automated manner (e.g., with machine learning), or manually; and (iii) use model inference to formulate options for outcomes. AI systems are designed to operate with varying levels of autonomy.[3]

In less technical terminology, the OECD's (2022) AI Policy Observatory adds:

> AI is a general-purpose technology that has the potential to improve the welfare and well-being of people, to contribute to positive sustainable global economic activity, to increase innovation and productivity, and to help respond to key global challenges. (…) AI also raises challenges for our societies and economies, notably regarding economic shifts and inequalities, competition, transitions in the labour market, and implications for democracy and human rights.

For Lipsey et al. (2005: 98), a general-purpose technology is 'a single generic technology, recognizable as such over its whole lifetime, that initially has much scope for improvement and eventually comes to be widely used, to have many uses, and to have many spillover effects.' Other scholars, however, reject the idea that AI is a *single* technology – or even a *technology* at all. On the one hand, there is a vast number of interventions positing that AI is an umbrella term for a variety of technologies, rather than a single one (Dyer-Whiteford et al. 2019; Crawford 2021). On the other hand, for some critical voices arguing against AI being a technology in the first place, the term is not a description of the present moment. Rather, it is being deployed to underpin the

[3] The fact that this OECD definition has changed since Krafft et al. (2020) cited it neatly shows the veritable difficulty of pinning down AI as a clearly defined policy issue.

discursive construction of a speculative future that seems *almost* there yet, nevertheless, continues to remain out of reach (Broussard 2018; Natale 2021).

While it is not possible to provide an all-encompassing review of contemporary definitions of AI, the most important point to make here is that we need to consider the distinction between using the term as an object of study vis-à-vis using it as a concept. In this thesis, I decisively approach artificial intelligence as an object of study, rather than as a concept. This means that it is not my ambition to contribute to philosophical debates on whether machines can (or should) be intelligent and whether those machines are beneficial or rather detrimental to humanity's flourishing as a whole. Instead, I will examine a specific constellation of services, techniques and models that rely on the use of artificial neural networks to make sense of the forces that enact, and are enacted by, AI's geographies of production. Artificial neural networks represent a subset of AI and not all forms of AI are powered by them. I argue that a spotlight on the role of artificial neural networks provides a valuable point of entry to defy issues of definitional ambiguity and situate my arguments historically and theoretically.

Many researchers attribute the AI hype that the world is experiencing today to the visual database ImageNet. Initially published in 2009, the dataset evolved into an annual competition, whereby researchers could compare and benchmark the accuracy of their object recognition algorithms. Participants could train their models on the object dataset, which consisted of 3.2 million human-labelled images, separated into 5,247 categories and 12 subtrees (Gershgorn 2017). In 2012, a neural network architecture developed by University of Toronto researchers (AlexNet) famously outcompeted the other object recognition models in the competition by a percentage point margin of 10.8. The eye-catching result prompted academics and practitioners worldwide to adopt the use of artificial neural networks in many domains, including but not limited to natural language processing, robotics, autonomous vehicles, machine translation, medical imagery, warfare, and cybersecurity (Alom et al. 2018). While my focus is not on the inner workings of the neural network, it is worth stating how ImageNet's creators were able to assemble their benchmark dataset in the first place. As Gray and Suri (2019: 31) state, the answer lies in their use of Amazon Mechanical Turk:

> They realised that the MTurk API[4] gave them a way to automatically distribute image-labelling tasks to people […] Ultimately, they were able to use about 49,000 workers from 167 countries to accurately label 3.2 million images. After two and a half years, their collective labour created a massive, gold-standard dataset of high-resolution images, each with highly accurate labels of the objects in the image.

Given the affordances of distributing menial tasks of data annotation at a global scale, Irani and Silberman (2013: 17) argue that Amazon Mechanical Turk transforms workers into 'a form of infrastructure, rendering employees into reliable sources of computational processing.' Although I use this example to indicate the links between two key terms around which this thesis turns, artificial intelligence and infrastructure, this thesis is *not* an exploration of the hidden collective labour without which artificial neural networks would not exist. In the next two sections, I will elaborate on this point by clarifying my understanding of the terms infrastructure and platform.

### *2.1.2 Defining "Infrastructure"*

Resembling the definitional ambiguity surrounding the term artificial intelligence, the notion of infrastructure now has such a multidisciplinary reach that it is essential to be

[4] Amazon Mechanical Turk (MTurk) is a digital labour platform that mediates value-creating interactions between consumers and service-providing workers. Application programming interfaces (APIs) 'define the list of instructions that a programme will accept and what will happen after each instruction is executed' (Gray & Suri 2019: xiv).

specific about how it is defined. I will start by shedding light on its role as a concept. Subsequently, I will describe its function as an empirical object of study by emphasising the term's fruitful intersections with the term platform.

Common definitions of infrastructure vary and are highly context-dependent. Cambridge Dictionary (2022b), for example, defines the term as 'the basic systems and services, such as transport and power supplies, that a country or organization uses in order to effectively.' Merriam Webster (2022b) uses three definitions: first, 'the system of public works of a country, state, or region'; second, 'the underlying foundation or basic framework (as of a system or organization); third, 'the permanent installations required for military purposes.' Those definitions neatly demonstrate the term's double articulation as both an abstract concept and as a signifier of a set of particular empirical phenomena such as power supplies. Against this backdrop, I explicitly situate my use of the concept in the interdisciplinary scholarship of John Durham Peters, Lisa Parks, Paul N. Edwards, and Geoffrey C. Bowker. As Peters (2015: 31) puts it:

> Infrastructures can be defined as "large, force-amplifying systems that connect people and institutions across large scales of space and time" or "big, durable, well-functioning systems and services." Often they are backed by states or public-private partnerships that alone possess the capital, legal, or political force and megalomania to push them through. […] Because of their vast technical complexity and costs, infrastructures are often cloaked from public scrutiny, their enormous risks and unintended consequences shielded from open debate. […] The bigger the infrastructure, the more likely it is to drift out of awareness and the bigger the potential catastrophe. […] The greatest thinkers of infrastructure were never interested only in the gear; they always wanted to know why awareness of essential things so quickly fades into "beaten paths of impercipience."

Infrastructure, in this scholarly tradition, is not defined by *what* it is, but rather by *how* it is: large, force-amplifying, durable, opaque, shielded, impercipient, etc. In my eyes, this reorientation from describing the essence of things (regardless of whether or not things have an essence in the first place) to describing the features of things is not a weakness but the key strength of the term. The fact that different interpretations of the term are conceivable does by no means make it elusive or vague; it makes it elastic and lucid. Peters' (2015: 29) ability to develop a holistic yet thoroughly coherent argument that complicates the relations between 'ships, fire, night, towers, books, Google, and clouds' is an apt example here. And so is Crawford's (2021: 15) invitation to take the reader on a 'series of journeys to places that reveal the makings of AI.' The explanatory power of deploying the concept of infrastructure in this line of thought does not stem from the orthodox application of established frameworks or methods[5] but the creative synthesis of ideas, sources, phenomena, feelings, and patterns.

Focusing on the implications of this assumption with respect to common objects of analysis in media studies, Parks (2017: 160) reminds us about the stakes of adopting such an infrastructural disposition. 'When consuming or critiquing media', she argues, 'it is vital to think not only about what media represent […] but also more elementally about what they are made of and how they arrived.' Along similar lines, Bowker et al. (2009: 103) note that 'when dealing with infrastructures, we need to look to the whole array of organisational forms, practices, and institutions which accompany, make possible, and inflect the development of new technology.' As such, instead of theorising technological systems as isolated, stable, or monolithic entities, infrastructural thinking highlights their dynamic relational nature. Infrastructures are not a passive backdrop of human action but are firmly entangled with a staggering multitude of social, political, cultural, ecological, and elemental relations. Some of them only come to light at particular moments of infrastructural disruption. For example, in signifying that

[5] Crucially, this is not to dismiss the importance of the ethically sound use of frameworks and methods. Chapter 3 engages with those methodological issues in more detail.

Google's infrastructural role in Project Maven only became publicly visible when a critical point was reached triggered by employee protests, this case illustrates one of Star's (1999) defining characteristics of infrastructure: moments of disconnection, failure, and breakdown serve to make infrastructure more visible.

I argue that this spotlight on always-unfolding relations and the fierce rejection of seeing complex technological systems as isolated objects of analysis is indispensable for examining AI's emerging geographies of production. To think infrastructurally, in other words, is to probe the actually-existing networks and relations of contemporary AI systems rather than speculating about potential future outcomes. Given that no facet of infrastructure is 'inherently either visible or invisible' (Star & Strauss 1999: 9), it is vital how infrastructures are operationalised. For Larkin (2013: 330), infrastructures 'are not, in any positivist sense, simply "out there.' The act of defining an infrastructure is a categorizing moment.' He concludes that, when argued with due consideration and thought, the concept of infrastructure draws attention to the epistemological and political commitments involved in determining what one perceives as infrastructural and what one excludes from consideration. Consequently, a decision to empirically focus on a particular infrastructural dimension of artificial intelligence (e.g., the role of artificial neural networks or computational resources) is also a decision against focusing on another infrastructural dimension (e.g., the extraction and modification of rare earth minerals for producing the hardware of computational devices).

This leads us to consider the relationship between infrastructure and platforms. While it has long been a focus of infrastructure studies to understand the materiality of things, sites, scales, people, and processes within networked systems of power (Parks & Starosielski 2015), an influential body of literature has grown around the argument that a handful of digital platforms have acquired a range of features that were traditionally in the realm of infrastructure[6] (Plantin et al. 2018; Nieborg & Helmond 2018; Langlois & Elmer 2019; Peck & Phillips 2020; Gray et al. 2020). Plantin and Punathambekar (2019: 2), for example, contend that technology giants such as Google, Tencent, Alibaba and Amazon have reached such tremendous levels of indispensability, scale, ubiquity, and reliability, so that they 'function as vital infrastructures in the world at large, such that living without them shackles social and cultural life.' What are platforms and why is it pivotal that they now seem to be characterised by infrastructural features?

### *2.1.3 Defining "Platform"*

It is not straightforward to define what a platform is; there is no singular definition. In a much-cited article on the politics of online service providers, Gillespie (2010: 349) notes that the term's 'discursive positioning depends on terms and ideas that are specific enough to mean something, and vague enough to work across multiple venues for multiple audiences.' For example, the term platform can refer as much to a physical structure on which people or things can stand as it can designate a figurative platform or a political platform for the exchange of ideas (Gillespie 2010). Much has been written about how companies strategically exploit the term's ambiguity to immunise themselves from regulatory oversight (van Dijck et al. 2018; Langley & Leyshon 2017; Gillespie 2018). I deploy the platform concept in order to theoretically underpin my analysis of how technology giants act as lead firms within AI's geographies of production. For this reason, I situate my understanding of the concept in the cross-disciplinary literature at the intersection of media studies and political economy. McKelvey (2018: 220) presents

[6] This focus on questions of scale, power and relationality is indicative of an infrastructural turn in sociology, media and communication studies, geography, and related disciplines. Critics of certain aspects of this turn point to its alleged 'tendency towards banality and vagueness – including dubious defences of vagueness itself' (Hesmondhalgh 2021: 2).

a useful description of how a platform can be conceptualised:

> Platforms are a "convergence of different technical systems, protocols and networks that enable specific user practices and connect users in different and particular ways." The concept of the platform helps explore the operations of control in heterogeneous systems with horizontal and vertical factors.

Scholars that deploy a critical political economy lens[7] for making sense of platforms often note that a platform is not a stable or fixed monolith but instead an evolving entity that may expand horizontally and vertically (Nieborg & Helmond 2018). The process of horizontal integration has been theorised as the 'infrastructuralization of platforms' (Plantin et al. 2018: 295). It encompasses the mechanisms by which platform orchestrators 'morph into infrastructures for users by establishing themselves as vital obligatory passage points' (Grabher 2020: 252). For example, platforms extract and circulate user data through application programming interfaces (APIs) to third parties and app developers, infamously demonstrated by the Cambridge Analytica scandal. The process of vertical integration, by contrast, can be understood as the 'platformization of infrastructure' (Plantin et al. 2018: 295). This dimension indicates the integration of platform features such as cloud computing infrastructure, hardware, and proprietary data analytics services as part of a corporate ecosystem (van Dijck et al. 2018). An example

[7] According to Mosco (2018), 'political economy approaches examine the power relations that comprise the production, distribution, and exchange of resources.' Some scholars argue that platforms call into question conventional frameworks to analyse these power relations, necessitating an analysis of the discontinuities of what has been called the conjuncture of 'platform capitalism' (Srincek 2016). Others, however, rather see the rise of platforms as being part of a longer history of 'rentier capitalism' (Christophers 2020). As I will argue in Chapter 6, those perspectives are not mutually exclusive. But for now, it is paramount to clearly spell out the value of a critical political economy lens and the analytical distinction between horizontal and vertical integration as platform features.

of this is Apple's payment service Apple Pay, which does not work without a built-in NFC chip as a proprietary hardware resource in iPhones. Crucially, those two modes of expansion are not mutually exclusive but instead may occur simultaneously, shifting the nature of how corporate power operates (Plantin et al. 2018).

Consequently, a review of this scholarship does not only expose a tendency to point out the blurring boundaries between platforms and infrastructures but also the difficulty of distinguishing between using the term platform as a concept and as an object of study. When media scholars use the term, it is not always clear as to whether they mean a platform in an abstract sense, such as a 'convergence of different technical systems' (McKelvey 2018: 220), or rather in an organisational sense, as a way to denote a particular type of corporate entity. Within public discourse, however, the terms tend to be used in relation to firms. For example, many of the world's most valuable technology giants, including Alphabet, Amazon, and Microsoft, are often called 'platform giants' (Foroohar 2021). In harmony with my relational use of the concept of infrastructure, I explicitly deploy the term platform for operationalising the *mechanisms* by which AI's geographies of production are brought into being. I refrain from using it as a positivist seismograph with the aim to unearth the essence of platforms. In a nutshell, a critical political economy approach to platforms is useful for probing how technology giants act as lead firms in border-crossing networks of economic production.[8]

Against that backdrop, what concepts and methods are needed to adequately grasp the spatial dimensions of AI's geographies of production? How can particular spatial ontologies be mobilised in order to make sense of the infrastructural role of technology giants in cases such as Project Maven? How to claw apart the conceptual

[8] In Chapter 4, I will expand on this aspect by clarifying the relation between AI and technology giants, showing that the infrastructural provision of AI is highly centralised.

distinction between material and immaterial forces of production, and how useful are those dichotomies in the first place? Those questions refer to the spatial sites and scales at which different kinds of value are created and captured; how technology giants operate as lead firms to expand their corporate ecosystems; and how they articulate and consolidate their power. As the next section aims to show, the conceptual vocabulary of economic geography is helpful for addressing these questions.

### 2.2 Global Production Networks

Arguably, deploying a concept that includes the ubiquitous-yet-evasive term *network* risks a similar definitional ambiguity as AI, infrastructure, and platform. I aim to alleviate this ambiguity by situating the concept in economic geography. I will begin by outlining the global production network framework and contextualising it vis-à-vis other ideas in economic geography. Subsequently, I will illustrate the analytical strengths of the global production network framework by using it to examine previous research on the geographies of AI, infrastructures, and platforms. Building on this work, I will then introduce a new heuristic device, the notion of neural production networks, which provides a coherent framing for the later chapters of this thesis.

At its most general level, a global production network can be defined as 'the nexus of interconnected functions and operations through which goods and services are produced, distributed, and consumed' (Hess 2018: 2). The concept has been frequently iterated by economic geographers (Coe & Yeung 2019; McGrath 2018; Werner 2019; Grabher & van Tuijl 2020). They have taken inspiration from several disciplines and intellectual legacies, including economic sociology, international political economy and world-systems theory, and actor-network theory (Table 2.1). It is not my ambition to provide a comprehensive etymological history of the term (Coe & Yeung 2015), but instead to apply its analytical categories to the study of AI's geographies of production.

As Werner (2019: 948) puts it, global production network studies typically examine the 'value, power and embeddedness of transactionally linked but geographically dispersed production arrangements.' Before outlining these dimensions in more detail and putting them in conversation with literature on the geographies of infrastructures and platforms, it is worth highlighting what distinguishes the idea of global production networks from other established conceptual frameworks in economic geography.

Table 2.1: Ascendants of global production networks. Source: Coe et al. 2008.

| | *Global commodity/value chains (GCC/GVC)* | *Actor-network theory (ANT)* | *Varieties of capitalism (VoC)* | *Global production networks (GPNs)* |
|---|---|---|---|---|
| Disciplinary background | Economic sociology<br>Development studies | Sociology of science | Political science | Economic geography |
| Object of enquiry | Interfirm networks in global industries | Heterogeneous networks of association between human and nonhuman actors | Variations in national institutions and systems of capitalism | Global network configurations and regional development |
| Orienting concepts | Value-adding chains<br>Governance models<br>Organizational learning | Relational networks<br>Human and nonhuman actors | Organization of production regimes | Value creation, enhancement, and capture |
| | Industrial upgrading and rents | Immutable mobiles<br>Topological surfaces | Market-related institutions<br>Modes of coordination | Corporate, collective, and institutional power<br>Societal, network, and territorial embeddedness |
| Intellectual influences | MNC literature<br>International business<br>Trade economics | Sociology of science<br>Post-structuralism | Classical political economy<br>Institutionalism | Relational economic geography<br>GCC/GVC, ANT, VoC |

The heterodox field of economic geography possesses a diverse vocabulary to operationalise relations and processes that link disparate places to exploit dependencies and in which one firm provides the infrastructural foundation for other actors, and, in so doing, enacts highly opaque spatial relations and formations. To name a few examples, one can refer to 'global value chains' (Bair 2005), 'global supply chains' (Tsing 2009), as well as 'global commodity chains' (Gereffi et al. 2005). What unites all those ideas and spatial metaphors is their spotlight on the analysis of transnational configurations of economic production, distribution, and consumption. At first glance, however, the direct applicability of those ideas to the analysis of the forces that enact AI's geographies of production seems limited because scholars initially developed them to underpin the empirical analysis of *physical* or *tangible* goods and products.

Against that backdrop, for the purposes of this thesis, the key strength of the global production network framework lies in its rethinking of conceptualising economic production as a 'process of sequential transformation from inputs, through stages of transformation to outputs' (Coe et al. 2008: 274). This means that the heuristic device of global production networks challenges conceptions of the production process as a linear progression, a step-by-step process with clearly distinguishable stages on a pre-determined spatial or temporal axis. As Coe et al. (2008: 272) write,

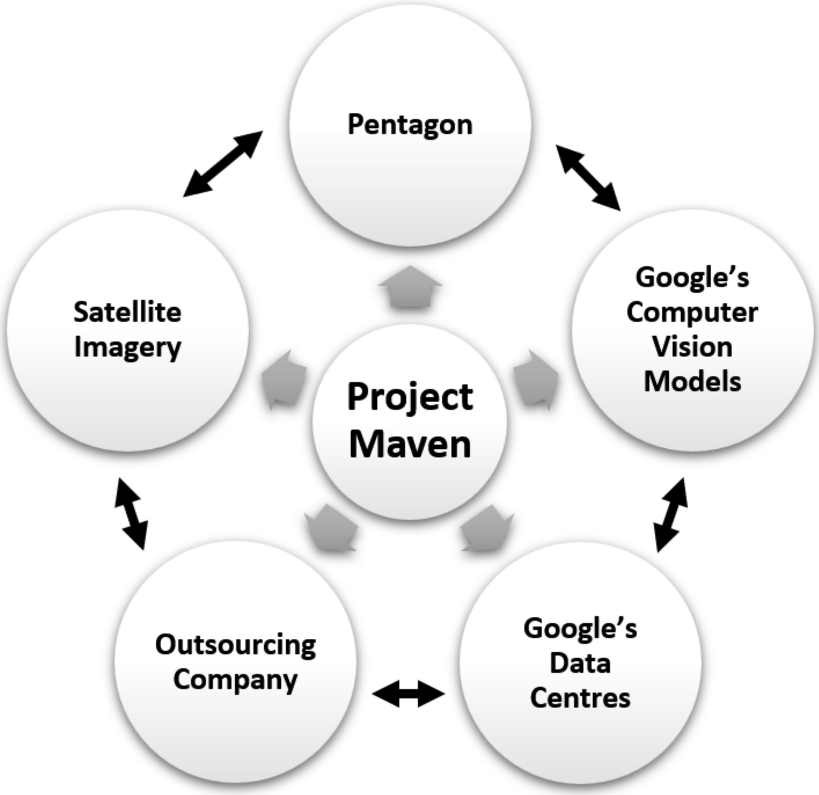
> global production networks are inherently dynamic; they are always, by definition, in a process of flux – in the process of becoming – both organizationally and geographically. The spatio-temporality of production networks, therefore, is highly variable and contingent. […] [As a result of this], economic processes must be conceptualized in terms of a complex circuitry with a multiplicity of linkages and feedback loops rather than just ''simple'' circuits or, even worse, linear flows.

Therefore, what brings together global production networks and infrastructural thinking is that both strands of literature represent deeply relational ways of seeing the world. Due to its grounding in actor-network theory[9], the global production network framework implies that its modes of analysis are by no means limited to fixed or stable points in the two or three-dimensional geometry of Euclidean space – that is, the idea that space must be conceptualised in relation to discrete horizontal territories, 'in which positions can be taken' (Mol 2002: 144). Returning to the example of Project Maven,

[9] I do not wish to imply that actor-network theory provides a coherent framework or a unified mode of research in general. Instead, I refer to it as one of the theoretical influences of global production networks to stress the concept's cross-disciplinary roots.

one might map the circulation of satellite images from the Pentagon to Google's data centres; from there to the outsourcing company and to its data annotation workers and vice versa (Figure 2.1). This circuitry of production is not a *sequence* of consecutive transactions but rather a *convergence* of heterogenous relations that are so opaque and interlocked that it seems futile to disentangle them meticulously.

Figure 2.1: Simplified circuitry of Project Maven's production network.




I use this example to visualise the value of transcending dualisms such as material/immaterial, embodied/virtual, and structure/agency when conceptualising AI's geographies of production. Project Maven's spatial sites and scales are material *and* immaterial, embodied *and* virtual, shaped by structure *and* agency, making it impossible to operationalise them based on simplistic dichotomies. For this reason, I argue that an integration of the global production network framework and infrastructural thinking is not only hypothetically possible but also conceptually appropriate.

Before specifying my original approach of synthesising the combined insights of the previous sections into a new conceptual framework, I argue that the dimensions of value, embeddedness, and power (Werner 2019) provide a fruitful roadmap to analyse previous research on the blurring boundaries between platforms and infrastructures in the context of artificial intelligence. By articulating the state-of-the-art in this domain, I set the stage for clarifying the most important contributions of my thesis. My discussion of gaps within the cross-disciplinary literature in this area is designed to open up fruitful avenues to enrich it, rather than to unfairly critique it.

### *2.2.1 The Dimension of Value*

Scholars have long shown that cross-border movements of labour, money, technology, and products produce unequal geographies of value production and distribution (Dicken 2004; Peck & Tickell 2002; Harvey 2006). Those strands of work challenged neoliberal narratives that the world and geography itself are now 'flat' (Friedman 2005) due to the fruits of technological progress. As such, the rationale of geographers for considering the dimension of value when it comes to conceptualising global production networks lies in its potential to encompass both the extraction of *surplus value* created through the labour process of workers and different forms of *rent* that companies can capture and enhance within production networks (Hess 2018; Werner 2019).

Typically, production network scholars deploy a broad understanding of rent by distinguishing between different forms, such as 'technological rents, achieved through access to advanced product and process technologies; brand rents, realised through a strong market presence […], organisational rents, […] relational rents […] and trade policy rents.' I intentionally use a simplistic way to introduce the contested notions of value and rent, which are at the heart of longstanding debates (Havice & Pickles 2019; Bair & Werner 2011; Pickles & Smith 2016). In the wake of those controversies, value has been interpreted in different ways, echoing tensions that stem from the intellectual legacies of Marxian value theory and Ricardian rent theory.

In recent years, literature on patterns of value distribution in global production networks has started to consider insights from platform and infrastructure studies. Lai and Samers (2020: 6), for example, examine the financial technology (FinTech) sector to argue that platform-based service providers such as Foxconn are aimed at 'locking-in users by providing a growing ecosystem of business and financial services.' With the entry of new business models to mediate financial services, they argue, the line between what constitutes *production* and what denotes *finance* is becoming blurred. In a similar vein, in concluding one of the most sophisticated papers on the production networks of platforms, Coe and Yang (2022: 327) note that platforms raise 'a series of compelling economic–geographical research questions in relation to ongoing processes of industrial restructuring and shifting patterns of value creation and capture therein.' In other words, these scholars see a strong potential for new research on how the features of platforms call into question well-established theories of value and rent.

With that in mind, it is worth noting that the cross-disciplinary literature on the nexus between platforms and value points to different strategies of how value creation and capture occur in this context. Fernandez et al. (2020), for example, present a highly useful summary, differentiating between three interrelated strategies:

(1) network effects (i.e., the larger the platform, the more value can be captured, which represents the most commonly discussed strategy);
(2) intrafirm cross-subsidisation (i.e., a company accepts financial losses in one particular domain to increase its market shares and network effects);
(3) optimising user engagement (i.e., a platform company maximises data collection and value capture by keeping users on the platform for as long as possible).

With respect to AI's geographies of production, research on the distribution of value remains in its infancy. This is partially due to the fact that the relevant literature typically focuses on the role of outsourcing and hidden labour in relation to artificial intelligence. Tubaro and Casilli (2019), for instance, demonstrate that the automotive industry creates a staggering demand for human-labelled training data that serves as the backbone for the artificial neural networks of autonomous vehicles. Car manufacturers such as Tesla and Daimler are the most important lead firms in this area. Along similar lines, Schmidt (2019) shows that during the country's economic crisis, Venezuela has become a key site for outsourced data work to take place. The emergence of economic relations between workers in Venezuela and the producers of autonomous vehicles neatly exemplifies Massey's (1995: 3) insistence that spatial divisions of labour give rise to 'whole new sets of relations between activities in different places and new spatial forms of social organisation.' More generally, this strand of research shows that the integration of artificial neural networks into the circuits of economic production is a collective effort that relies on a globe-spanning division of labour – and by no means the isolated work of well-paid data scientists at lead firms.

Nonetheless, no comprehensive study exists that analyses the mechanisms of value capture within AI's geographies of production by focusing on the infrastructural role of technology giants acting as lead firms in global production networks. This area of enquiry remains a crucial blind spot in scholarly debates. While established notions of value are useful entry points to underpin the endeavour of changing this status quo, the corporate mechanisms of platforms cast doubt on the boundaries between categories like 'technological rents' and 'brand rents' (Hess 2018: 3). Building on these insights, I will now focus on the role of embeddedness as a productive bridging concept between considerations of value and power in global production networks.

### *2.2.2 The Dimension of Embeddedness*

Embeddedness was taken up in economic geography by Dicken and Thrift (1992) to challenge the misleading neoclassical view of economic actors as abstract entities that are detached from any social relations – the rational *homo economicus* that lives in an ethereal sphere separated from the rest of us. Foster and Graham (2016: 4) write that embeddedness helps understand 'the coupling between networks of global production and the diverse locations within which production networks are grounded.' As Weller (2006) argues, for an adequate analysis of embeddedness, it is pivotal to consider the shifting and uneven character of embedded relations at multiple scales. With that in mind, a common way to make sense of how production networks are embedded at multiple scales is by investigating three potential layers (Hess 2004):

(1) territorial embeddedness (i.e., the spatial sites where particular value-creating or value-capturing activities are geographically/physically 'anchored');
(2) societal embeddedness (i.e., the implications of institutional, social, or cultural influences on the configurations of global production networks);
(3) network embeddedness (i.e., the social ties between heterogenous actors or entities that collectively make up the global production network);

More recently, an emerging body of literature has explored the geographies of the platform economy through the analytical lens of embeddedness, arguing that platforms can be simultaneously embedded and disembedded from the space-times they mediate (Wood et al. 2019). In South Africa, for example, the ride-hailing platform Uber could avoid a legal challenge by workers; not because the workers had no valid claim, but rather because they made it against the wrong company: Uber International Holding(s) BV, a Netherlands-based company. Uber immunised itself from legal claims by South

Africa-based workers, who are unlikely to take up their case in a Dutch court (Graham 2020). Nevertheless, Howson[10] et al. (2021: 14-15) note that 'despite their deliberate elusiveness, neither are platforms in any way immaterial: they are always embodied and grounded in different places and social relations.' In short, embeddedness remains a relevant category, but its empirical manifestations are in flux.

A spatial metaphor to demonstrate this aspect for the purposes of this thesis is the imaginary of the cloud. Peck and Phillips (2020: 82), for instance, note that 'the matrix-like capacities of platforms mean that they can seem to be everywhere but at the same time remain placeless – their preferred address, appropriately enough, being "the cloud."' The cloud is a way to reduce complexity. According to Amoore (2018: 5), for 'describing the advent of processes at scales that appear to transcend the observational paradigm, […] the visualisation of a figurative cloud stands in for the complexity of the internet.' Clouds typically float above our heads; they are intangible yet ubiquitous. Their data flows within the cloud seem to be detached from material infrastructures, transcending physical laws and spatial constraints. But the heuristic of the cloud hides the fact that most data centres are placed in places with plenty of land and water for cooling, low taxes, and cheap energy (Amoore 2018). Against that backdrop, dissecting the cloud by focusing on its scales and sites of embeddedness can unveil its constitutive elements: undersea cables, pipelines, satellites, metals, and more (Starosielski 2015). Similarly, for grasping the coupling between AI's geographies of production and their manifestations of spatial grounding, there is a need to think creatively about how the category of embeddedness may present itself in this context.[11]

[10] I am a co-author of this paper. But the objectives of this thesis substantively differ from my collaborative work in this area, which has focused on the geographies of work.

[11] At the same time, previous attempts to enumerate the embeddedness of AI systems show how 'categorising moment[s]' (Larkin 2013: 330) characterise manifestations of infrastructural

Juxtaposing the corporate trajectories of AI in North America, Europe, and Asia, Dyer-Whiteford et al. (2019: 80) postulate that 'how AI is introduced around the planet varies according to capital's spatial organisation.' Although there is a lack of work that analyses this spatial organisation, research on the geographies of data centres provides valuable insights. Brodie (2020: 15), for example, examines Irish data centres to argue that the 'global data centre industry manages the movement of financial and data circulations like supply chain managers, adapting to spatial and cultural dynamics' while navigating messy on-the-ground contingencies. In their study of Singaporean data centres, Neilson and Notley (2019: 25), highlight that 'because the concepts of chain, flow and network are metaphors that seek to describe complex material relations, they have limited applicability in studying the different kinds of connectivity enabled by data centres and related infrastructures.' Such arguments not only illustrate the usefulness of synthesising key insights from platform/infrastructure studies and the global production network framework; they also point to the need for a refined conceptual vocabulary that accounts for AI's geographies of production as objects of study. Before substantiating this point, the dimension of power deserves attention as well.

### *2.2.3 The Dimension of Power*

To chase the history of the term power in geographical thought is to chase a ghost. As Clegg (2015) puts it, 'power has been a core concept for as long as there has been speculation about the nature of social order.' Power is ubiquitous, omnipresent, and difficult to pin down. In relation to work on global production networks, the dimension

thinking. In their map, *Anatomy of an AI System*, Crawford and Joler (2018: 5) admit that 'the true scale required to build AI systems is too complex, too obscured by intellectual property law, and too mired in logistical complexity to fully grasp.'

of power focuses on 'the way in which some firms use their power over other firms to control or 'drive' the overall system' (Coe 2016: 332). Empirically, most studies focus on inter-firm relations, such as between coffee retailers, roasters, and suppliers. Foster and Graham (2016: 7) add that global production networks scholars typically imagine power as being 'exercised in three different ways – corporate (firms), institutional (state, global institutions) and collective (unions, NGOs), all of which imply wilful agent-led activities with little room left to conceptualize the digital.' Given these limitations, it is important to theorise how the features of platforms as lead firms reshape the articulation and negotiation of power in global production networks.

Drawing on the example of freely accessible software development kits, Blanke and Pybus (2020: 11) argue that 'those platforms are the most powerful ones that are able to generate services […] other digital industries cannot do without anymore, as they depend on them to make profits.' This argument implies that those companies that control the deepest infrastructural layer that figuratively sits beneath other industries should be in the spotlight of power analyses. In other words, those companies form 'the heart of the ecosystem upon which many other platforms and apps can be built' (van Dijck et al. 2018: 12). Those insights imply that platforms exert power in ways that are qualitatively different from other lead firms in global production networks. As Plantin et al. (2018: 307) put it, 'system builders that dominated infrastructures of the past may be replaced by ecosystem-builders that leverage programmability and interconnection to achieve control.' What is of interest is not simply the provision of a service or product but instead the virtuous cycles commensurate with platforms' 'jurisdiction-spanning, infrastructural presence in the economies of market exchange and everyday life' (Peck & Phillips 2020: 77). In the light of this expanded analytical focus, the always-existing question of how power works has not lost any of its relevance.

But it is important to note that platforms *replace* global production networks; they *enact* global production networks (Coe & Yang 2022). Weatherby (2018) writes that ‘the platform has gone from an operating and gaming system to a global economic factor, from a metaphor to the mechanism of cultural production.’ Although scholars have begun to enumerate the implications of this hypothesis (Grabher & van Tujil 2020; Kenney & Zysman 2020; Grabher 2020), its consequences for how platform power can be understood remain understudied. It is not an exaggeration that technology giants are not simply new types of intermediaries that facilitate ‘diverse interests and objectives of firms and extra-firm actors’ (Coe & Yeung 2015: 50). Instead, platform-powered lead firms may be able to use their dominant economic position to consolidate their power in ways that go beyond the conventional scope of inter-firm relations.

In summary, the study of power in AI’s geographies of production remains in its infancy. Yet, understanding power’s changing nature is a precondition for regulating the operations of industry-dominating lead firms. Building on the combined insights of the previous sections, it is time to introduce the notion of neural production networks as the original theoretical framework that is at the heart of this thesis.

### 2.3 Neural Production Networks

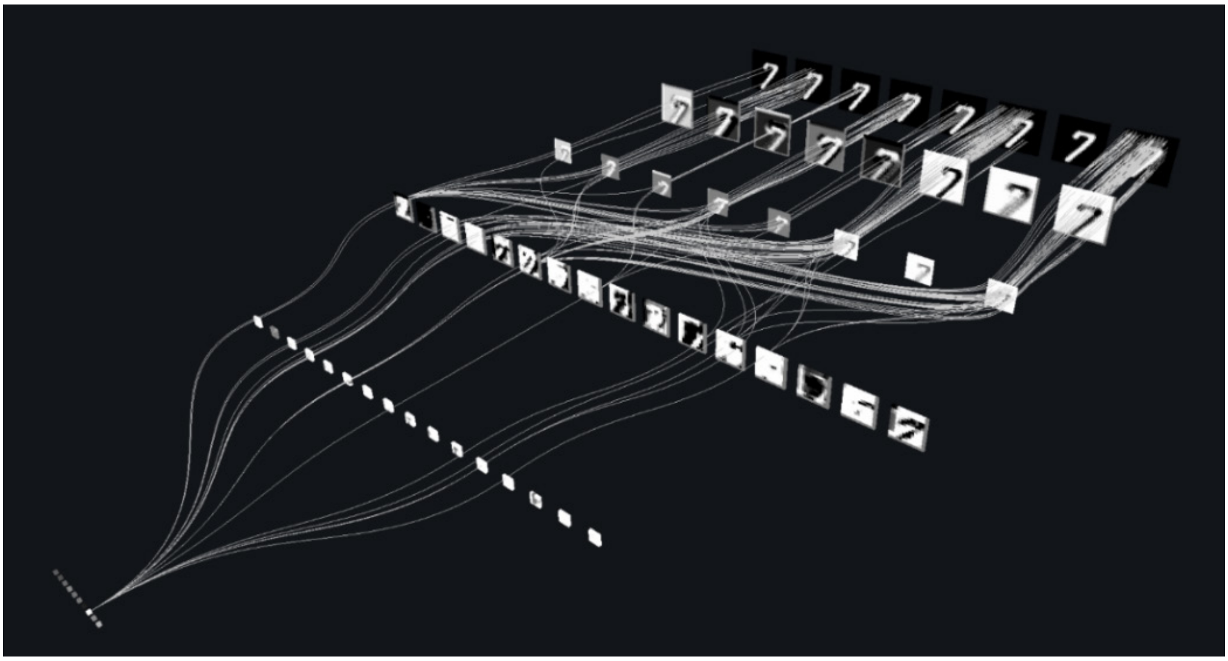
The most important definitional criterium of global production networks is that they are *global*: they may operate at a global scale. By contrast, the most important definitional criterium of neural production networks is that they are *neural*: they are powered by artificial neural networks. Paraphrasing Werner (2019: 948), I define neural production networks as *geographically dispersed but computationally enveloped production arrangements powered by artificial neural networks*.

In contrast to global production networks, neural production networks signal a shift from focusing on global network configurations and uneven inter-firm relations to problematising the integration of artificial neural networks into the geographies of production. In developing a theoretical framework to grasp this shift in focus, I am inspired by Dyer-Whiteford's et al. (2019: 62) argument that AI has now become 'ubiquitous, distributed by a network infrastructure, just as electricity and internet access.' What is noteworthy about this argument is that it brings together two aspects that seem to be at odds with each other. On the one hand, there is the mundane nature of infrastructure: a barely recognisable feature of everyday life. On the other hand, there is the future-oriented, out-of-reach nature of AI: a complex set of technologies that are not only imaginary or speculative objects but already serve as the computational backbone of actually-existing modes of industrial production. It is precisely this interplay that the notion of neural production networks captures and conceptualises.

But why are artificial neural networks geographically relevant in the first place? How can their abstract computational features be related to less abstract social, cultural, and economic relations? Although research on the bourgeoning spatialities of artificial neural networks across different domains of life remains in its infancy, we can already see highly innovative theoretical approaches in fields such as political geography. For instance, in her analysis of the role of deep learning technologies in contemporary border control systems, Amoore (2021: 3) disentangles the peculiar links between the mathematical understanding of depth and a 'broader political imaginary of a deep reach into diverse sources of available input data, and a mapping of non-linear relations in line with 'output' policy objectives.' Amoore critically engages with the work of influential computer scientists such as Yoshua Bengio, Yann LeCun, and Geoffrey Hinton to argue that deep learning is not simply a novel instrument of border control, but crucially a 'means of reordering *what the border is*, what it could be, and how it imagines and bounds political community' (Amoore 2021: 7). For unfolding the idea of neural production networks, Amoore's (2021: 3) description of the spatiality of artificial neural networks is particularly relevant – a structure that is visualised by Figure 2.2:

> Each of the layers in a neural net computes one partial function of a much larger whole of the representation of the input data – for example, a layer in an image recognition might represent one small edge in a group of pixels in the image. Here, depth means a capacity to abstract and to represent the relationships in high-dimensional data. […] Put simply, the imagination of depth in computer science is not only a means of model building, but a form of world-making.

Figure 2.2: Visualisation of an artificial neural network. Source: Broad 2019.



Consequently, rather than foregrounding the *territorial expansion* of the space within which production takes place (i.e., its globe-spanning scale), the notion of neural production aims to contribute to our understanding of the *computational envelopment* of the space within which production takes place (i.e., its infrastructural dependency on computational processes). This is not to imply that neural production networks replace global production networks, that the latter are no longer relevant, or that some of them do not require computational processes. Neither do I wish to suggest that those types of production networks are mutually exclusive. Instead, my rationale for coining a new term for the purposes of this thesis is to take seriously the conceptual strengths of global production networks while also accounting for the particularities of AI's geographies of production as objects of study (Table 2.2). In other words, I deploy this idea in order to structure, sharpen, and connect the key arguments of this thesis.

Table 2.2: Global production networks vs. neural production networks.

| | **Global production networks** | **Neural production networks** |
|---|---|---|
| **Disciplinary background** | Economic geography<br>Economic sociology | Platform and infrastructure studies<br>Economic geography |
| **Analytical spotlight** | Global network configurations and uneven inter-firm relations<br><br>Territorial expansion of the space within which production takes place<br><br>Regulatory and cultural implications of border-crossing production networks | Integration of artificial neural networks into geographies of production<br><br>Computational envelopment of the space within which production takes place<br><br>Regulatory and cultural implications of AI's proliferation and sophistication |
| **Theoretical influences** | Relational economic geography<br>Global value/commodity chains<br>Actor-network theory (ANT)<br>Varieties of capitalism | Infrastructural thinking<br>Global production networks<br>Critical political economy<br>Platform capitalism |

Given that the later chapters will explore and elucidate the different facets of neural production networks, it is pivotal to use this section to clarify my understanding of computational envelopment as their defining feature. In the context of video game interfaces, Ash (2015: 3) defines envelopes as 'localized foldings of space-time that work to shape human capacities to sense space and time for the explicit purpose of creating economic value.' My interpretation is slightly different. I use the notion of envelopment to emphasise that, because they are computing systems, artificial neural networks, by definition, require computational processes. This means that there are no neural production networks without computation. While global production networks can, in theory, exist without computation, neural production networks cannot. Neural production networks, however, are neither free-floating and beyond the constraints of geography, nor can they be understood as fixed infrastructures that are simply rooted in particular parts of the globe. Instead, like other forms of networked production, they are 'constituted through a variety of circuits and (non-linear) flows linking a variety of sites and spaces' (Hudson 2008: 422). But beyond operating at the conjuncture of multiple 'foldings of space-time' (Ash 2015: 3), the geographies of neural production networks remain enveloped by a constitutive outside: the necessity of computation.

This envelopment entails a number of implications that go beyond the question of how firms create economic value as part of border-crossing production arrangements. Recapitulating an argument that I made in relation to Project Maven and *The Gospel of Wealth*, I am not only interested in addressing issues of industry consolidation, but also in examining the regulatory dimensions and cultural ramifications of AI's geographies of production. Herein lies the strength of the global production networks framework as an inspiration for my original idea of neural production networks: its holistic analytical scope. This wide-ranging focus is exemplified in the conceptual groundwork undertaken by Hudson (2008: 438), who argues in favour of pursuing a cultural political economy approach to the analysis of global production networks:

> GPNs [i.e., global production networks] could become an extremely powerful way of representing 'the economy' in its essential complexity (that is with emergent effects, an entity that is not just complicated but complex), of representing the richness of the economy in terms of the links between the affective, cognitive and material, between circuits of value, meaning and matter, between the moments of production, exchange and consumption and between political economies grounded in different concepts of value and processes of valuation.

According to Hudson (2008: 422), three analytical registers can be identified in this context: first, a political-economic register ('encompassing labour processes as well as processes of value creation, exchange and realization'); second, a semiotic register ('relating to flows of knowledge and information and to the culturally endowed meanings that things come to acquire); third, a material register ('conceptualising the economy in terms of materials transformations – biological, chemical and physical – as well as flows of energy, matter and materials'). For Hudson (2008), it is important to probe the relations between those co-constitutive registers and steer clear of thinking about production processes as unidirectional or linear dynamics.

Table 2.3: Territorial expansion vs. computational envelopment.

| | **Territorial expansion of space (global production networks)** | **Computational envelopment of space (neural production networks)** |
|---|---|---|
| **Industry concentration** | Strategic use of uneven geographies to cut and supress costs as the 'elemental rationale' of lead firms (Peck 2017) | Morphing of 'infrastructure features' and 'platform features' of lead firms (Plantin et al. 2018; Grabher 2020) |
| **Regulatory dimensions** | Mismatch between scales of regulation (e.g., national/local) and scales of production (e.g., transnational) | Emergence of new regulatory tools in response to industry consolidation (e.g., the EU's Digital Markets Act) |
| **Cultural ramifications** | Commodity fetishism due to relative distance between sites of production and sites of consumption (Castree 2001) | Dissolution of commodity form, proliferation of 'X-as-a-service' business models (Sadowski 2019) |

Inspired by Hudson's (2008) imaginative synthesis, I theorise the implications of computational envelopment as the defining feature of neural production networks in relation to three registers: industry concentration, regulatory dimensions, and cultural ramifications. To justify the necessity of coining a new term to coalesce those registers (rather than resorting to an already-established concept), a top-level contrast between *territorial expansion* and *computational envelopment* is illustrative (Table 2.3). The purpose of this exercise is not to offer a comprehensive discussion of all aspects and facets of those definitional features, but to encapsulate the paradigmatic discontinuities of neural production networks as a powerful conceptual instrument.

First, the underlying theoretical assumptions of what counts and what does not count as a networked production arrangement determine how geographies of production can be related to the analysis of industry concentration. If the territorial expansion of the space within which production takes place is used as the key frame of reference to study a specific industry, research questions may turn around how firms exploit geographical inequalities to expand their market power. Take the case of outsourcing. As Peck (2017: 16) writes, 'while the outsourcing industry likes to convey the impression […] that it has matured away from the cheap-labour model, the hard-to-escape reality is that cost suppression remains an existential condition.' Lead firms may achieve cost suppression by using labour pools in low-wage locations to produce goods and conduct services at a cheaper cost. Uncovering those links between labour arbitrage and patterns of industry concentration remains highly relevant, particularly in an era of digital labour platforms as lean optimisers of cost suppression (Howson et al. 2021).

By contrast, if we prioritise the computational envelopment of the space within which production takes place as the theoretical entry point, the ways in which lead firms consolidate their industry dominance do not hinge on their *moments of grounding*, what Hess (2018) calls 'territorial embeddedness.' Instead, the spotlight is on the facilitation of infrastructural dependencies by firms that own the computational resources that are needed to deploy artificial neural networks. Although this practice may require the exploitation of territorial unevenness (e.g., the strategic location of data centres), neural production networks seem far less spatially rooted than networks that gravitate around the creation, circulation, and consumption of physical commodities. Here, I am inspired by Bratton's (2016: xvii) argument that 'maps of horizontal global space cannot account for all the overlapping layers that create a thickened vertical jurisdictional complexity, or how we already use them to design and govern our worlds.' This is not to imply that value, embeddedness, and power are no longer relevant categories to analyse production networks but rather that they may matter in new ways. In a nutshell, the proliferation of the platform business model (van Dijck et al. 2018) and the infrastructural presence of platform companies (Plantin et al. 2016) call into question conventional ways to relate patterns of industry concentration to specific production geographies.

Second, a shift in focus from territorial expansion to computational envelopment is also reasonable when it comes to questions of regulation. This is a crucial area, as it points to the bilateral relationship between conceptual abstraction on the one hand and regulatory intervention on the other hand. Far from being restricted to scholarly debates, notions and ideas emanating from economic geography and political economy have had direct impacts on regulatory frameworks around the world. Most notably, the idea of 'global value chains' fed into numerous corporate disclosure and due diligence laws on the social and environmental impacts of economic production (Salminen & Rajavuori 2019). Crucially, such laws may apply to domestic lead firms even if their production activity itself takes place outside a specific jurisdiction. Theorising that there *are* global production networks (or global value chains, for that matter) has been a prerequisite not only for grasping their problematic outcomes but also for appraising the efficacy of regulatory proposals, which aim to mitigate those outcomes.

Along similar lines, theorising that there *are* neural production networks will be a prerequisite not only for studying how lead firms consolidate their dominance but also for appraising the efficacy of regulatory frameworks, which aim to curb this dominance.

This task, however, goes beyond emphasising the mismatch between the national and local scale of regulatory frameworks and the globe-spanning scale of certain production activities, which continues to complicate the implementation of effective regulatory regimes across the globe (Salminen & Rajavuori 2019). Beyond such considerations of scalar mismatches, the notion of neural production networks also affords the possibility to theorise the complex positioning of states as co-creators of AI's spaces of production. The European Commission, for example, is a crucial geopolitical actor in the field of digital policy. But its sweeping policy proposals have not yet been problematised in relation to empirical findings on how technology giants act as orchestrators of neural production networks. It is time to change this status quo.

Third, a reorientation from expanded territories to enveloped computation as an analytical frame of production geographies also raises a number of intriguing questions regarding their cultural ramifications. If it is not primarily the relative distance between the sites of production and sites of consumption that constitutes the cultural impacts of production geographies, then what else does? Is a focus on commodities still justified? How do artificial neural networks reshape the cultural perception of a blurring boundary between reality and simulation on a more general level? Before catalysing those points into an answerable research question, it is worth emphasising the commodity-centric nature of spatial thought in this context. As Castree (2001: 1520) notes, in their efforts to rejuvenate the Marxian critique of commodity fetishism[12], scholars have followed an 'urge to ground commodities in a specific site and a particular constituency: namely, the site of production and the constituency of spatially dispersed labour.' However, an issue

[12] While there are different understandings, commodity fetishism can be seen as 'the mistaken view that the value of a commodity is intrinsic and the corresponding failure to appreciate the investment of labour that went into its production' (Buchanan 2018).

with this tendency is that 'the trope of "unveiling" not only underplays the positivities of consumption but – as Jean Baudrillard showed – also fails to take seriously the semiology of commodity surfaces' (Castree 2001: 1520). In other words, a narrow focus on the territorial journeys of commodities as a set of cultural phenomena risks legitimising the objects of their critique: the ontological stability of commodities and their perceived centrality in determining the cultural impacts of production networks.

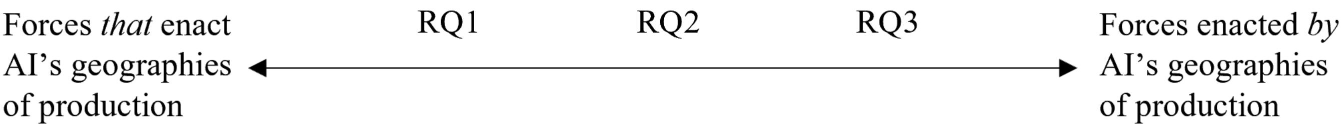
In recent years, a range of scholars has argued that we witness a dissolution of the commodity form in certain industries – a development that coincides with changing cultural practices. Bernevega and Gekker (2021: 53) argue that modern video games, such as the highly successful free-to-play online game Fortnite, are 'neither produced nor monetized as a commodity.' Instead, game studios aim to extract value from micro-transactions, whereby players buy virtual assets to equip their avatars with fancy clothes or eccentric hairstyles. Those virtual assets, in turn, are computationally reproducible and thus extremely profitable sources of revenue for game producers. This argument, however, is by no means restricted to games. As Sadowski (2019: 6) states, 'the surge of companies that describe themselves as "Uber-for-X" or "X-as-a-service" – whether start-ups in search of funding or incumbents looking to rebrand – are creating rentier relations by another name.' It is precisely this strand of critical work that inspired the theoretical framework of neural production networks.

In the next section, each of these three analytical registers will serve as a basis for posing an answerable research question. Together, those analytical registers and research questions congeal into a holistic yet coherent strategy to deploy the idea of neural production networks for addressing the research problem of this thesis. I argue that this approach is flexible enough to cast a wide net of empirical phenomena, but also narrow enough to avoid the pitfalls of conceptual blurriness.

**2.4 Research Questions**

The central research question of this thesis is: **What forces enact, and are enacted by, AI's geographies of production?** The thesis is designed to answer this central research question by (1) analysing the role of lead firms within AI's geographies of production; (2) juxtaposing these empirical findings with contemporary regulatory developments on AI governance in the European Union; and (3) problematising the cultural ramifications of AI's geographies of production by dissecting the logics of deep generative models as burgeoning forces of media production. As such, the empirical chapters of this thesis are framed as answers to the following three research questions:

1. How do technology giants act as lead firms within AI's geographies of production?
2. How does state action shape, and is shaped by, AI's geographies of production?
3. What do AI's geographies of production imply for platformised cultural production?




I operationalise my object of study, AI's geographies of production, by using the conceptual framework of neural production networks to answer my research questions. The logical flow of the research questions can be visualised as a spectrum that ranges from the forces *that* enact AI's geographies production to the forces enacted *by* AI's geographies of production. In answering the questions, the empirical chapters build on each other in order to tell a coherent story that is bigger than the sum of its parts. With that in mind, the next section provides a comprehensive overview of the epistemological assumptions and methodological approaches of this thesis, elaborating on its case study-based research strategy and data analysis techniques.

# Chapter 3: Methodology and Research Strategy

## 3.1 Introduction: The Force of COVID-19

> Do you know what Life is to me? A monster of energy . . . that does not expend itself but only transforms itself. . . . [A] play of forces and waves of forces, at the same time one and many . . . a sea of forces flowing and rushing together, eternally changing (Nietzsche 1967 [1901], cited in Bennett 2010: 540).

In December 2019, a few weeks before the first positive COVID-19 case was detected in Europe, I attended a workshop at Cardiff University. Hosted by the Centre of Law and Society and the Environmental Justice Research Unit, the workshop's theme was "Imagining the Eco-Social: New Materialist Reflections for the Anthropocene." It was co-led by two political theorists: Jane Bennett and William Connolly. Retrospectively, I see the workshop as a major turning point in my time as a doctoral researcher. For two reasons, the workshop had a significant impact on this thesis. Neither the relevance nor the existence of those two reasons could have been foreseen in 2019 – a fact that makes it even more important to reflect upon them at the start of this chapter.

First, it sparked my interest in what it means (and what is required) to 'produce guides to action appropriate to a world of vital, crosscutting forces' (Bennett 2010: 38). Before attending the workshop, I had not explicitly considered key ideas and strands of social theory that were discussed in Cardiff, such as Bennett's notion of vibrant matter. But despite this lack of theoretical background knowledge, I did not feel out of place there. On the contrary, when I presented my initial research plan for this thesis, I was fortunate to receive valuable suggestions to consider how the agential capacities of non-human forces present themselves in my objects of study. However, there was just one problem that turned out to stand in the way of implementing those fruitful conceptual suggestions: I was not able to execute my initial research plan.

This leads to the second reason as to why I see the workshop as a turning point: it was the last time I attended an event before the COVID-19 pandemic hit the world with full force – and the last time I presented fieldwork plans that never materialised. Having to change one's research plan cannot be seen as a problem in the grand scheme of misery and human tragedy caused by the COVID-19 pandemic. But even if it is not a problem in the literal sense, it still needs to be addressed in this chapter.

The plan that I presented at the Cardiff workshop involved several months of ethnographic fieldwork in Bangalore, India. Substantively, its focus was on the first historical parallel between Project Maven and the *Gospel of Wealth* that I discussed in Chapter 1: the discursive eradication of collective labour from societal discourses about the driving forces of technological progress. At that time, many debates around artificial intelligence not only ignored the crucial role of outsourced labour (e.g., data annotation) but also suggested that such menial tasks will become obsolete thanks to technological advances. *The opposite is true*, I hypothesised, and designed a research project with the aim to investigate the division of labour between those who design the artificial neural networks of AI systems (e.g., Project Maven's computer vision models) and those who manually annotate their training datasets. I will return to this initial research plan later on. But for now, it is sufficient to note that it existed.

In essence, the research strategy of this thesis is the result of a synthesis of those two reasons. On the one hand, it represents a feasible way to answer the central research question, whose formulation was inspired by the fruitful exchange in Cardiff: *What forces enact, and are enacted by, AI's geographies of production?* On the other hand, it is also a pragmatic alternative to plans that were cancelled due to the COVID-19 pandemic. Those two reasons are inextricably related. Their synthesis lies in the fact that answering the central research question *necessitates* a shift in focus away from the study of outsourced labour through ethnographic fieldwork. In other words, I aim to demonstrate in this chapter that the above-mentioned turning point did not undermine but rather reconfigure the research strategy of this thesis: it was an impetus to craft a more coherent and comprehensive methodological approach than the one I initially developed. This turning point expanded the scope of this thesis.

To develop this argument, this chapter proceeds as follows. The first step is to show that the key advantage of formulating the research question in this way lies in the possibility to decompose it into answerable elements that co-constitute the phenomenon in question (i.e., rather than representing analytical silos). The second step is to justify my rationale for analysing paradigmatic cases as 'carefully selected examples extracted from phenomena' (Pavlich 2010: 645). In this regard, I combine a subjective account of the research process with theoretical considerations. After explaining how I triangulated document analysis, event observations, and informal conversations, the chapter outlines the ethical considerations and limitations of this methodological approach.

### 3.2 Enacting Infrastructural Geographies

> Method is not […] a more or less successful set of procedures for reporting on a given reality. Rather it is performative. It helps to produce realities. […] The consequence is that method is not, and could never be, innocent or purely technical. If it is a set of moralisms, then these are not warranted by a reality that is fixed and given, for method does not 'report' on something that is already there. Instead, in one way or another, it makes things more or less different. The issue becomes how to make things different, and what to make (Law 2004: 143).

It is important to start this section by laying bare my belief that methods do not only describe their objects of study but are involved in producing them. One of the most lucid expressions of this belief can be found in Law's (2004) *After Method: Mess in Social Science Research*. Concepts, ideas, abstractions, and frameworks cannot be separated from the methods, instruments, and analytical approaches that probe them by studying a given phenomenon. A research strategy is thus not an objective or value-free *representation* of reality, but rather a subjective and value-laden *enactment* of reality. This, however, does not mean that it cannot be rigorous and reflexive. The purpose of this chapter is to offer a comprehensive account of how my research strategy 'make[s] things different' (Law 2004: 143). But to do so, it is indispensable to take a step back and answer the question: What things are being made different?

Because infrastructures are 'are not, in any positivist sense, simply "out there"' (Larkin 2013: 330), infrastructural geographies are not simply "out there" either. The 'categorizing moment' (Larkin 2013: 330) of this thesis, therefore, lies in enacting the infrastructural geography of artificial intelligence by focusing on three of its disparate elements: patterns of industry concentration; the role of state actors and state action; and the cultural ramifications of AI-generated media objects. To conceptually underpin this enactment, Chapter 2 has introduced an original heuristic device that problematises how those three constitutive elements comprise a larger phenomenon that is bigger than the sum of its parts: the notion of neural production networks.

However, I do not wish to argue that this research strategy is the one-and-only way (or the "right" way) to go about enacting AI's infrastructural geographies, or that no other constitutive elements are conceivable. For this reason, I intentionally use the singular form "infrastructural geography" instead of the plural form in the subtitle of this thesis[13]. As I stressed when discussing the term infrastructure, there is a wide range of *other* elements that may be problematised in light of their "infrastructural" properties

[13] That said, it could be argued that the definite article "the" is inappropriate and that it should be replaced with "an." However, I hold that it is possible to appreciate the existence and possibilities of enacting AI's infrastructural geographies in other ways without having to water down or relativise the most important contributions of this thesis.

for artificial intelligence (e.g., elemental processes that underlie computational devices). Yet it is not only the complexity of such other elements with respect to AI but also the ontological features of infrastructures more generally that complicate the development of a rigorous and coherent research strategy. As Larkin (2013: 239) reminds us, the peculiar ontology of infrastructures 'lies in the facts that they are things and also the relation between things.' Engaging with this duality is certainly a laborious and *messy* undertaking – to use Law's (2004) terminology – but it also opens up new grounds for otherwise-unelaborated arguments, abstractions, and juxtapositions.

### 3.3 The Paradigmatic Case Study

Methodologically, my research strategy to craft those arguments is informed by the idea of 'paradigmatic cases' (Pavlich 2010; Flyvbjerg 2011) as a distinct and fruitful way to select and analyse case studies. On the most general level, a case study 'investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used' (Yin 1984: 23). In challenging the conventional misunderstanding that the insights of case studies are not relevant beyond the individual case units (i.e., the objects of analysis), Flyvbjerg (2006: 229) asserts that the 'generalisability of case studies can be increased by the strategic selection of cases.'

With that in mind, the relevant methodological literature points to a variety of case selection strategies, ranging from random sampling to extreme cases (Table 3.1). An important aspect in this context is the distinction between random sampling and purposive sampling. While the random selection of cases is especially pertinent for quantitative methods that aim for representative samples (e.g., surveys), the purposive selection of cases operates 'on the basis of expectations about their information content' (Flyvbjerg 2011: 307). Therefore, this thesis uses a purposive case selection strategy.

Because the purpose of a paradigmatic case is to ‘develop a metaphor or establish a school for the domain that the case concerns’ (Flyvbjerg 2011: 307), it is the most suitable case selection strategy to answer the central research question. Having laid out this decision, what are paradigmatic cases and why are they useful?

Table 3.1: Common case selection strategies. Source: Flyvbjerg 2011.

| *Type of Selection* | *Purpose* |
|---|---|
| A. Random selection | To avoid systematic biases in the sample. The sample's size is decisive for generalization. |
| 1. Random sample | To achieve a representative sample that allows for generalization for the entire population. |
| 2. Stratified sample | To generalize for specially selected subgroups within the population. |
| B. Information-oriented selection | To maximize the utility of information from small samples and single cases. Cases are selected on the basis of expectations about their information content. |
| 1. Extreme/deviant cases | To obtain information on unusual cases, which can be especially problematic or especially good in a more closely defined sense. To understand the limits of existing theories and to develop new concepts, variables, and theories that are able to account for deviant cases. |
| 2. Maximum variation cases | To obtain information about the significance of various circumstances for case process and outcome; e.g., three to four cases that are very different on one dimension: size, form of organization, location, budget, etc. |
| 3. Critical cases | To achieve information that permits logical deductions of the type, "If this is (not) valid for this case, then it applies to all (no) cases." |
| 4. Paradigmatic cases | To develop a metaphor or establish a school for the domain that the case concerns. |

Linking the etymologies of the two terms *paradigm* (i.e., “to show side by side”) and *case* (i.e., “state of affairs” or “event”), Pavlich (2010: 645) defines a paradigmatic case as a ‘singular event that involves placing an exemplar alongside a phenomenon; by virtue of so placing, it shows or reveals key elements of that phenomenon.’ Flyvbjerg (2011: 308) uses a slightly different definition, interpreting paradigmatic cases as ‘cases that highlight more general characteristics of the societies in question.’ The issue with this second definition is that it – in one way or another – may apply to all case selection strategies displayed in Figure 3.1. A randomised survey sample, for instance, might also aim to unearth such more general characteristics. Consequently, it is worth following

Pavlich's (2010) definition for now by considering it in relation to a clear example of a paradigmatic case. While there are key definitional differences, both authors refer to Foucault's panopticon in this context. As Pavlich (2010: 647) explains:

> Even though Bentham's all-seeing (panopticon) blueprint was never actually built […], Foucault considers it a paradigmatic case for his analysis of a specifically modern form of power: discipline. […] Foucault treats the panopticon as a paradigmatic case that elucidates new disciplinary technologies that include hierarchical observation, indiscernible surveillance, normalizing judgment, and so on. This paradigmatic case, by virtue of its being placed side by side with the phenomenon of modern power, is at once an element thereof and yet serves to make such power intelligible. […] [B]y opening out a conceptual space that situates a paradigmatic case (the panopticon) alongside the phenomenon (modern disciplinary power), and charting the exemplary elements of that case (e.g., invisible, continuous surveillance that encourages individuals to regulate themselves), Foucault renders the phenomenon of modern discipline intelligible.

In other words, the purpose of selecting a paradigmatic case is to bring to light the fundamental characteristics of a phenomenon by placing it alongside (rather than on top of, or subordinated to), the phenomenon. There are, of course, many other research strategies and methodological techniques that rely on similar approaches, without using case study terminology. It is questionable whether Foucault would have agreed that his mode of reasoning should be classified into systematic methodological categories in the first place. Still, the example of Foucault's panopticon is illuminating insofar as it lays bare the mechanics of paradigmatic case studies as consisting of three elements that are relevant for crafting a coherent and rigorous case study design:

(1) The phenomenon in question (e.g., modern disciplinary power);

(2) the paradigmatic case placed alongside it (e.g., panopticon);

(3) what this case elucidates (e.g., indiscernible surveillance).

Against that backdrop, it is pivotal to clarify what those three dimensions denote with respect to my central research problem: that is, *uncovering the forces that enact, and are enacted by, AI's production geographies*. I purposively phrased this research problem (and the research questions that derive from it) in a way that avoids implying a deterministic or unidirectional cause-and-effect relationship. I take my cue for this from Hepworth (2014: 1121-1122), who fruitfully combines a document analysis with the epistemological standpoint of 'enacting logistical geographies':

> I use the term enactment to describe how particular realities are brought into being. […] In considering how logistics enacts new geographies, I analyse how the technological and physical interventions into Port Botany are guided by and represented in government reports, policy documents, and even promotional materials. Each of these documents suggests particular modalities of management and regulation of urban space and logistical sites. […] These interventions into lived space are considered relative to a single city – Sydney – […] in order to highlight the frictions and partial optimisations that result when the archetypal, imagined space–times of logistics encounter a particular city.

While Hepworth (2014) does not use paradigmatic case study terminology, her approach still resonates with Pavlich's (2010) methodological considerations. It is worth elaborating on this resonance. Hepworth's (2014) phenomenon in question is logistics. Her paradigmatic case is Port Botany, Sydney. The aim is to render logistics intelligible by charting the exemplary elements of the case. The case elucidates frictions between the imagined geographies of logistics and lived urban space.

My phenomenon in question is artificial intelligence. Specifically, as outlined in Chapter 1, I focus on forms of AI that are underpinned by artificial neural networks. My paradigmatic case to render artificial intelligence intelligible, however, is not a *singular* unit, as in the case of the panopticon or Port Botany. Instead, the nature of my research question requires an analysis of *multiple* cases that entail paradigmatic properties vis-à-vis the phenomenon in question. To stay with the metaphor of forces, I argue that there is not just one force that enacts AI's geographies of production, like the infrastructural role of technology giants. Rather, problematising the tensions between multiple forces is pivotal in light of my research question. Crucially, the paradigmatic discontinuities of neural production networks go beyond questions of industry concentration (Table 2.3). Therefore, it is necessary that the research design conforms to this expanded focus of the conceptual framework by considering not only matters of industry concentration, but also AI's regulatory dimensions and cultural ramifications.

That, however, warrants the key question: How to select cases with paradigmatic properties without risking superficiality or losing coherence? As Flyvbjerg (2011: 308) puts it, 'the paradigmatic case transcends any sort of rule-based criteria. No standard exists for the paradigmatic case because it sets the standard.' Pavlich (2010: 647) adds that 'the effectiveness of the [paradigmatic case study] approach depends on the careful selection and extrapolation of examples.' Both assessments are highly relevant and can be directly linked to Larkin's (2013) considerations regarding the role of categorising moments in the study of infrastructures. The lack of rule-based criteria reinforces the need for transparency and reflexivity regarding the case selection strategy.

As a roadmap for the remainder of this chapter, Table 3.2 offers an overview of how the case study choices of my research design relate to the research questions. The number of units per case study ranges from one (Chapter 5) to three (Chapter 4). Before outlining data collection and analysis methods in detail, the next three sections elaborate on the methodological reasons for my case study choices. This is a fruitful endeavour insofar as it provides the opportunity to not only showcase a subjective reflection on the research process, but also to situate the final research design in relation to relevant work that also resorts to the analysis of paradigmatic cases.

Table 3.2: Overview of paradigmatic case study research strategy.

| | **Chapter 4: The Rise of Neural Production Networks** | **Chapter 5: The Regulation of Neural Production Networks** | **Chapter 6: Neural Production Networks and Platformised Cultural Production** |
|---|---|---|---|
| **Research question** | How do technology giants act as lead firms within AI's geographies of production? | How does state action shape, and is shaped by, AI's geographies of production? | What do AI's geographies of production imply for platformised cultural production? |
| **Focus of case study analysis** | Dominant role of technology giants as lead firms in neural production networks | Strategies and constraints of state actors to regulate neural production networks | Implications of neural production networks for platformised cultural production |
| **Units of case study analysis** | Amazon, Google, Microsoft | The EU's Digital Single Market Strategy | Unity Technologies, Epic Games |
| **Corpus of case study analysis** | Industry documents (e.g., technical papers, marketing materials, media articles), event observations, informal conversations | Policy documents (e.g., legislation, reports, whitepapers, speeches), event observations, informal conversations, media articles | Industry documents (e.g., technical papers, marketing materials, media articles), event observations, informal conversations |
| **Why is this a paradigmatic case study?** | Market capitalisation, dominance of global infrastructure-as-a-service market, highly visible AI research (e.g., AlphaGo) | World's first comprehensive approach to AI regulation and *ex-ante* digital policy: spill-over effects to other countries | Prototypes of AI-generated media and new revenue streams due to confluence of game engines and deep generative models |

### *3.3.1 The Paradigmatic Lead Firm*

The purpose of analysing paradigmatic lead firms in neural production networks is to elucidate the mechanisms of how their 'centralisation of control' (Crawford 2021: 223) renders intelligible my phenomenon in question: artificial intelligence.

But what is a lead firm, and why is it an important object of study to address my research problem? According to Yeung and Coe (2015: 39), a lead firm is defined by its 'capacity to coordinate and control directly its production network – be it in the role of a buyer, producer, coordinator, controller, or market-maker, or a composite of one or more of these roles.' They refer to a number of prominent examples that exhibit this capacity, such as Vodafone in telecommunications, Wal-Mart and Tesco in retail, UPS in logistics, and Nestlé in nutrition. The methodological advantage of centring the lead firm in the study of production geographies is that this focus enables the identification of industry structures without having to map the *entire* industry. For instance, if the aim is to understand power asymmetries in global food chains, a spotlight on the operations of Nestlé may help ensure methodological clarity.[14]

Along similar lines, for making sense of how patterns of industry concentration manifest in AI's geographies of production, the identification of lead firms is a useful methodological starting point. Given that Chapter 4 provides more detail about why I selected Amazon, Google, and Microsoft (instead of Apple or Alibaba), for now, it is pivotal to stress the two major benefits of studying lead firms as paradigmatic cases. First, this way of identifying the objects of analysis augments what Flyvbjerg (2011: 307) calls the 'validity claims' that can be placed on the study 'and the status these claims obtain in dialogue with other validity claims in the discourse to which the study

[14] For Coe and Yeung (2015: 40), each lead firm constitutes its own production network, which makes it easier to identify similarities and differences between those networks.

is a contribution' The notion of the lead firm represents an already-established research paradigm that is a reference point for scholarly debates. Instead of having to use 'one's intuitions' (Flyvbjerg 2011: 308) to justify the selection of paradigmatic cases, metrics such as market capitalisation and market shares can be consulted.

Second, the features of technology giants as units of case study analysis can be leveraged to underpin the necessity of new conceptual frameworks and new paradigms. For example, Helmond's (2015: 1) notion of platformization, defined as the 'rise of the platform as the dominant infrastructural and economic model of the social web and its consequences', is derived from the analysis of one paradigmatic case study: Facebook. Subsequently, it was used to study a range of phenomena, from influencers to protests. As Pavlich (2010: 646) notes, 'the paradigmatic case simultaneously, if paradoxically, emerges from, and constitutes, the set to which it belongs.' Pragmatically speaking, the necessity of developing the original framework of neural production networks becomes evident through elucidating its empirical manifestations.

### *3.3.2 The Paradigmatic State Action*

To explain why I see the EU's Digital Single Market Strategy as a paradigmatic case to understand the positioning of state actors in neural production networks (Table 3.2), it is worth recapitulating the subjective decision-making process that preceded this selection. Initially, I planned a comparative case study analysis of how two state actors approach the topic of AI regulation: the European Commission and Canada.

In preparation for this work, I secured a one-year affiliation with a university in Montréal, gained visiting researcher status, and began to map the Canadian ecosystem of relevant AI policy experts to plan semi-structured interviews with them. But when I began my time as a visiting researcher in Montréal in August 2021, two factors led me to change the direction of my research plans toward what I now see as a more coherent and theoretically relevant approach. First, at that time, COVID-19 cases rapidly surged across Canada, especially in the province of Québec, resulting in a situation where face-to-face interviews were not feasible due to a general sense of caution as well as official social distancing policies. In my experience, it is very difficult (not to say impossible) to establish a sufficient level of trust in remote expert interviews to be able to generate rich empirical data – that is*, data that cannot be gathered in any other way*, such as through analysing publicly available policy documents and reports. While the later sections of this chapter engage with practical matters of data gathering and analysis in more detail, it is important to note that I decided to exclude formal interviews as a methodological approach due to the unpredictability of the COVID-19 pandemic.

Second, even though there are intriguing similarities between how the European and the Canadian regulatory regimes approach AI policy (e.g., by considering AI's risks and harms), I assessed my contextual background knowledge of institutional processes in Canada as being too limited to allow for a methodologically sound comparative case study. Additionally, given that one case unit is a supranational state actor and the other case unit a national state actor, I decided to adjust my research plans and focus only on one case: the EU's Digital Single Market Strategy. This, however, does not mean that my work in preparing the comparative case study did not generate important insights that could inform this reconfigured focus. On the contrary, my initial preparatory work in analysing Canadian AI policy reports can be repurposed to elucidate the paradigmatic properties of my final case study choice. For example, as the authors of a report released by the Law Commission of Ontario (2021: 3-4) note:

> Globally, the European Commission was the first regulatory body to attempt a comprehensive legislation to address AI. Others will follow suit soon, selecting regulatory approaches that are best suited for their specific context. […] [T]he Commission created a Digital Single Market, which aims to harmonize digital

> services in its 27 countries, to allow for interoperability of data and digital innovation. […] [I]n the European Union, the initiative for regulation lies with the European Commission. It has chosen to propose new rules within the proposed regulation of April 21, 2021, and not just interpret existing ones. It has also decided to have a broad regulatory approach and not a sectoral approach[15] […].

The perception that the European Commission's policy proposals with respect to AI regulation have a 'metaphorical and prototypical value' (Flyvbjerg 2011: 307) for debates in Canada was echoed by informal conversations with researchers and experts. As the above-mentioned quote demonstrates using the example of Canada, the point is not that the EU's regulatory approach is simply being replicated by other jurisdictions, but rather that it serves as a *prototypical* reference point in legal and policy debates. For example, a pattern that I observed was that the contours of specific proposals, such as the Government of Canada's Directive on Automated Decision-Making (ADM), were clarified by reports and commentators *vis-à-vis* the European Commission's proposals. In methodological terms, 'paradigmatic analysis involves a relation between particular cases to other particular cases' (Pavlich 2010: 646).

In summary, in contrast to paradigmatic lead firms, the paradigmatic state action is not defined by its 'capacity to coordinate and control' (Yeung & Coe 2015: 39) neural production networks but rather by its capacity to enable and constrain the operations of said lead firms. But as I argue in Chapter 5, the empirical manifestations of this capacity require an expansion of focus beyond exclusively considering state actors as *regulators*. After all, as the introductory example of Google's Project Maven emphatically shows, state actors like the US Department of Defence are also *buyers* of lead firms in neural

[15] This quote is also helpful to justify my unit of analysis because it contextualises the European Commission's regulatory efforts with respect to AI as part of a broader political strategy to create a Digital Single Market. Crucially, AI regulation does not emerge out of a historical or ideological vacuum. In Chapter 5, I clarify this point in more detail.

production networks. Like other clients, states are thus enveloped by the computational resources of technology giants, which might impact their points of regulatory leverage. With that interdependency in mind, analysing the paradigmatic case of the EU's Digital Single Market Strategy provides a feasible way to render intelligible a different facet of the broader phenomenon in question (i.e., artificial intelligence): that is, the capacity of state actors to co-constitute AI's geographies of production.

### *3.3.3 The Paradigmatic Media Object*

Arguably, the third research question (*What do AI's geographies of production imply for platformised cultural production?*) could be answered in a plethora of ways. For good reasons, there is not a singular and pre-defined methodological toolkit for the study of digital culture. Unlike the selection of dominant lead firms, this case study selection cannot be justified with quantitative metrics such as market shares or market capitalisation. Along similar lines, compared to particularly influential expressions of state action, it is much harder to retrospectively explain the paradigmatic properties of a cultural phenomenon. To defy those challenges, the aim of this section is to combine a subjective account of the research process with theoretical considerations.

In early 2021, when completing the first draft of Chapter 4, I developed a strong interest in a particular type of neural network architecture developed by the Microsoft-backed AI research lab OpenAI: DALL-E. As a so-called text-to-image model, DALL-E is designed to generate digital images based on text queries. In the context of Chapter 4, my research focus was on how headline-generating and resource-intensive AI models like DALL-E relate to how lead firms in neural production networks consolidate their industry dominance – in this case, Microsoft. But beyond addressing such matters of industrial organisation, I also got interested in the cultural ramifications of those text-to-image models. While DALL-E's source code is not publicly available, developers soon shared their attempts to replicate the model's underlying computational and algorithmic techniques by releasing less potent models as free versions. *WIRED* called one of those replications the 'Internet's Favorite AI Meme Machine' (Knight 2022). In a nutshell, the creation of AI-generated images has become a matter of a few clicks; there is no need to have any machine learning expertise for using text-to-image models.

In many ways, text-to-image models are intriguing objects of analysis. However, *in isolation*, they do not have paradigmatic properties that could be analysed to spell out the cultural ramifications of AI's production geographies. Instead, of particular interest to me was their *integration* into the circuits of platformised cultural production. That is, the 'reorganization of cultural production and circulation' (Nieborg & Poell 2015: 15) around the logic of platforms. In their theory-building article on the platformisation of cultural production, Nieborg and Poell (2015: 15) similarly resort to what they describe as 'paradigm cases of platform-dependent contingent commodities': free-to-play games and platform-native news. Inspired by how Nieborg and Poell (2015) select those cases in order to render intelligible broader shifts in cultural production based on the platform business model, my aim was to select paradigmatic cases that render intelligible broader shifts in cultural production based on AI-generated media[16].

To account for this aim, I began to probe how the underlying machine learning techniques that give rise to contemporary forms of AI-generated media, such as images

[16] In addition to the notion of paradigmatic cases (Pavlich 2010), a different methodological terminology to describe such a purposive case selection strategy relates to the distinction between *intrinsic* and *instrumental* cases. According to Stake (2005: 445), the intrinsic case study is 'not undertaken primarily because the case represents other cases […] but instead because, in all its particularity *and* ordinariness, this case itself is of interest.' In contrast, the instrumental case 'plays a supportive role, and it facilitates our understanding of something else' (Stake 2005: 445). While there is no clear-cut boundary between those ideal types, paradigmatic cases are arguably closer to instrumental cases.

generated by DALL-E (i.e., deep generative models, see Chapter 6), integrate with, and rely on, already-existing infrastructures of cultural production. A crucial insight of this phase of the research process was that technology giants such as Amazon, Google, and Microsoft are by no means the only firms that own those infrastructures. Instead, when it comes to particular computational challenges, lead firms are actually dependent on the services of smaller companies, including Unity Technologies and Epic Games. The key assets of those firms – assets than cannot be reproduced by digital behemoths – are their powerful video game engines. Initially developed for the production of video games like the above-mentioned Fortnite, in recent years game engines were repurposed as *training grounds* for artificial intelligence[17] (Freedman 2018). By training grounds, I mean that game engines now act as customisable tools for creating synthetic environments within which the parameters of artificial neural networks can be optimised.

Herein lies the paradigmatic property of game engines as a case to elucidate the cultural ramifications of neural production networks: their cross-sectoral variability. In *The Language of New Media*, Manovich (2002: 56) aptly notes:

> A new media object is not something fixed once and for all but can exist in different, potentially infinite, versions. […] Old media involved a human creator who manually assembled textual, visual and/or audio elements into a particular composition or a sequence. This sequence was stored in some material, its order determined once and for all. […] New media, in contrast, is characterized by variability. Instead of identical copies a new media object typically gives rise to many different versions. And rather being created completely by a human author, these versions are often in part automatically assembled by a computer.

[17] According to Freedman (2018: 318), video 'game engines have been used by a wide array of industries to contour play, educate, train, and shape our perception of the world.' As 'an object of critical inquiry, [game engines] can be used to illustrate the synthetic power of computing and the instrumentality of code' (Freedman 2018: 326).

As Chapter 6 demonstrates, the ways in which this variability expresses itself across cultural industries are contingent and context-dependent. Nonetheless, although there is a wide range of use cases that deploy game engines for AI-generated media – from autonomous vehicles to virtual influencers – I identified an intriguing pattern: The confluence of game engines and deep generative models coincides with the emergence of new revenue streams for Unity Technologies and Epic Games, potentially affecting power relations in the 'platform society' (van Dijck et al. 2018).

In summary, a focus on game engines as paradigmatic cases is suitable not only due to their variable and modular properties as media objects (Manovich 2002) but also in light of their prototypical value for illustrating the cross-sectoral expansion of rentier relations (Sadowski 2019; Bernevega & Gekker 2021). Analysing new revenue streams to monetise game engines alongside the phenomenon in question, artificial intelligence, provides a glimpse into the different ways in which cultural practices are mediated by, and interact with, burgeoning forms of AI-generated media.

### 3.4 Triangulation of Empirical Evidence

To empirically substantiate my theoretical arguments, I planned the paradigmatic case study design in a way that allows for the triangulation of multiple sources of evidence. As Cox and Hassard (2010: 944) summarise, 'triangulation is an approach derived from navigation, military strategy, and surveying; it is based on the logic that researchers can move closer to obtaining a "true" picture if they take multiple measurements.' As such, it is important to stress that the notion of triangulation carries a positivist etymological and methodological history: an objectivist conception of the world as a fixed and stable reality. By contrast, given my firm belief that methods do not *describe* their objects of study but rather *co-constitute* them (Law 2004), I do not use the term triangulation as a positivist instrument to 'reduce bias' (Cox & Hassard 2010: 944) and thus increase the case study's objective verisimilitude. Instead, as suggested by Cox and Hassard (2010: 946) I decisively deploy the term in a reflexive manner, as an impetus to be transparent and upfront about the 'dilemmas, inclusions, and exclusions of the research process as a social and relational act.' With that in mind, the boundary between reflecting upon data collection and analysis methods, and ethics, positionality and limitations is not as much of a hard-and-fast line as it is occasionally presented in the literature. As Denzin (2012: 82), an influential scholar in the realm of qualitative inquiry, notes:

> The use of multiple methods, or triangulation, reflects an attempt to secure an in-depth understanding of the phenomenon in question. Objective reality can never be captured. We only know a thing through its representations. Triangulation is not a tool or a strategy of validation but an alternative to validation […]. The combination of multiple methodological practices, empirical materials, perspectives, and observers in a single study is best understood as a strategy that adds rigor, breadth, complexity, richness, and depth to any inquiry.

Triangulation, in this understanding, is a set of iterative processes that construct and thereby modify their objects of analysis by gathering data in multiple ways. This implies that 'researchers cannot definitively state the unit of analysis at the outset of the research; it must come into focus as the research progresses' (Vanwynsberghe & Khan 2007: 9). This spotlight on the iterative nature of triangulation is especially pertinent in the context of COVID-19 disruptions. As mentioned at the beginning of this chapter, my initial units of analysis looked very different from the units of analysis displayed in Table 3.2. Rather than focusing on lead firms, state actors, and cultural practices, my two groups of case units consisted of outsourcing firms that provide AI data annotation services and clients that make use of their services.

To gather data, I planned multiple months of on-the-ground fieldwork in India. My aim was to triangulate data gathered through a multi-sited workplace ethnography, including observations and semi-structured interviews, with a document analysis of how firms and clients discursively articulate their positionality within this digital division of labour. I prepared my fieldwork with a start date in late April 2020, secured affiliation with a local university in Bangalore, was granted a one-year visa and ethical approval, found accommodation, and generously received fieldwork funding from the Economic and Social Research Council (ESRC). Unfortunately, the fruits of my preparatory work could not be taken up and I was forced to cancel those plans. In 2020, I then completely restructured my research strategy, methodologically and theoretically.

Specifically, rather than utilising theories from cultural geography and science and technology studies (i.e., to craft arguments that needed to be substantiated with on-the-ground fieldwork and ethnographic observations), the focus of my research shifted towards questions of political economy, regulation, and media theory. A reconfiguration of data gathering and analysis methods was necessary to account for this shift in focus. Retrospectively, it is not an exaggeration to state this reorientation phase represented the most challenging period during the research process. In the spirit of being upfront about 'dilemmas, inclusions, and exclusions' (Cox & Hassard 2010: 946), the central dilemma that I faced related to the question: How to find a balance between developing original arguments and triangulating appropriate data to underpin them?

Before acknowledging the fact that my answer to this question coincides with a range of limitations that need to be addressed, it is worth specifying that answer. The paradigmatic case study design triangulates document analysis, event observations, and informal conversations. The main data source is publicly available industry and policy documents. Those data sources are supplemented by event observations and informal conversations with industry and policy experts. The next section begins with document analysis and then proceeds to supplementary techniques.

### *3.4.1 Document Analysis*

On the most general level, document analysis can be defined as a 'systematic procedure for reviewing or evaluating documents' (Bowen 2009: 27). In the context of case study approaches, document analysis is one of the most common ways to gather and interpret empirical data. As Olson (2010: 318) puts it, 'document analysis can provide a window into a variety of historical, political, social, economic, and personal dimensions of the case beyond the immediacy of interviews and observations.' Due to the broad range and scope of available documents with regard to my research focus, this instrument of data collection is well-suited as part of a paradigmatic case design.

By selecting document analysis as my primary data source, I posit that there is not a *lack of data* on AI's geographies of production, but rather a *lack of theory* to make sense of it. For example, Amazon, Google, and Microsoft regularly publish technical papers and documents that allow for rich insights into how they operate as lead firms in neural production networks. However, no economic-geographical study exists that uses those sources to rethink and rework established ways to map and examine production geographies[18]. Along similar lines, given that there is a myriad of speeches, reports, and whitepapers on the EU's Digital Single Market Strategy – such as about specific policy proposals such as the Digital Markets Act and the Artificial Intelligence Act – those rich sources provide useful evidence on the relationship between state actors and lead firms.

[18] In production network scholarship, document analysis is a well-regarded methodological tool for understanding industry contexts that are difficult to access through fieldwork (or when fieldwork is not possible). When it comes to tracking corporations across space and time, Goldstein and Newell (2020: 4) usefully distinguish between what they call 'ex-situ' and 'in-situ' approaches. While ex-situ approaches can help 'to construct linkages and identify hotspots', in-situ techniques 'tend to focus one or a few participants or study sites.' Given my research questions, an ex-situ approach is feasible and appropriate.

This argument also applies to the repurposing of game engines as new synthetic training grounds for artificial neural networks. In short, the analysis of documents is necessary to answer all three research questions that are featured in Table 3.2.

As with any methodological tool, document analysis dovetails with a range of challenges, the most notable one being that ‘documentary data may be irrelevant to the present research purpose, and the sheer volume of documentary data can become overwhelming’ (Olson 2010: 320). I aimed to defy this challenge by refraining from relying on ‘a predilection for a natural science approach (and positivism in particular) [and] an objectivist conception of social reality’ (Bryman 2012: 160) when analysing documentary evidence. During the research process, I attempted different analytical techniques and strategies to make sense of empirical evidence, including the use of qualitative research tools such as NVivo to create coding schemes. However, in my experience throughout the three case studies, such attempts at *objectively* systematising modes of data analysis by using software-based coding were not fruitful. This is not to devaluate the value of a ‘systematic procedure’ (Bowen 2009: 27) but rather a reflexive account of the research process.[19] For being transparent about sources without creating a division between corpus and case studies, I include them as appendices.

As Table 3.2 and the appendices indicate, I relied on a heterogenous corpus of documentary sources, including technical papers and research articles, media articles, marketing materials, and various policy materials (e.g., legislation, speeches, reports, and whitepapers). A corpus can be defined as ‘an ensemble of inscriptions or traces that

[19] As C. Wright Mills (2000: 72) provocatively argues in *The Sociological Imagination*, ‘those in the grip the methodological inhibition often refuse to say anything about modern society unless it has been through the fine little mill of The Statistical Ritual.’ But I do not wish to generally deny the strengths of objectivist approaches to document analysis. Instead, I aim to emphasise that they were not fruitful for my research purposes.

have undergone the process of selection, cleaning and refining necessary to prepare them for scientific analysis' (Venturini et al. 2018: 4200). For the purposes of my thesis, the processes of cleaning and refining were not necessary as I did not seek to modify the documentary sources myself. Given that the methodological rationale was to conduct a close reading and interpretative analysis in order to inform the conceptual arguments of the case studies, selecting appropriate sources was the most challenging task. Although the time period to which those documents refer ranges from 1979 to 2022, most of the analysed documents were published between 2012 and 2022 – a timespan that coincides with the mushrooming interest in artificial neural networks following the ImageNet success in 2012 (see section 2.1.1, Chapter 2). In this context, my strategy to defy the definitional ambiguity of artificial intelligence by focusing on a specific constellation of services, techniques and models that rely on the use of artificial neural networks offered a useful guiding structure when assembling and analysing the corpus.

Consequently, I situate my methodological approach as part of an interpretivist paradigm. This approach 'assumes that there are many points of entry into any given reality' and it 'emphasises an often story-like rendering of a problem and an iterative process of constructing the case study' (Vanwynsberghe & Khan 2007: 8). Each of the following chapters begins with a story that sets the stage for the respective case study. My strategy to iteratively craft those analyses is inspired by what Glaser and Strauss (2006: 101) call the 'constant comparative method of qualitative analysis.' According to them, the defining rule of this analytical approach is: 'while coding an incident for a category, compare it with the previous incidents in the same and different groups coded in the same category' (Glaser & Strauss 2006: 106). In less commanding terms, it can also be understood as 'much more fluid search for themes or meanings that recur in a variety of documents leading to categories' (Olson 2010: 319). Although the substance of those categories may be informed by the insights of previous research, a high degree of analytical openness vis-à-vis potential frictions and tensions between the empirics of the case studies and already-existing theory was pivotal for this method. In other words, the theoretical framework of neural production networks did not exist prior to the study but was constructed on the basis of the empirical materials. Combining an interpretivist paradigm with a constant comparison of insights within and across case studies (Glaser & Strauss 2006) worked well for answering my research questions.

To explain how I interpreted this approach in practice, it is worth using Chapter 4 as an example. Due to my focus on how Amazon, Google, and Microsoft act as lead firms in neural production networks, I began the analysis of the documentary materials with a focus on which company the respective source relates. I then created different categories to craft the structure of the case study, such as "computational resources" and "open-source tools." When finding an interesting piece of empirical evidence for one of those categories with respect to a particular lead firm, I compared it to the other lead firms. This iterative comparative approach to data analysis helped me to find intriguing structural similarities. Those similarities offer a structure for the flow of the case study analysis – and an empirical basis for my conceptual argument that neural production networks represent a distinct form of platformised industrial organisation.

In this regard, the major point of difference in the modes of analysis of the three chapters relates to the substance of those categories. When working on Chapter 5, for example, I began by using categories emerging from economic-geographical research on state roles in global production networks: "regulator", "facilitator", "producer", and "buyer." During the research process, it occurred to me that those state roles should not be understood as discrete categories but rather through the prism of the contradictions or frictions that play out between them. In other words, while Chapter 4 finds a number of structural similarities *between* different cases, Chapter 5 problematises tensions *within* one case. Before addressing the limitations of this flexible interpretation of Glaser and Strauss' (2006) constant comparative method and documentary evidence more broadly, it is worth discussing supplementary data-gathering techniques.

### *3.4.2 Event Observations*

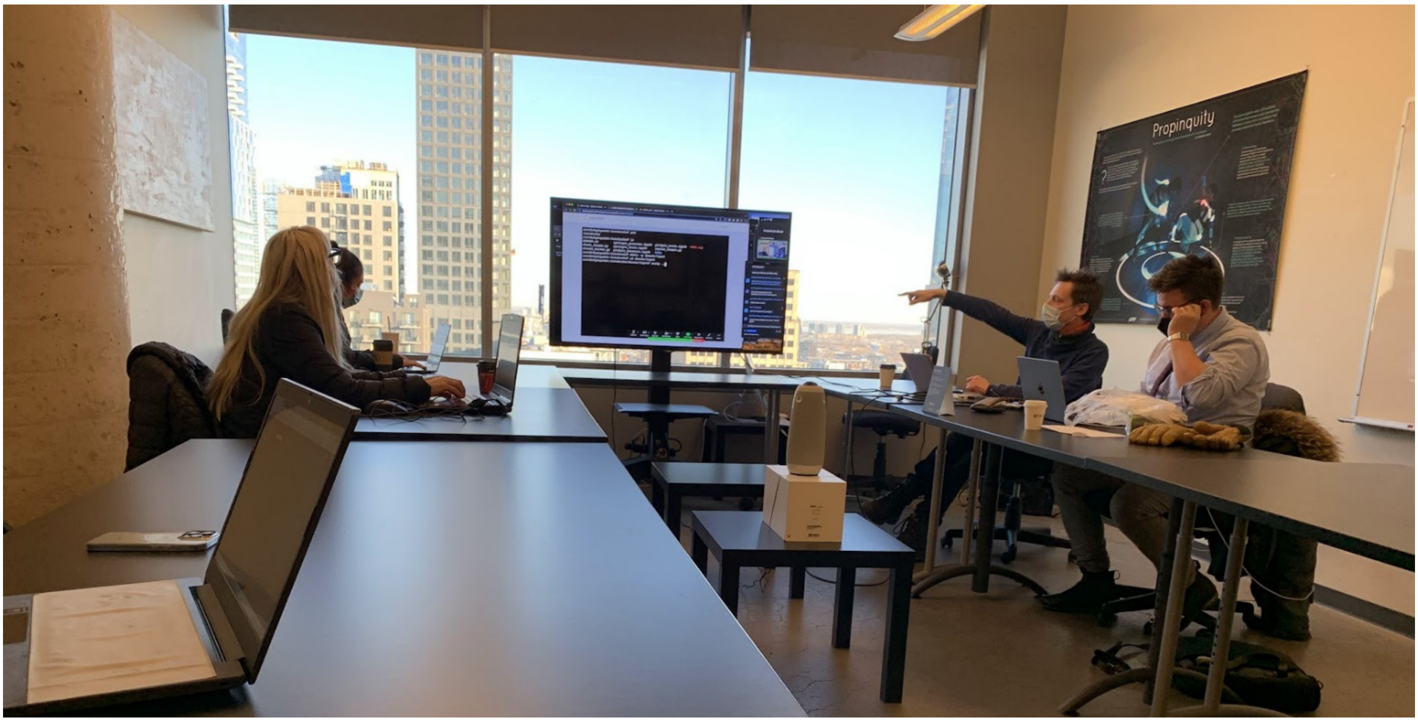
The empirical insights gathered through document analysis were supplemented by the attendance of industry and policy events, workshops, and informal conversations with industry and policy experts. The main purpose of those data-gathering techniques was to inform the primary mode of data collection (i.e., document analysis). As Olson (2010: 318) states, 'research participants can point to valuable documents the researcher might not have thought of, especially those that are not completely public.'

Consequently, the three research questions that structure my paradigmatic case study design (Table 3.2) required different supplementary techniques. With respect to Chapter 4, attending in-person industry events was not possible because the principal proportion of research for this chapter coincided with the first year of the COVID-19 pandemic (mid-2020 to early 2021). During this time, no in-person events and trade conventions took place that could have provided entry points to understand patterns of industry concentration. I thus had to rely on observing virtual industry events, such as Google's annual developer conference in 2021, in addition to recorded versions of earlier events. Naturally, the quality of data that can be gathered in these ways is limited as no interaction with participants is possible and the rationale of lead firms is to use those events for marketing purposes. But I still gained a number of useful insights. For example, although I was already aware of the role of proprietary hardware assets such as Google's tensor processing units (see Chapter 4), observing how Google executives framed those computational resources reinforced the necessity to investigate how such assets are positioned within AI's geographies of production.

In terms of my research for Chapter 5, I was fortunate that I could gain access to EU-level policy events and experts as a member of a pan-European network of doctoral researchers that I joined in 2020. My initial research plan entailed a comparative study of Canadian and European regulatory frameworks, but I eventually decided to focus on the EU's Digital Single Market Strategy as a paradigmatic case study. My membership in a network organised by the Foundation for European Progressive Studies – a Brussels-based think tank that is affiliated with the social-democratic party family in the EU Parliament – led to my participation in a range of virtual workshops, debates, and informal conversations about AI policy developments. As Swain and King (2022: 1) argue, such informal conversations 'create a greater ease of communication' than formal interviews and can 'complement and add to more formal types of data.' In addition to pointing me to relevant policy proposals and legal developments, those observations also provided me with a better understanding of the contradictions between different state roles: a key argument developed in Chapter 5.

Regarding Chapter 6, it is worth noting the entire qualitative research for this case study took place during my time as a visiting researcher in Montréal (August 2021 to July 2022). I began this research phase with a general interest in the nexus between synthetic environments and AI-generated media, but had yet to define and delineate my case units. For achieving this 'circumscription of the unit of analysis' (Vanwynsberghe & Khan 2007: 9), event observations and informal conversations provided intriguing sources of inspiration. For example, while in-person events were not possible due to social distancing policies, I observed a range of virtual industry events organised by the Synthetic Futures community, which defines itself as a group of 'diverse individuals, companies, and organizations who share our mission to shape a positive future for synthetic media.' Those observations led me to problematise the links between specific discursive strategies (e.g., the metaverse) and platform-based business models. Once social distancing policies were eased, a particularly inspiring workshop focused on the algorithmic and computational techniques that underlie AI-generated media. By training a deep generative model myself (Figure 3.1), I developed a much better understanding of the technical dimensions of this phenomenon. Chapter 6 is the result of a confluence between those discursive, economic, and technical insights.

Figure 3.1: Participation in an interactive workshop on AI-generated media in Montréal.




In summary, the purpose of this section was to demonstrate that the development of conceptual arguments cannot be separated from the methods and analytical tools that inform them. Rather than seeing my research strategy as an objectivist tool for knowing reality, I explicitly position it as a socially constructed entry point into AI's geographies of production. And this entry point has its limitations.

### 3.5 Research Ethics and Limitations

In this concluding section, I focus on questions of research ethics and limitations. Those two elements are closely related, as my interpretation of ethical research entails a need to be transparent and reflexive about limitations. By explaining how my positionality does not only shape empirical findings but also conceptual arguments, I aim to provide essential context to situate the contributions of this thesis.

Ethical approval was granted by the Oxford Internet Institute DREC on behalf of the Social Sciences and Humanities Inter-Divisional Research Ethics Committee at the University of Oxford (Ref: SSH_OII_CIA_20_005). The topic of research ethics offers yet another occasion to reflect upon the unpredictability of the COVID-19 pandemic. I revised and resubmitted my ethics form several times, each time in the hope that formal in-person interviews would become a feasible and sustainable option. Eventually, I had to make the decision to entirely plan without semi-structured interviews, as I did not perceive remote expert interviews as a viable alternative to establish a sufficient level of trust. With those formal ethics procedures in mind, the discussion of limitations can be separated into three broad categories: methodological limitations (i.e., due to my data gathering and analysis tools), epistemological limitations (i.e., due to my use of holism), and disciplinary limitations (i.e., due to my cross-disciplinary approach). Of course, there are also intersections between those categories.

To start with methodological limitations, it is pivotal to acknowledge the pitfalls of resorting to documentary evidence as the primary source of data gathering. Each of the documents analysed in this thesis fulfils a particular function for those who created it. For example, a marketing document, by definition, serves to exaggerate the product or service that it promotes. A scientific paper published by one of the lead firms is only publicly available because it has successfully passed multiple rounds of internal review, ensuring that it remains in line with official corporate policy[20]. The rationale for this is to omit particular details to prevent journalists, competitors, and regulators from gaining insights into their proprietary operations. Thus, it is important to go beyond exclusively considering the *content* of documents (as advocated by some quantitative approaches to content analysis) and probe their *relational embeddedness* vis-à-vis corporate interests, political ideologies, and bureaucratic procedures. As Hull (2020: 253, my emphasis) convincingly argues, 'to look *at* [documents] rather than *through* them, is to treat them as mediators, things that "transform, translate, distort, and modify the meaning or the elements they are supposed to carry" (Latour 2005: 39).'

In different ways, each of the following chapters attempts to take seriously those relational and mediating features of documents, rather than perceiving the data gathered through them as taken-for-granted facts. For example, the role of intellectual property rights and non-disclosure agreements that obfuscate trade secrets is directly related to the transformation of things into assets. As Birch and Muniesa (2020: 6) explain, those procedures have been 'legally instituted to give owners both exclusion rights to the use of the asset itself and to the use of any copies derived from the asset.' To return to the paradigmatic case of video game engines, this means that – in contrast to commodities – owners of those engines can exert control over the ways in which those media objects are deployed. Exerting this form of control is much more difficult in the case of, say, a smartphone that can be modified by changing its hardware or hacking its software. In

[20] An illustrative example of how those corporate checks and balances play out in practice can be found in the case of Timnit Gebru and Margaret Mitchell, two former Google AI ethics researchers. Both were fired by the company after Gebru refused to retract her name from a paper that Google did not approve for publication. The paper stressed the environmental impacts of models used for natural language processing, which is one of the most common application domains of artificial neural networks (Hao 2020).

short, even though I acknowledge the limitations of documents as socially constructed expressions of particular interests, I argue in favour of attempting to turn this limitation into a strength by explicitly theorising those expressions and interests.

Speaking of methodological limitations also requires a consideration of potential alternative techniques of data gathering and analysis. As I emphasised throughout the chapter, I do not perceive my methodological strategy as the one-and-only approach. Conceptually, the empirical analysis of AI's infrastructural geographies could also be firmly grounded in cultural geography, financial geography, or potentially more-than-human geographies. Methodologically, a range of alternative options is conceivable, such as a multi-sited ethnography of data centres, the quantitative analysis of financial transaction data, or an auto-ethnography of how artificial neural networks reconfigure artistic practice. Given my understanding that infrastructure is not defined by *what* it is but rather by *how* it is (Peters 2015), I do not wish to imply that my interpretation of the term is better than alternative interpretations. Additionally, there are also alternative ways to answer my specific research questions, such as by selecting other case units or by significantly reducing the analytical scope. Those considerations are an impetus to state the epistemological limitations of my approach.

In this regard, the most obvious limitation with respect to my research strategy relates to its strategic prioritisation of *scope* over *depth*. It would not only be reasonable but also productive to flag one of, or a combination of, the following points. First, the number of case units is too high. Second, it is not possible to juxtapose those cases as they relate to different things (i.e., lead firm, state action, media object). Third, the point of a paradigmatic case is that there is only *one*, not multiple cases (Pavlich 2010). Fourth, the cases are not adequately temporally and spatially bounded. Fifth, the goal of the case study approach, a thick description, cannot be reached with document analysis.

The rationale of listing those points is not to set the stage for a comprehensive response to all of them, but to engage with the spirit of their critique by providing transparency about the limitations of my paradigmatic case study design.

When consulting the methodological literature building on what Rogers (2013) has called 'digital methods', common advice is that 'researchers should concentrate on specialised questions (e.g., whether peer-to-peer renting has professionalised in a given city and in a given period of time) and on restricted subsets of the traces generated by the medium they investigate' (Venturini et al. 2018: 4209). Certainly, this argument is highly relevant if the goal is to study the 'repurposing of the inscriptions generated by digital media for the study of collective phenomena' (Venturini et al. 2018: 4195). But for research designs whose purpose is to open out new conceptual spaces to study how different enactments of a phenomenon (i.e., artificial intelligence) relate to each other, 'a broader approach is required to understand phenomena in their totality or situ where the whole is greater than the sum of its parts' (MacQuarrie 2010: 442). How does this point relate to Law's (2004) epistemological scepticism and Bennett's (2010) emphasis on non-human forces mentioned at the start of this chapter?

When I wrote the first version of this methodology chapter in mid-2019 – before the COVID-19 pandemic caused me to change my research plans – I did not foresee that I would resort to holism (MacQuarrie 2010) to contextualise my research design. On the contrary, inspired by science and technology studies[21], I argued in favour of pushing

[21] In particular, I was influenced by Mol's (2002) *The Body Multiple*. In the book, Mol shows how a disease, atherosclerosis, has diverging meanings for different actors in the interlaced processes of medical practice. From Mol's perspective, the shape of the disease is not *pre-given* or prior to practices of bodies, patients, doctors, technicians, and technologies, but rather a *consequence* of their entangled relations. When designing a research strategy to grasp the division of labour between data scientists and data labellers, I attempted to formulate a similar set of arguments in relation to artificial intelligence. For example,

back against modes of knowledge creation that seek to extrapolate from the study of situated practices (e.g., outsourced data annotation for AI systems) to higher levels of conceptual abstraction. I was intrigued by Saldanha's (2003: 427) interpretation of Kwa's (2002) distinction between two ideal types of problematising and representing complexity, 'romantic complexity' and 'baroque complexity':

> Romantic complexity comes from a Platonic tradition and is holistic**,** wishing for a transcendent unity (e.g., 'system,' 'society,' 'class,' 'organism,' 'subject') […]. Baroque complexity, on the other hand, is a 'bottom up' approach to matter. Every atom or constituting element is a reality by itself; although elements do work together to form larger entities, there is no 'blue-print' prior to this formation.

In appreciation of the fact that the point of ideal types is to delineate a spectrum of possibilities that are positioned in-between them, it can be inferred that what started as an enquiry of baroque complexity morphed into an enquiry of romantic complexity. In other words, holism was not a deliberate choice at the outset of this research process but rather became a methodological necessity in direct response to an exogenous force: the COVID-19 pandemic. For articulating the limits of this gradual shift in focus, it is helpful to recapitulate Bennett's (2010: vii) arguments about the agential capacities of non-human forces: 'By "vitality" I mean the capacity of things, edibles, commodities, storms, metals – not only to impede or block the will and designs of humans but also to act as quasi agents or forces with trajectories, propensities, or tendencies of their own.' Although I do not explicitly think through the notion of vibrant matter for answering my

---

while workers in outsourcing firms enact AI by annotating training datasets, data scientists enact AI by writing code and fine-tuning the parameters of artificial neural networks. As Mol (2002: 5), 'objects come into being – and disappear – with the practice in which they are manipulated. And since the object of manipulation tends to differ from one practice to another, reality multiplies.' Crucially, adequately developing such conceptual arguments would have necessitated extensive ethnographic fieldwork.

research questions, it influences how I situate my conceptual arguments. Put differently, I decided against portraying the force of the COVID-19 pandemic as an epistemological *limitation*. Instead, I perceive it as a *prerequisite* for my conceptual arguments. Without a fundamental shift in focus, they would not exist in the first place.

However, this does not mean that I argue seek to establish a ‘totalizing account wherein the multiple, particular arrangements of capitalist accumulation are reduced to their parts in the system’ (Werner 2019: 953). Problematising the forces that enact, and are enacted by, AI’s geographies of production does not imply that the idiosyncrasies of those constitutive elements dissolve ‘while wishing for a transcendent unity’ (Saldanha 2003: 427). For example, just because I relate the EU’s Digital Single Market Strategy to the operations of lead firms in neural production networks, I do not deny the former or the latter their existence outside of this theoretical framework. Conceptually, the idea of neural production networks does not purport to subsume all economic, political, and cultural processes with respect to AI under one pre-defined umbrella. As a result of this, my use of holism can be seen as a ‘conceptual procedure, rather than an empirical or conceptual premise in which the whole is discovered through the analysis of the mutual conditioning parts’ (McMichael 1990, cited in Werner 2019: 953). To use MacQuarrie’s (2010: 442) terminology, due to the fact that ‘holism is an interpretative standpoint of the researcher,’ there is a need for transparency and reflexivity about ‘the interpreter’s design for what the big picture is and what that entails.’

Finally, in addition to methodological and epistemological limitations, there are also disciplinary limitations. By this, I refer to the fact that the research strategy of this thesis is situated at the intersection of economic geography, media studies, and political economy. Throughout Chapter 2 and this chapter, I attempted to emphasise the benefits of approaching the phenomenon in question by thinking beyond and across disciplinary boundaries. Nonetheless, I acknowledge there are also advantages of adhering to those boundaries more strictly. For example, a more in-depth focus on the role of technology giants as lead firms could have resulted in expanded economic-geographical arguments about the financial details of their value capture mechanisms. Similarly, a restriction of scope to media objects could have led to a comparison of a range of media objects that underpin AI-generated images or a more comprehensive discussion of their histories or their technological evolution. Additionally, a stronger commitment to the theories and approaches of critical political economy could have generated more detailed insights into the relationship between corporate power and state power.

Although it is important to mention those disciplinary limitations, I still prefer a research strategy that combines such considerations, rather than categorising them into clearly defined and hermetically sealed analytical silos. If disciplinary boundaries were to be taken as irreconcilable borderlines, *other* disciplinary limitations would arise. For instance, an approach that is exclusively rooted in media studies might neglect to probe the spatialities of artificial intelligence. By contrast, an approach that limits itself to the vocabulary and frameworks of economic geography might underestimate that platform-powered lead firms represent a new and distinct form of industrial organisation. Further, an approach that is exclusively rooted in the established procedures of critical political economy and historical materialism might not be appropriate for grasping the cultural ramifications of emerging phenomena such as AI-generated media.

To sum up this section, I provided a reflexive and comprehensive account of my research strategy, introducing and problematising my paradigmatic case study design. Chapter 7 returns to some of those considerations by laying out ideas and trajectories for future research that builds on my methodological approach. The next chapter marks the beginning of the empirical part of this thesis.

# Chapter 4: The Rise of Neural Production Networks

## 4.1 Introduction: AI and Technology Giants

In October 2015, AlphaGo, a computer programme developed by the London-based Google subsidiary DeepMind, famously became the first AI-powered software to beat a professional human Go player, Fan Hui. Five months later, AlphaGo outcompeted Lee Sedol, a victory that attracted even more attention due to the fact that he was the world champion. In 2018, Lee retired from professional play, highlighting that 'with the debut of AI in Go games, I've realized that I'm not at the top even if I become the number one […] there is an entity that cannot be defeated' (Vincent 2019).

While AlphaGo is commonly regarded as a milestone in AI research, Broussard (2018: 36) recalls the fact that its programmers relied on thirty million human games to train AlphaGo's artificial neural networks, 'including data from the world's best players who invisibly and over the course of years worked (without compensation) to create the training data.' AlphaGo's creators extracted the collective work of thousands of players without their permission in order to develop a system that eventually caused the world champion to retire. This story might be interpreted as a compelling case study about a discursive spotlight on the model's computational magic and an undervaluation of the accumulated human work without which it would not exist in the first place. But beyond such considerations of labour exploitation, I begin with the example of DeepMind's Go-playing AI software to show the importance of another, largely unexplored, question associated with the proliferation of artificial neural networks: *How do technology giants act as lead firms within AI's geographies of production?*

In answering this research question, the central argument of this chapter is that technology giants such as Google do not operate as lead firms of neural production networks in the same ways as lead firms of global production networks. I contend that this discontinuity manifests itself in three ways. First, technology giants combine the extraction of rent by providing infrastructural services with the cross-subsidisation of strategic nodes within their own ecosystems. Second, technology giants expand these proprietary ecosystems horizontally and vertically, thereby operating at the conjuncture of multiple geographies and production topologies. Third, technology giants consolidate their power by decentralising the production of AI and centralising the provision of AI, taking up an infrastructural role that goes beyond the offering of a service in exchange for money – and beyond the realm of economic production.

But why is it so important to grasp the relationship between AI and technology giants in the first place? In recent years, a wide range of scholarship has argued that the resources to develop and implement contemporary manifestations of AI, particularly those powered by artificial neural networks, are extremely concentrated and centralised (for example, Crawford 2021; Srnicek 2020a; Luitse & Denkena 2021; Steinhoff 2022; Whittaker 2021[22]). As Dyer-Whiteford et al. (2019: 43) specify, the 'control of cloud computing facilities, ownership of large data sets, and wealth to hire the best from a limited pool of AI talent' are crucial factors that compound the 'centralization and concentration of capitalist power.' Scrutinising the idea that data is the *only* competitive advantage for a handful of technology giants, Srnicek (2020a) points to the relevance of their planetary-scale computational infrastructure to make a case that 'far from being immaterial companies, [technology giants] are significantly embodied companies.' In a nutshell, it is rather uncontroversial in the critical political economy literature that AI dovetails with the centralisation of power in the hands of a few firms. Nonetheless, this

[22] What brings together this body of interdisciplinary political economy work is a shared interest in the ramifications of AI's role as the basis of a 'deep and extensive infrastructural reorganization of the capitalist economy' (Dyer-Whiteford et al. 2019: 31).

centralization of power in the context of AI has yet to be examined in terms of how it relates to potentially reconfigured geographies of production.

To do so, the chapter proceeds as follows. Providing a basis for the chapter's key arguments around the discontinuities of lead firms in neural production networks, I will first briefly elaborate on two ideal types of global production networks: buyer-driven networks and producer-driven networks. After justifying my empirical focus on Amazon, Google, and Microsoft as paradigmatic case studies, I will problematise how technology giants act as lead firms act in neural production networks. The chapter ends by pointing out the wider implications of these empirical findings.

### 4.2 Lead Firms in Global Production Networks

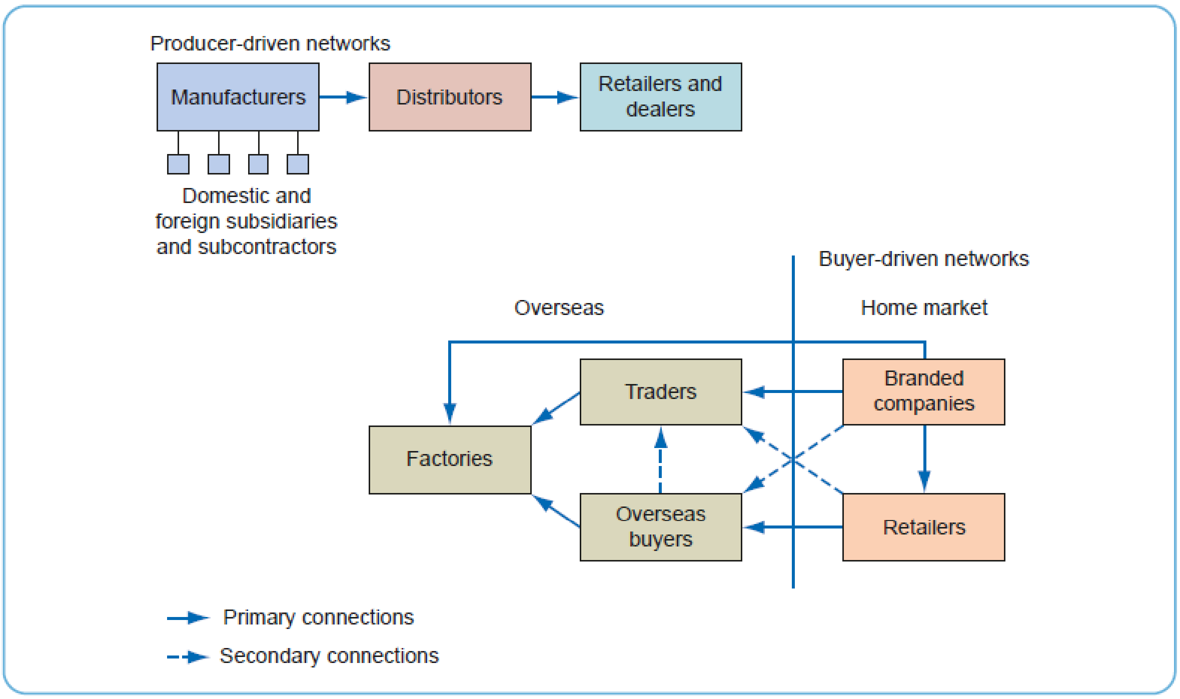
An apt summary of the role of lead firms in production networks is by Coe (2016: 323), who points to the established differentiation between networks that are buyer-driven and networks that are producer-driven[23] (Figure 4.1). While buyer-driven networks suggest that buyers, retailers, or merchandisers have the most significant influencers in shaping the geographies of production, producer-driven networks indicate the dominance of producers and manufacturers in shaping economic outcomes.

As Coe (2016) points out, buyer-driven networks are especially prevalent in work-intensive industries such as apparel, handicrafts, and sectors with respect to other consumer items that need a large amount of manual labour. In this regard, lead firms of buyer-driven networks outsource most of their actual production activities to low-income countries and aim to 'extract substantial profits from bringing together their

[23] This distinction goes back to a paper by Gereffi (1999). Since then, Gereffi et al. (2005; 2018) and others (Bair 2005; Coe et al. 2008) presented more complex typologies of production network governance. However, for the purposes of this chapter, it is useful to condense the analysis of previous work and focus on those two 'ideal types' (Coe 2016: 334).

design, sales, marketing and financial expertise with strong brand names and access to large consumer markets' (Coe 2016: 334). Examples of this type of lead firm include multinational companies such as Nike and Rebook, whose operations have been seen as paradigmatic examples of what Tsing (2009: 148) theorises as supply chain capitalism: that is, a 'model for understanding both the continent-crossing scale and the constitutive diversity of contemporary global capitalism.' Simply put, lead firms in buyer-driven networks consolidate their market dominance by orchestrating transnational production arrangements, rather than producing goods and commodities themselves.

Figure 4.1: Producer-driven networks and buyer-driven networks. Source: Coe 2016.




Producer-driven networks, by contrast, are more common in capital- and technology-intensive industries such as automotive, aviation, and semiconductor manufacturing (Coe 2016). Car manufacturers such as Volkswagen and Toyota, for instance, rely on a plethora of subsidiaries in places with low minimum wages and proximity to precious minerals that are needed for the production process. It is worth using the example of car manufacturers to clarify the trajectories of value in producer-driven networks, commonly understood as value creation, value enhancement, and value capture (Hess 2018). For value creation, lead firms combine the conversion of workers' labour-power into actual labour with generating possibilities for the extraction of rent, exemplified by opening up a new plant. When it comes to value enhancement, the focus is on how lead firms maximise this value. A car manufacturer might negotiate favourable export conditions with the local government. The dimension of value capture addresses the question of who retains and appropriates this economic value. A lack of local ownership and control by local actors and detrimental consequences of a potential plant closure would typify this analytical category in our example.

However, a problem that complicates the uncovering of how value is created, enhanced, and captured in both producer- and buyer-driven contexts is the opacity of global production networks. As Coe (2016: 322) puts it, 'almost all the commodities that we consume have complicated histories and geographies and yet, as a system, capitalism seems to conceal these.' The spatialities of capitalism, in other words, are inherently only ever *partially* visible. Although the analytical repertoire of scholars has matured to encompass a more granular range of organisational structures and dynamics (Gereffi et al. 2005), efforts to probe the spatial complexity of firms offer an incomplete picture (Werner 2019). This issue is exacerbated by the blurriness of the notion of the firm itself. While it serves as a quintessential reference in a plethora of papers, Maskell (2001: 330) argues that it still remains undertheorised:

> When the "firm" appears as an entity in the index of a book on economic geography, it is usually only in context-specific combinations like "firms, interrelationships between," "firm-environment relations," and "firm-level economics of scale." Although highly idiosyncratic aspects of the function or activities of the firm are thus selected for consideration, the firm as such seems to be regarded as self-evident, requiring neither definition nor discussion.

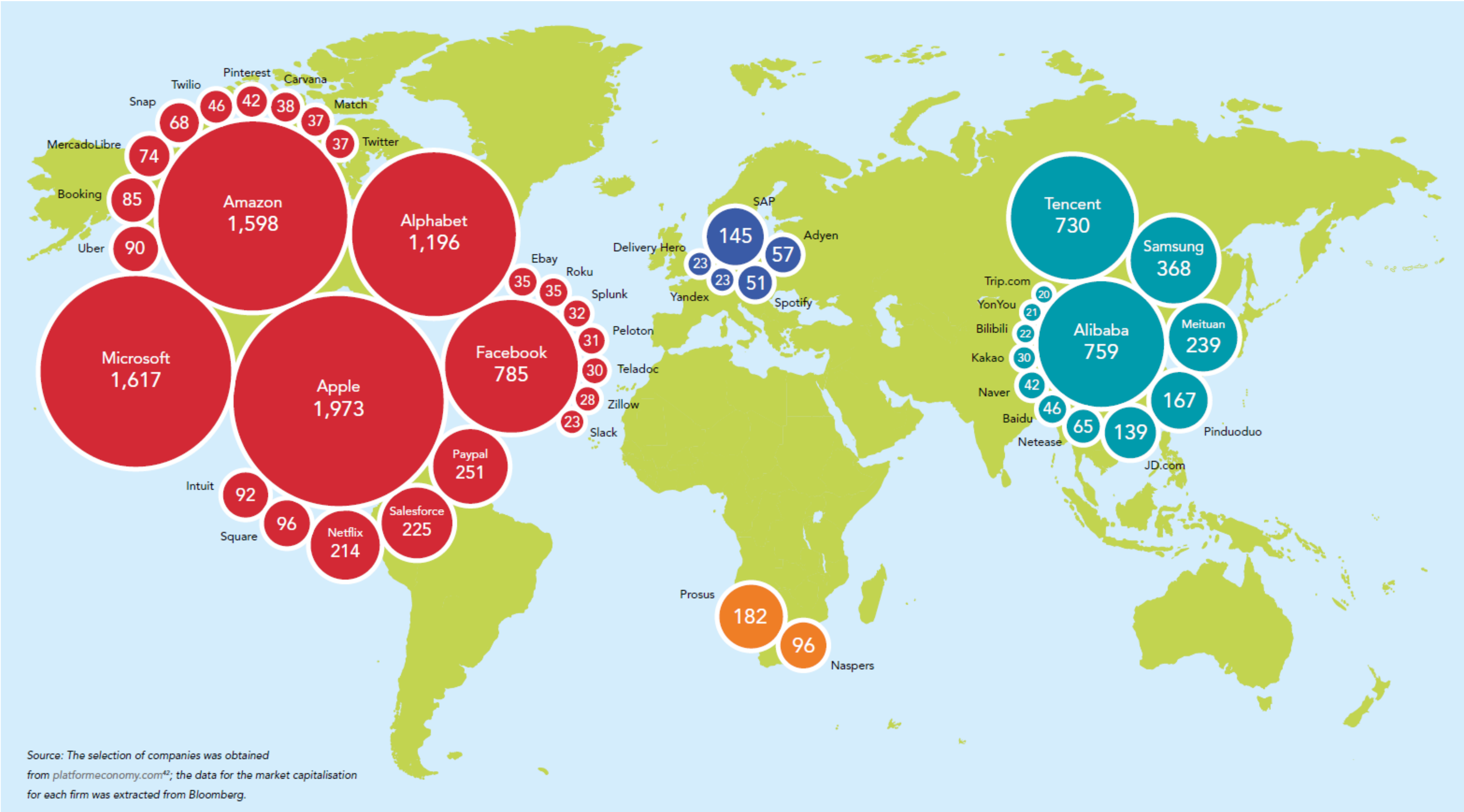
In the next section, I argue that an empirical focus on the paradigmatic role of technology giants as lead firms in neural production networks is useful to anticipate and minimise the pitfalls of this definitional blurriness. This is not to imply that technology giants are not complex organisations but rather a claim that the necessity of justifying a spotlight on them requires spelling out their defining feature as lead firms.

### 4.3 Overview of Case Studies: Amazon, Google, Microsoft

Within policy discourse and academic debates, several terms are deployed to refer to the world's largest and most dominant platform-powered technology companies. Prominent terms are 'Big Tech' (Whittaker 2021), 'gatekeepers' (European Commission 2022), 'internet giants' (Sandbu 2018), 'Big Four' (Galloway 2017) made up by 'GAFA' (i.e., Google, Amazon, Facebook, Apple), and 'Big Five' (van Dijck 2020), which includes Microsoft. Although all those terms prioritise different aspects, such as the number of companies included or the fact that their success is related to the internet or their role as gatekeepers, what unites them is an emphasis on the *scale* of specific firms.

One way to operationalise the scale of technology firms is by considering their market capitalisation, that is, 'the value for a listed company arrived at by multiplying the outstanding common stock or ordinary shares by the market price' (Moles & Terry 1997a). Using this metric to map the world's most valuable platform companies with a market capitalisation above US$ 20 billion, Fernandez et al. (2020: 20) geographically visualise the dominance of seven companies: Apple, Microsoft, Amazon, Alphabet (Google), Facebook, Tencent, and Alibaba (Figure 4.2). According to the authors, it is pivotal to distinguish those 'seven companies – Big Tech's "infrastructural core" – from smaller tech firms, which typically rely on their infrastructures' (Fernandez et al. 2020: 8). This distinction is crucial for the purposes of this chapter.

Figure 4.2: Market capitalisation of platform companies (above US$ 20 billion, December 2020). Source: Fernandez et al. 2020



Before relating this to neural production networks, it is worth clarifying what this reliance on 'Big Tech's infrastructural core' (Fernandez et al. 2020: 9) means in practice. For example, rather than running its own data centres, Netflix commissions Amazon's cloud computing subsidiary, Amazon Web Services, for most of its storage, data analytics, recommendation, video transcoding, and computing needs. Amazon's key competitive advantage over smaller cloud computing providers is its 'modular technological architecture composed of a core and a periphery' (Gawer 2014: 1240). While the *core* of this architecture is comprised of components with low variability (e.g., fixed material infrastructures such as data centres), components as part of the *periphery* (e.g., data analytics solutions) are characterised by a higher variability. A Netflix software engineer cited by Amazon Web Services (2017) emphasises that they 'experimented with multiple designs and used many AWS products' before eventually finding the most efficient cloud computing architecture.[24]

However, the crux is that there is only a handful of technology giants that can enact those core-periphery dynamics at an industry-dominating scale, especially with regard to the provision of computational infrastructure. Market research provides useful empirical context to substantiate this claim. Subsumed under the term 'infrastructure-as-a-service', the market research firm Gartner (2021a) analysed the market shares of what it sees as 'standardized, highly automated offerings in which computing resources owned by a service provider, complemented by storage and networking capabilities, are offered to customers.' Although the scope of this definition goes beyond AI because it

[24] Another creative way to imagine the relationship between *core* and *periphery* is by using the metaphor of a 'platform tree' (van Dijck 2020). The core units of a platform architecture, such as computational resources and data centres, are entrenched like the trunk of a tree. A platform's customisable services and modular solutions, such as tailored data processing pipelines as in the case of Netflix, can be thought of as the tree's trenches.

includes offerings such as cloud storage, website hosting and data analytics, it usefully indicates the market power of three American companies (Amazon, Microsoft, Google) and two Chinese companies (Alibaba, Huawei). As demonstrated by Table 4.1, Amazon accounts for nearly half of the worldwide infrastructure-as-a-service market, followed by Microsoft (19.7 per cent), Alibaba (9.5 per cent), and Google (6.1 per cent). Thus, as a basis to answer this chapter's question, the defining feature of technology giants as lead firms is their *dominance as providers of infrastructural services*.[25]

Table 4.1: 'Infrastructure-as-a-service' market shares (millions of USD). Gartner 2021a.

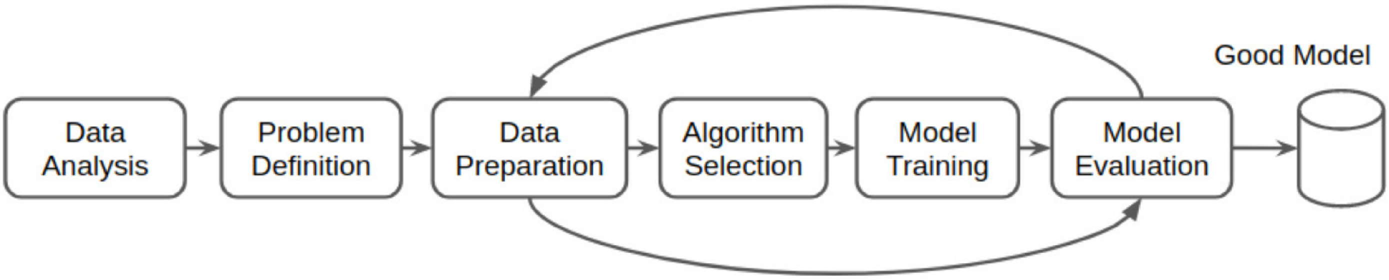
| Company | 2020 Revenue | 2020 Market Share (%) | 2019 Revenue | 2019 Market Share (%) | 2019-2020 Growth (%) |
|---|---|---|---|---|---|
| Amazon | 26,201 | 40.8 | 20,365 | 44.6 | 28.7 |
| Microsoft | 12,658 | 19.7 | 7,950 | 17.4 | 59.2 |
| Alibaba | 6,117 | 9.5 | 4,004 | 8.8 | 52.8 |
| Google | 3,932 | 6.1 | 2,367 | 5.2 | 66.1 |
| Huawei | 2,672 | 4.2 | 882 | 1.9 | 202.8 |
| Others | 12,706 | 19.8 | 10,115 | 22.1 | 25.6 |
| **Total** | **64,286** | **100.0** | **45,684** | **100.0** | **40.7** |

Against this backdrop, a constructive way to link this market power to the idea of neural production networks is to consider their computational envelopment not as a natural law, but as a result of economic-geographical processes enacted by lead firms. By definition, neural production networks inevitably remain enveloped by the necessity

[25] This is the central reason why my case study analysis does not include Apple and Facebook (now Meta Platforms). Although Apple has a higher market capitalisation than Amazon, Google, and Microsoft (Figure 4.3), the company's infrastructural services are restricted to their proprietary products (e.g., iCloud for Apple devices such as iPhones). Additionally, iCloud runs on Amazon's and Microsoft's computational infrastructure.

of computation. Consequently, those lead firms that control the computational resources that underpin and facilitate this envelopment are in a powerful position. In this respect, outlining the typical stages of building a machine learning model (Figure 4.3) provides essential context to begin analysing the relationship between computational resources and artificial neural networks.[26] The first three stages involve collecting data, defining a problem (e.g., classifying data into a set of categories), and preparing the data (e.g., annotating it with labels such as ‘cat’ or ‘dog’). Stages four and five relate to selecting an appropriate algorithm and training the model. The last stage is evaluating and fine-tuning the model based on how well it solves the problem.

Figure 4.3: Basic stages of building a machine learning model. Source: Das et al. 2021.




For example, a marketing agency that aims to make predictions from its social media data to inform future campaigns might not have the expertise and computational resources to perform such stages on its own. Consequently, the firm might resort to the same company that provides its cloud storage, as it is familiar with the provider’s user interface and payment structure. Due to the stark concentration of the infrastructure-as-a-service market, it is likely that this provider is one of the companies listed in Table 4.1. In recent years, Amazon, Google, and Microsoft have developed tailored services

[26] To clarify this relationship, artificial neural networks underpin most contemporary machine learning models and associated infrastructure-as-a-service offerings. But this is not to imply that *all* types of machine learning rely on artificial neural networks.

and infrastructure offerings for such situations. This emerging domain of 'AI-as-a-service' (Dyer-Whiteford et al. 2019: 42) includes both proprietary and open-source hardware and software solutions, spanning from fine-tuned machine learning models for computer vision tasks (e.g., Project Maven) to advanced speech recognition services. Table 4.2 offers an overview of a preliminary mapping of the breadth of those services, which complicates a condensed case study analysis. The point of this mapping is not to conduct an all-encompassing summary of all existing AI-as-a-service offerings. Instead, the purpose of Table 4.2 is to showcase the structural similarities between Amazon, Google, and Microsoft as lead firms in neural production networks.

Table 4.2: Structural similarities between lead firms in neural production networks.

| | **Amazon** | **Google** | **Microsoft** |
|---|---|---|---|
| **Relevant divisions** | Amazon Web Services<br>Amazon Science | Google Cloud<br>Google Brain, DeepMind | Microsoft Azure<br>Microsoft Research |
| **Specialised platforms** | Amazon SageMaker<br>AWS Marketplace | Google Cloud Platform<br>Google Cloud Marketplace | Microsoft Azure AI<br>Azure Marketplace |
| **Open-source frameworks** | AWS SaaS Boost<br>Deep Learning AMIs | TensorFlow, ML Kit<br>Fairness Indicators | Microsoft Cognitive Toolkit, ML.NET, ONNX |
| **Proprietary AI services** | SageMaker Autopilot<br>AWS Inferentia | Cloud AutoML Vision<br>Cloud TPU Service | Azure Cognitive Services<br>Azure Machine Learning |

The most important structural similarity is that AI's strategic importance extends beyond the scope of research and development (e.g., public displays such as AlphaGo), as AI represents an opportunity for technology giants to consolidate their dominance of the infrastructure-as-service market (Table 4.1). If more companies and organisations adopt AI to solve their problems (irrespective of whether that is appropriate in the first place), providers of the underlying infrastructure will benefit from this technological expansion. Speaking of computational envelopment is especially useful in this context as it enables us to grasp both *company-internal* and *company-external* implications of this infrastructural dependency. On the one hand, technology giants can leverage their edge when it comes to computational resources and complex artificial neural networks to improve and optimise their own services (e.g., Amazon's product recommendation algorithm or Google's advertising systems). On the other hand, they can use those assets to enclose users by acting as 'gatekeepers' (European Commission 2022) without whom it is not possible to deploy potent AI technologies. It is pivotal to consider lead firms in neural production networks not only as infrastructure providers for *other* actors but also as infrastructure providers for *themselves.* Mucha and Seppala (2020: 3) aptly describe this intriguing interplay between internal and external dynamics:

> Long before battles for mobile platform domination were fought, many digital platform companies were already using machine learning algorithms in their internal business processes […] Microsoft introduced spam filtering based on machine learning in 2003. […] However, [the] proliferation of AI technologies has also enabled [a] platformization of this new middleware layer in digital platform technology stack[s], thus effectively creating a platform inside a platform.

For example, all three technology giants provide marketplaces on which third-party developers can sell their AI software (Table 4.2). Similar to the role of third-party sellers on Amazon Marketplace, lead firms benefit from transactions between different user groups (i.e., sellers and clients) by taking a commission from each sale. Beyond merely bringing together different user groups as a multi-sited market (Rochet & Tirole 2003), Amazon, Google, and Microsoft also offer their own proprietary services, which are typically more prominently advertised than third-party services. But the emergence of this 'new middleware layer' has neither been investigated in conjunction with company-internal processes, such as research and development or strategies of cross-subsidisation (Fernandez et al. 2020) nor as being a constitutive part of AI's production networks. In other words, I argue that it is immensely fruitful to disentangle the Janus-faced characteristics of lead firms as both *external-facing* infrastructure-as-a-service providers and *internal-facing* distributors of their staggering computational resources. The advantage of using the metaphor of computational envelopment is that it recognises this duplicity as a defining feature of neural production networks, rather than side-lining its significance as a neglectable side effect of corporate power.

In comparing the operations of Amazon, Google, and Microsoft, I follow Plantin et al.'s (2018: 307) call that methods of juxtaposition are useful to reach the 'goal of comparative analysis that is often missing from some accounts of single platforms.' My rationale for choosing those three technology giants over Alibaba and Huawei is a rather pragmatic one, given that the number of English-language sources and documents about the modus operandi of Chinese technology giants is limited. Furthermore, in line with the methodological conventions regarding the paradigmatic case selection design (see Chapter 3), I argue that a sample of three lead firms is reasonable.

To sum up this section, the defining feature of technology giants as lead firms is their dominance as providers of infrastructural services – not just for others, but also for themselves. Building on this groundwork, the next section discusses the findings of the case study analysis that is at the heart of this chapter.

### 4.4 Lead Firms in Neural Production Networks

In different ways, lead firms in global production networks attempt to gain control over other actors by *exploiting territorial inequalities*. In producer-driven networks, this occurs by 'exerting control over 'backward' linkages to raw material and component suppliers, and 'forward' linkages with distributors and retailers' (Coe 2016: 333). In buyer-driven networks, this occurs by exerting control at the retail end (e.g., branding and design), while outsourcing actual production activities to other places. By contrast, lead firms in neural production networks attempt to create and sustain infrastructural dependencies by *enveloping the space of production activities*. This is a fundamental difference that justifies the necessity of a new theoretical framework.

Table 4.3: Patterns of lead firms in neural production networks.

| | **Global production networks** | **Neural production networks** |
|---|---|---|
| **Empirical focus** | Ability of lead firms to gain control over other actors in production networks by exploiting territorial inequalities<br><br>Inter-firm relations (B2B) | Ability of lead firms to create and sustain infrastructural dependencies by enveloping the space of production activities<br><br>Platform economics and governance |
| **Analytical categories** | Linear trajectories of value creation, value enhancement, and value capture<br><br>Societal, network, and territorial embeddedness of economic activities<br><br>Relations of corporate, collective, and institutional forms of power | Intrafirm cross-subsidisation, facilitated by rentier relations and network effects<br><br>Horizontal and vertical integration of proprietary ecosystems<br><br>Expansion of power by decentralising AI production and centralising AI provision |

In this section, I show that the empirical implications of this reconfigured focus are profound. Established ways to operationalise the three analytical categories of value, embeddedness, and power cannot capture the mechanisms of how technology giants enact AI's geographies of production. Nonetheless, this threefold structure offers a useful roadmap to organise the discussion of results in a logical way.

As a signpost for the following text, Table 4.3 summarises the most important frictions between the empirics of my case study analysis and previous interpretations of how those categories manifest in production geographies. To unearth those frictions, the discussion uses selected excerpts from the case study analysis, rather than descriptively presenting all results. The discussion begins by considering issues of cross-subsidisation and rentier relations, before proceeding to analyse the horizontal and vertical expansion of platform ecosystems. Finally, the third part of the discussion builds on those findings to argue that technology giants consolidate and expand their power as lead firms by decentralising AI production and centralising AI provision.

### *4.4.1 Cross-Subsidisation and Rentier Relations*

Let us return to the example that I mentioned at the beginning of this chapter: the Go-playing AI software that attracted significant media attention. When Lee Sedol – the defeated world champion – announced his retirement from professional play in 2018, AlphaGo had already been succeeded by a more potent version: AlphaGo Zero. As its developers at DeepMind emphasise in a *Nature* paper, unlike its predecessor, this newer version 'is based solely on reinforcement learning[27], without human data, guidance or domain knowledge beyond game rules' (Silver et al. 2017: 354).

While the operations of technology giants are often seen as being black-boxed (Pasquale 2015), especially with regard to their proprietary AI systems, AlphaGo Zero's creators are rather transparent about the computational resources that were used to train the Go-playing models. To demonstrate the efficiency of their newest model, they note: 'AlphaGo Zero used a single machine with 4 tensor processing units (TPUs), whereas AlphaGo was distributed over many machines and used 47 TPUs' (Silver et al. 2017: 356). What may seem like a peripheral technical detail opens up the study of a bigger

[27] Reinforcement learning can be understood as a 'a subset of machine learning that allows an AI-driven system (sometimes referred to as an agent) to learn through trial and error using feedback from its actions' (University of York 2022). Due to its iterative characteristics, this technique is particularly potent in game contexts, such as for playing Go.

phenomenon – one that characterises a key pattern of technology giants as lead firms in neural production networks. Seen this way, AlphaGo is not a fairy tale of superhuman, alien-like intelligence fuelled by mathematical complexity. AlphaGo provides, rather unglamorously, an illuminating demonstration of intrafirm cross-subsidisation. More specifically, Google does not act here as an infrastructure-as-a-service provider to *other* companies – the company instead subsidises a strategic node within its *own* production network: DeepMind's resource-intensive research. How does this work?

Figure 4.4: An aisle of rack-based servers containing TPUs. Source: Google.

In 2016, Google announced that the company developed a proprietary hardware chip to accelerate the training of neural networks that had already been used in its data centres for over a year, Tensor Processing Units (TPUs). As Outeiral (2021) explains, 'unlike GPUs [i.e., graphics processing units, which are used to train artificial neural networks], TPUs have been designed from the ground up for deep learning, and they have been featured in most of DeepMind's recent successes.' Crucially, DeepMind has been a loss-making business in recent years, with operating losses of £477 million in 2019. In the same year, Google waived the repayment of DeepMind's intercompany loans and all accrued interest, which amounted to a total of £1.1 billion (Heller 2020). Most of DeepMind's revenues were generated by 'applying deep reinforcement learning within Google to reduce power costs for cooling its servers' (Heller 2020).

The fact that Google financially backs its AI research lab is not an exceptional phenomenon. After all, cross-subsidisation is a pivotal feature of platform economics (Srincek 2016; Fernandez et al. 2020; Klinge et al. 2022) – that is, using profits in a sector that a digital giant already dominates (e.g., advertising or retail) to finance its entry and expansion into new sectors (e.g., health care or education). But beyond such flows of capital, the rarely discussed aspect that Google also operates as a provider of computational resources for its subsidiary is more relevant for answering this chapter's research question. This new manifestation of cross-subsidisation goes beyond financial support – it occurs on an *infrastructural level*. And this level cannot be understood by exclusively considering financial statements. Dickson (2020) stresses the links between DeepMind's day-to-day research work and Google's TPUs:

> DeepMind's "technical infrastructure" runs mainly on Google's huge cloud services and its special AI processors, the Tensor Processing Unit. DeepMind's main area of research is deep reinforcement learning, which requires access to very expensive compute resources. […] There are no public details to indicate how much Google charges DeepMind for access to its cloud AI services, but Google is most likely renting its TPUs at a discount. This means that without Google's support and backing, the company's expenses would have been much higher.

Therefore, by operating as the infrastructural backbone for its London-based subsidiary, Google defies conventional articulations of value in production network scholarship in favour of a delayed realisation of gains through DeepMind's expected future scientific breakthroughs. Unlike a car manufacturer in a producer-driven network, whose processes of value creation and capture occur consecutively[28] (i.e., the firm must first *create* value before it can *enhance* and *capture* it), I argue that lead firms in neural production networks perform all these processes contemporaneously. As a result of this, it is necessary to present an alternative framing vis-à-vis those linear value trajectories. Building on the groundwork of Birch and Cochrane (2022: 6), I contend that lead firms are orchestrators of rentier relations 'that bring users into the respective enclaves where they can be monetised.' The fruits of this monetisation, in turn, can be cross-distributed to subsidise cost-intensive yet headline-generating research projects such as AlphaGo. Before clarifying the meaning of rentier relations in this context, it is worth taking a step back and digging deeper into the empirical case study material.

How expensive is the provision of computational resources for large-scale AI models? Following the success of AlphaGo and AlphaGo Zero, a third very prominently discussed DeepMind AI software was AlphaFold 2, presented in a *Nature* article in July 2021. The model's purpose diverges from its game-playing namesakes, as it aims to perform predictions of protein structures, also known as the protein folding problem. To optimise the model for this problem, Google granted DeepMind access to 128 TPUs for an undisclosed number of weeks. While the exact cost of training the model was not revealed, Wiggers (2020) calculates that 'Google charges Google Cloud customers $32 per hour per third-generation TPU, which works out to about $688,128 per week.' As training data for the model, DeepMind downloaded 170,000 protein structures from the Protein Data Bank, a freely accessible repository of biological molecules. In doing so, it

[28] This is not a claim about global production networks but an appraisal of the industrial features of physical commodities. The strength of global production networks remains its rejection of conceptualising economic activities as processes of sequential transformation.

benefitted from a database that only exists thanks to 'a large body of research paid almost exclusively by the taxpayers[29]' (Outeiral 2021).

Such blatant forms of privatising the fruits of publicly-funded collective labour cannot be easily identified with respect to Amazon and Microsoft – at least not when it comes to their practices of intrafirm cross-subsidisation. Nonetheless, a comparison between these case studies underlines the importance of proprietary hardware assets within neural production networks. In a similar way as the efficiency of data centre architectures is crucial to ensure the seamlessness of data supply chains (Brodie 2020), specialised hardware is key to optimising the training of artificial neural networks. Here, the term "training" refers to the iterative adjustment and fine-tuning of their layered architecture. The purpose of this is to improve a model's capacity to calculate accurate predictions. The goal is to 'efficiently and optimally set the weights and biases' (Yiu 2019) of individual neurons, which are tied together by different layers within the network (see Figure 2.2, Chapter 2). As Yiu (2019) explains:

> In a neural network, changing the weight of any one connection (or the bias of a neuron) has a reverberating effect across all the other neurons and their activations in the subsequent layers. That's because each neuron in a neural network is like its own little model. […] So, each hidden layer of a neural network is basically a stack of models (each individual neuron in the layer acts like its own model) whose outputs feed into even more models further downstream.

The most important point to be made here is that, given their hundreds of layers, training artificial neural networks by fine-tuning their neural parameters is an extremely computationally-intensive process. It is so resource-demanding that a bourgeoning area

[29] Besides illustrating the intricate relationship between economic and computational potency, this example also typifies Mazzucato's (2018) insistence on the key role of state-funded infrastructures for the value-creating practices of private corporations.

of competition between the technology giants included in this case study analysis has emerged around who owns the most potent hardware to facilitate this process. Echoing Google's TPUs, Amazon also developed its own proprietary AI hardware, called AWS Inferentia. Similar to TPUs, AWS Inferentia is deployed as an infrastructural tool for enhancing services and products within Amazon's corporate ecosystem, such as the company's voice assistant Alexa. At the same time (and this following quote introduces the role of rentier relations), AWS Inferentia 'is designed to provide high-performance inference in the cloud, to drive down the total cost of inference, and *to make it easy for developers to integrate machine learning into their business applications*' (Amazon Web Service 2022a, my emphasis). In other words, Amazon's proprietary AI hardware is by no means restricted to company-internal optimisation and subsidisation processes, and neither are Google's TPUs or Microsoft's specialised AI chips.

Using application programming interfaces (APIs) and software development kits (SDKs), all three lead firms provide gateways for users to build their own AI products and production pipelines *on top of* cloud-based infrastructure offerings. As Helmond (2015: 1) explains, 'the new architectural model of the platform explicitly opens up websites by enabling their programmability with a software interface.' The difference between APIs and SDKs is that the former is an interface for a service (e.g., uploading data to the cloud), whereas the latter is a more comprehensive set of tools and code fragments to build software (e.g., for creating a facial recognition system). Simply said, APIs and SDKs represent the figurative bridges between users and lead firms in neural production networks. Staying with the conceptual metaphor of this thesis, those bridges are the technical underpinnings of computational envelopment. For instance, to facilitate the uptake of AWS Inferentia chips by users, Amazon offers a designated SDK. Prices for those computational resources range from $0.228 per hour to $4.721 per hour, with discounts for multi-year ‘standard reserved instances’, while Amazon’s marketing pitch stresses reduced costs and increased data portability for developers.

Those proprietary infrastructure offerings can be understood as being part of the creation of ‘rentier relations far and wide, at different scales and intensities, while also concentrating control over the system and value captured from the system in a small number of large hands’ (Sadowski 2019: 8). The notion of rentier relations captures that technology giants occupy dominant positions as proprietors of computational services that even some of the world’s most powerful institutions (e.g., the Pentagon, Chapter 1), cannot do without. Birch and Cochrane (2021: 5) use the notion of ‘enclave rents’ to describe this envelopment, defined as revenues that lead firms generate by ‘controlling an ecosystem of devices, apps, platforms, and other products’ and setting and enforcing rules for ‘users, developers, and others.’ Building on this strand of work, I argue that the provision of AI hardware is an understudied expression of those rentier relations and it needs to be situated within broader platform ecosystems.

Before doing that, however, it is useful to underpin this endeavour by providing more detail on the links between technology giants and rentier relations. Within policy discourse, the usual example of such rentier relations is the role of third-party sellers on Amazon Marketplace. On the one hand, Amazon takes a specific commission for each transaction on its marketplace. Sellers can make use of the company’s network effects (i.e., accessing more clients than without Amazon) but have to pay a commission in exchange for gaining that competitive advantage. On the other hand, Amazon can also extract valuable data about those transactions for determining whether to it makes sense to circumvent third-party sellers and directly sell particularly well-performing products (without having to share profits with intermediaries). This structural power asymmetry is so striking that it has been used as a key piece of evidence to justify the necessity of new regulatory frameworks and instruments aimed at curbing Amazon's dual power as both an intermediating market-maker *and* a retail behemoth (Stolton 2022).

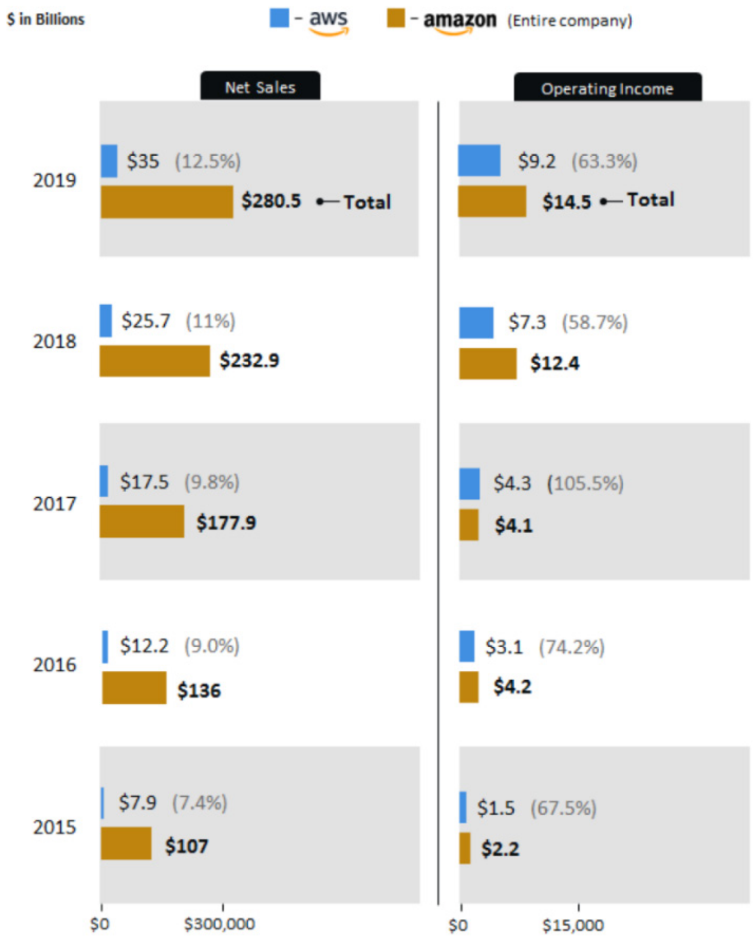
In neural production networks, however, uncovering those rentier relations is severely complicated by the impossibility of gazing backwards along the chains of production processes because distributed arrangements such as Project Maven's neural networks are continually unfolding and never fixed in nature[30] (see Figure 2.1, Chapter 2). To defy these methodological challenges, I argue that it is fruitful to subsume the company-external provision of AI-as-a-service offerings by lead firms as a feature of their wider operations as providers of cloud computing infrastructure. As echoed by a number of scholars (Srincek 2016; Narayan 2022), cloud computing is not only a highly lucrative engine of profit generation for the technology giants in general; it is also a reliable source of cash flows that those lead firms use to cross-subsidise their entry into new industries, sectors, and services (Khan 2016).

Against this backdrop, the analysis of policy documents offers valuable data to empirically substantiate this argument. An investigation by the House Democrats (2020: 379) states that all three technology giants included in this study (Amazon, Google, and Microsoft) are able to 'use supra-competitive profits from the markets they dominate to subsidize their entry into other markets.' Amazon Web Services (AWS), for example, accounted for 63.3% of Amazon's operating income in 2019 (Figure 4.5). In 2021, Google's cloud division accounted for 7.5% of the company's overall revenue, while advertising made up a staggering proportion of 81.3% (Statista 2022). In terms of Microsoft (2022), its cloud segment revenue was 35.4% of the overall revenue in 2021,

[30] Grounded in Whiteheadian process philosophy and relational materialism, this idea features prominently in studies of algorithmic systems, notably Bucher's (2018: 49) analysis algorithms as 'only knowable in their *becoming* as opposed to their *being.'*

with similar percentages for its operating income. In summary, although the domain of cloud computing is relevant across the board when it comes to the lead firms included in this chapter, there are differences in terms of their proportional revenues.

Figure 4.5: Comparison between AWS and Amazon. Source: House Democrats 2020.




The case study analysis also finds that the distribution of value in relation to the provision of cloud infrastructure can lead to puzzling outcomes when governmental institutions are involved. In 2018, Her Majesty's Revenue and Customs, a department of the UK Government that is responsible for the collection of taxes, for instance, spent £11m for commissioning infrastructural services from Amazon Web Services (AWS), ‘which was six times more than the amount it received in corporation tax from AWS’ (Kunert 2021). In 2020, the same government department signed a new ‘£41m two-year deal for hyper-scale compute cloud service provision' (ibid.). Between 2015 and 2019, the UK government awarded AWS a total of '36 public sector contracts worth £660m' (Kunert 2019). Such numbers are intriguing because they show how streams of taxpayer money flow via government departments into the hands of technology giants, despite their year-long efforts to minimise the amount of corporation tax they are paying (Neate 2021). The next section will problematise this aspect in more detail.

To summarise the key argument of this section, lead firms in neural production networks defy conventional articulations of linear value trajectories in global production networks. They do so by cross-subsidising strategic nodes in their ecosystems based on profits generated through rentier relations. At the core of neural production networks is not the creation of a physical commodity but the envelopment of the space within which *other* production activities take place. Instead of producing goods, lead firms own and control storage, processing, and computing capabilities that are needed to optimise the parameters of artificial neural networks. In the next section, I argue that the horizontal and vertical integration of ecosystems is a second important pattern that captures how technology giants enact AI's geographies of production.

#### *4.4.2 Horizontal and Vertical Integration of Ecosystems*

Artificial neural networks are already omnipresent in everyday social, economic, and cultural life. They underlie Google's search engine, Amazon's product recommendation systems, and Microsoft's speech recognition algorithms. However, the expertise to use or develop artificial neural networks is not nearly as distributed as their cross-sectoral and cross-industrial deployment. For the most powerful lead firms in neural production networks, this concentration of computer science expertise and technological know-how goes hand in hand with two major competitive advantages. On the one hand, they are able to attract new scientific staff not only by offering higher salaries than research institutes or publicly-funded universities but also by offering them the opportunity to use superior computational resources such as Google's TPUs and Amazon's Inferentia chips. In recent years, there has thus been a significant 'brain drain of researchers from academia to industry, particularly from elite institutions into technology companies such as Google, Microsoft and Facebook' (Jurowetzki et al. 2021). On the other hand, they may benefit from this asymmetrically distributed computer science expertise by offering AI-as-a-service offerings that promise to radically simplify (and even democratise) the uptake of artificial neural networks across more societal spheres.

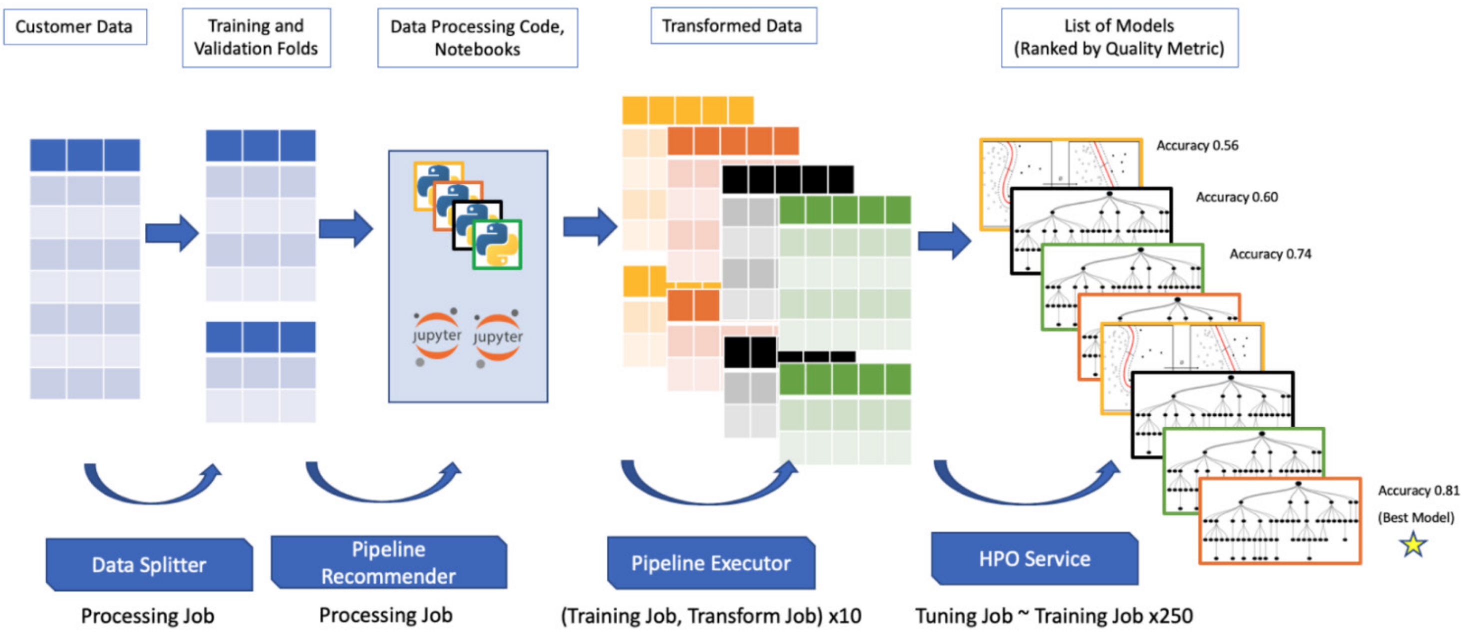
In this section, I argue lead firms have a great strategic interest in encouraging the widespread adoption of artificial neural networks – as long as they are the ones that provide the computational infrastructure on top of which others train their models. That said, I will prioritise the second competitive advantage (i.e., simplified AI-as-a-service products) over the first one due to its stronger explanatory potential for illustrating the horizontal and vertical integration of corporate ecosystems.

All three lead firms included in this case study analysis provide a wide range of infrastructure offerings that gravitate around the notion of 'automated machine learning' (Das et al. 2019), abbreviated by AutoML. In the simplest terms, AutoML applies the foundational logic of machine learning (i.e., making predictions on the basis of pattern recognition) to the development of machine learning itself. Different training datasets and application domains require different models and algorithms, resulting in a myriad of *potential* artificial neural network architectures. How should the model's parameters be fine-tuned? What is the most ideal mathematical weight for each individual neuron? For example, Project Maven's computer vision algorithms aimed at detecting objects in satellite imagery cannot be replicated in models that identify faces and match them with people's identities in police surveillance systems. While the former focus on objects and specialise in lower-resolution drone footage that was taken from a higher distance, the latter aim to identify faces in the higher-resolution camera footage.

With these questions and challenges in mind, AutoML services hold the promise to substitute laborious trial-and-error processes for building the most appropriate neural network architecture and creating well-performing models by automatically testing a multiplicity of potential training pipelines. As Das et al. (2019: 1), a group of engineers at Amazon's AI division, explain in a conference paper:

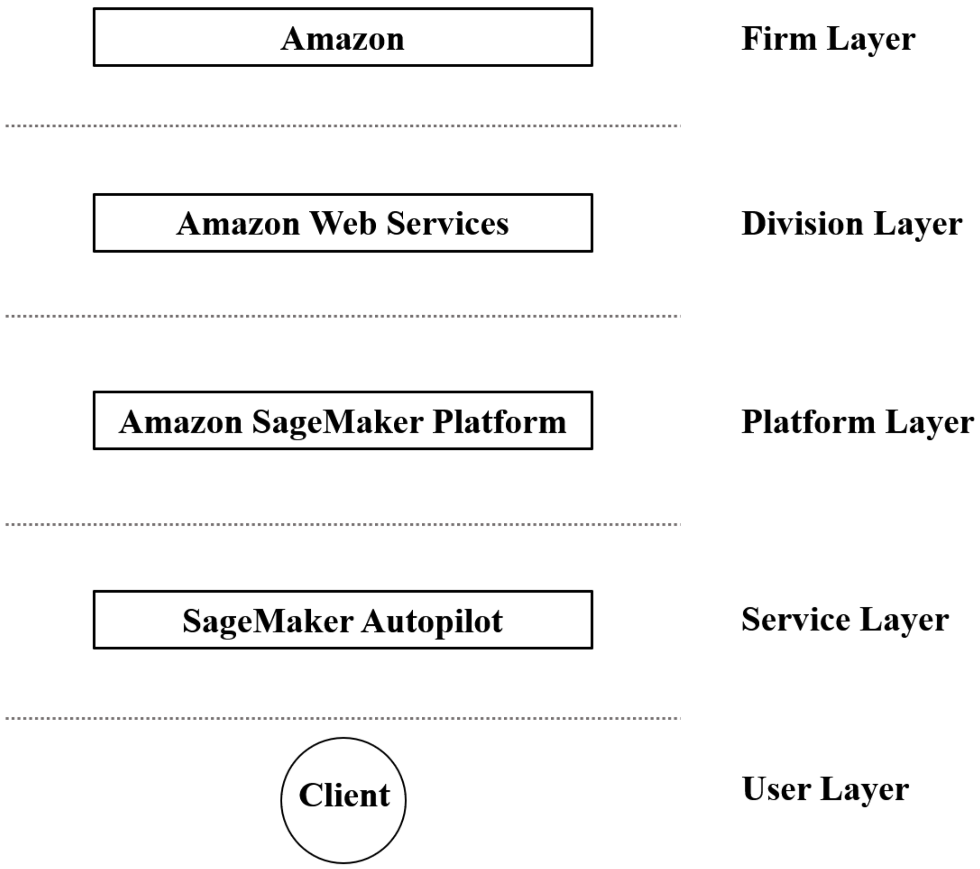
> Today, we have an abundance of advanced algorithms and infrastructure that allow us to build various ML-based smart solutions to benefit business and social needs. […] The philosophy behind Autopilot is to offer a white box AutoML solution, with the vision to democratize machine learning. It aims to make the power for ML accessible to non-experts without requiring to become an expert in this field first.

Figure 4.6: Visualisation of Amazon's AutoML service. Source: Das et al. 2019.




As the Amazon employees explain, there is not only an abundance of particular algorithms to choose from when building a model, but those algorithms also have a long list of parameters 'that need to be set "just right" if you want to squeeze every bit of extra accuracy' (Das et al. 2019: 1). Their product-based solution to this problem, called Amazon SageMaker Autopilot, 'generate[s] a leaderboard of candidate models that the customer can choose from.' They use Figure 4.6 to visualise this process, which starts by processing the training datasets of customers, uses that data to test different neural network architectures to make predictions, and ends by showcasing results. Of course, the reason for providing this level of detail is not to examine AutoML's mathematical assumptions. Rather, I use it as a contextual background for Figure 4.7, which shows how the AutoML service is situated within Amazon's corporate ecosystem.

Figure 4.7: Positioning of AutoML with layered hierarchy of Amazon's ecosystem.




As Figure 4.7 illustrates, SageMaker Autopilot is one of the service features of the machine learning platform, SageMaker, which is a business segment of the cloud computing division, Amazon Web Services. It is worth noting that the 'firm layer' is the only part of this graphic that only includes one entity: Amazon. The other layers refer to multiple things. AI-as-a-service offerings, for example, range from facial recognition to fraud detection. While all those offerings are features of the SageMaker platform, it is not the *only* platform in this ecosystem, as there is also Amazon's marketplace platform for physical commodities and the entertainment platform Prime Video.[31]

Figure 4.8: Topological visualisation of Amazon's neural production network.

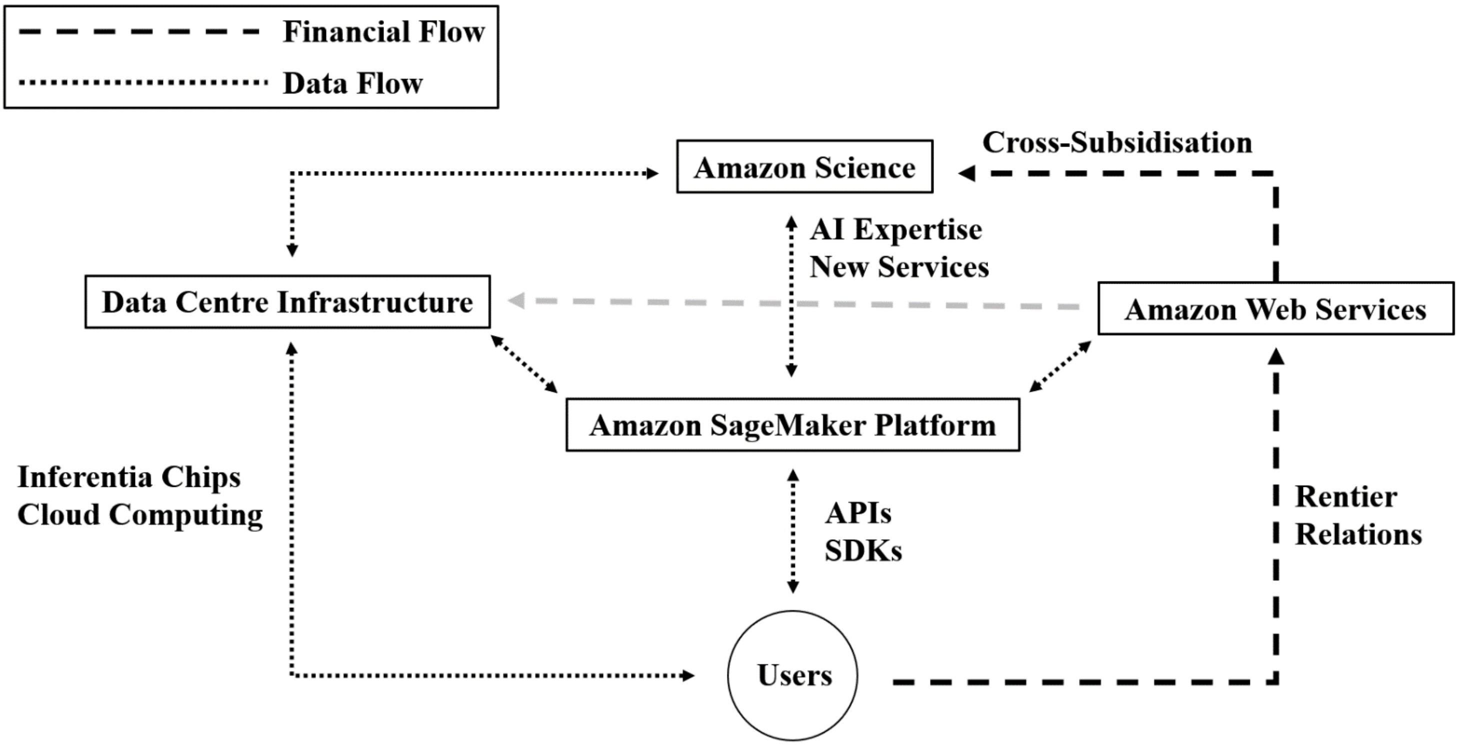
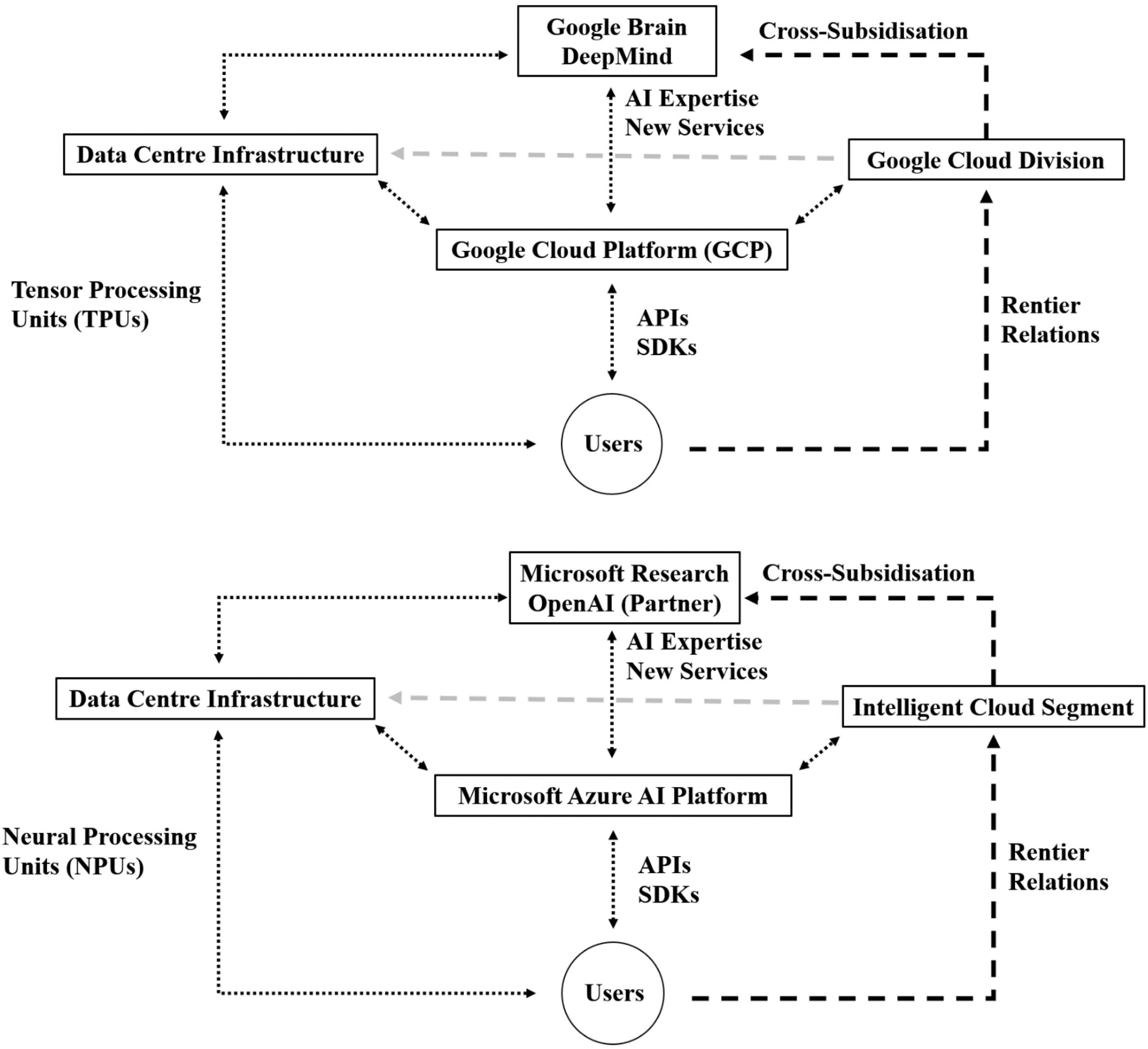



Figure 4.8 provides more detail about the financial flows and data flows within Amazon's neural production network. As highlighted in the previous section, lead firms are able to cross-subsidise both their research and development efforts (e.g., Amazon Science) and their physical data centre infrastructure with profits generated through

[31] Although Figure 4.7 and Figure 4.8 focus on Amazon, Google's and Microsoft's ecosystems also include a firm layer, a division layer, a platform layer, and a service layer. The minor differences are division, platforms, and services names (Figures 4.9). Thus, the structural similarity of lead firms in neural production networks is striking.

rentier relations. The fruits of those research efforts feed into consumer platforms (e.g., SageMaker), which allow users to integrate their datasets and train their models on top of Amazon's computational resources via application programming interfaces (APIs) and software development kits (SDKs). But the crux is that Amazon's AutoML service only integrates with *other services from the same company* in terms of data storage and model deployment components of AI production processes.

Figure 4.9: Visualisations of Google's and Microsoft's neural production networks.



For example, the above-mentioned marketing agency that aims to use AI-as-a-service offerings to analyse social media data could not combine AutoML tools from Amazon and cloud storage from Google or Microsoft (Figure 4.9). While lead firms promote their ambition to 'democratise machine learning' (Das et al. 2019: 1), their actually-existing services to do so are designed to create lock-in effects, whereby users remain enveloped within their proprietary ecosystems. In overshadowing this process with emancipatory discourses of democratisation, lead firms not only obfuscate existing rentier relations but also attempt to foster their cross-sectoral expansion. This is how the abstraction of computational envelopment becomes a material reality.

Before providing more detail on user lock-in effects, it is important to stress that the use of the ecosystem metaphor to describe the operations of technology giants is not uncontroversial. For Dyer-Whiteford et al. (2019: 43), the term ecosystem was 'adopted by those who like to naturalize capitalist activity.' Kenney et al. (2019: 877) write that the notion of ecosystems risks masking 'power dynamics that more closely resemble those for serfs on a feudal manor – always at the mercy of the lord who can expropriate their business without any compensation.' Undoubtedly, such criticism is justified in its rejection of equating capitalism and nature. But I argue that it is possible to postulate that a lead firm in neural production networks metaphorically 'resembles an ecological ecosystem especially because of its complex interdependent parts' (Merriam-Webster 2022c) without naturalising its capitalist modus operandi.

To avoid naturalising such patterns of industry concentration, it is productive to distinguish between the horizontal and vertical integration of those ecosystems.[32] On a

[32] Horizontal integration 'occurs when two companies, within the same industry and at the same stage of production, merge' (Daniels et al. 2016: 497). Vertical integration is 'a process whereby segments of the production process (usually in a vertically integrated business) are subcontracted out to smaller-scale producers' (Daniels et al. 2016: 504). In terms of

horizontal level, the embeddedness of Amazon's AutoML as a part of the SageMaker platform serves to create a 'lock-in to the platform's conception of users, functionality, and design values' (Plantin et al. 2018: 298). Such lock-in effects are key features of the 'infrastructuralization of platforms' (Plantin et al. 2018: 295) and therefore represent the strategic aims of Amazon, Google, and Microsoft 'to be the foundational infrastructure' (Rahman & Thelen 2019: 180) for AI. The qualitative analysis of industry documents is a useful source of evidence, as it can point to pieces of confidential information that were not supposed to be shared with the public. For example, an internal Google Cloud pricing strategy document notes (House Democrats 2020: 246):

> The question that we need to think about is whether we use our entry point with Big Query [i.e., storage service] to get a customer to use all the services such as Data Proc [i.e., analysis service], Data Flow [i.e., processing service], as a suite and give them a price break on the Analytics Suite because it will be much harder for them to migrate away from us if they use all the other services.

Simply put, Google's price experts strategised around how to make it harder for users to exit their proprietary 'walled gardens' (van Dijck 2013: 167) of data services. On a vertical level, such services are also part of a lead firm's arrangement of hardware configurations, compute power, and data storage capacity, exemplifying what Plantin et al. (2018: 295) call the 'platformization of infrastructure.' In 2022, Amazon's data centres were grouped into '84 Availability Zones within 26 geographic regions around the world, with announced plans for 24 more Availability Zones and 8 more AWS Regions in Australia, Canada, India, Israel, New Zealand, Spain, Switzerland, and

technology giants, a range of scholars convincingly combine the use of this distinction with the ecosystem metaphor (Grabher 2020; van Dijck 2020). Importantly, their use of horizontal integration goes beyond mergers of firms in that it refers to the constitution of technical barriers to switch providers, exemplified by AutoML offerings.

United Arab Emirates (Amazon Web Services 2022b). Therefore, to exclusively study Amazon's neural production network as a topological[33] relation of flows (Figure 4.9) would neglect their geographical moments of grounding. AI-as-a-service offerings are not only embedded within closed-off ecosystems but also anchored in particular spatial sites that provide the regulatory contexts for economic action. This anchoring, I argue, is best illustrated by the example of supranational tax laws.

The example of tax laws is helpful to show that lead firms in neural production networks do not make geography obsolete but rather take advantage of it. As indicated in the previous section, it was long speculated that Amazon pays little to no corporation taxes in the countries in which it is operating, putting this topic on the policy agenda of regulators. In 2017, the European Commission ordered Amazon to pay €250m in back taxes to Luxembourg. As Margarethe Vestager, the EU's Competition Commissioner, reasoned back then, 'Amazon was allowed to pay four times less tax than other local companies subject to the same national tax rules, [which is] illegal under EU state aid rules' (Neate 2021). In May 2021, however, this decision was overruled by the General Court of the EU, which decided in favour of Amazon. So, what is Amazon's loophole that immunised the lead firm in this legal procedure?

As Phillips et al. (2021: 5) show, 'Luxembourg is at the centre of Amazon's system of globally coordinated losses that simultaneously generate also unrepatriated profits', with the Luxembourgish set of entities accounting for ~75% of all Amazon's

[33] I take my cue from Neilson and Notley's (2019: 27) term 'production topologies', which they use to enumerate the differences between 'network architectures that structure operations within and between data centres.' The authors define the term topology as 'a distributed, dynamic configuration of practices [that] is organising the forms of social life.' Without a doubt, there could be a whole section (or chapter) about the term topology. I intentionally refrain from providing such an etymological contextualisation. The key point of using it is to go beyond theorising production in relation to discrete horizontal territories.

international sales. When making a case for Amazon's tax evasion, the EU Commission failed to take into consideration Amazon's 'accumulation of losses generated primarily by non-EU subsidiaries to create tax credits (instead of producing profits)' (Phillips et al. 2021: 6). Put differently, the systematic production of losses, instead of the handling of profits, is at the heart of Amazon's transnational tax minimisation strategy. By setting up subsidiaries in multiple jurisdictions, Amazon operates at the conjuncture of multiple geographies (Howson et al. 2021). Although the details of this tax strategy are beyond the scope of this chapter, it is important to note that such approaches are not limited to Amazon. They are also deployed by Google and Microsoft (Neate 2019). To reiterate, there are striking structural similarities between those three lead firms.

Nonetheless, a valid objection in this regard might be: Multinational companies, *by definition*, operate at the conjuncture of multiple geographies. Why are lead firms in neural production networks any different? Within public discourse, technology giants are typically compared to oil companies ('Big Tech' is a reference to 'Big Oil'), so it is useful to take this comparison seriously for a moment. The production networks of oil are a 'metabolism of human and natural production' (Bridge 2008: 415). That is, lead firms rely on elemental modes of production that precede their extractive operations: hydrocarbon processes occurring beneath the earth's surface. Lead firms anticipate and exploit both such biophysical processes as well as regulatory loopholes (Stephenson & Agnew 2016). This example illustrates the problems of classifying the embeddedness of production geographies into territorial and societal categories (Hess 2018). The role of subterranean elemental processes does not neatly fit into these categories, exposing the limitations of *perceivable* manifestations of oil production networks.

Along similar lines, service offerings such as Amazon SageMaker Autopilot and the artificial neural networks that underpin them also point to the shortcomings of such perceivable manifestations of neural production networks. In this respect, a comparison is definitely appropriate. But unlike oil with its fluid yet still tangible materiality, neural production networks do not gravitate around the extraction and commodification[34] of a scarce natural resource. Instead, the *imperceptibility* of neural network architectures is the key factor that complicates empirical efforts t production networks. To stress the complexity of this point, industry documents are insightful sources.

In a comprehensive survey of state-of-the-art AutoML offerings, He et al. (2021) explain that the computational challenge of Neural Architecture Search (NAS) is one of the most important domains in research and practice. The point of NAS is to search for a 'robust and well-performing neural architecture by selecting and combining different basic operations from a predefined search space' (He et al. 2021: 2). The ways in which Amazon, Google, and Microsoft deploy NAS, however, remain proprietary company secrets. As Liang et al. (2019: 401) state, 'Google AutoML system is a black-box that hides the network architecture and training from the user; it only provides an API by which the user can use to query on new inputs.' This opacity is striking as there is only a handful of machine learning techniques that are used to generate neural architectures. Consequently, a leak would provide Google's competitors with strategically important knowledge regarding its computational techniques.

While the practical deployment of sophisticated computational techniques such as Neural Architecture Search by Amazon, Google, and Microsoft remains concealed by trade secrets and non-disclosure agreements (NDAs), their *very existence* complicates any analysis of how those firms act as lead firms. Put differently, those companies do

[34] This is not to imply that commodities are not part of neural production networks, exemplified by the servers, laptops, and devices that make up data centre architectures. Some scholars see the personal data of users as a natural resource extracted by technology giants. I refrain from engaging with those debates as my focus is on infrastructural dependencies.

not only seem to integrate their ecosystems vertically (e.g., hardware stack and data centres) and horizontally (e.g., user lock-in effects), but also algorithmically. In 2017, Google AI researchers were the first to deploy reinforcement learning – the technique that underpinned DeepMind's AlphaGo – for the task of Neural Architecture Search (Zoph & Le 2016). A machine learning approach that is capable of winning against the world's best Go player might thus eventually help our hypothetical marketing company that aims to make predictions from its social media data to inform future campaigns. In short, considerations of ecosystem integration and cross-subsidisation (Figure 4.9) are inextricably related – they are two sides of the same coin.

In summary, this section argued that lead firms in neural production networks do not fully 'overcome the fetter of space' (Dyer-Whiteford et al. 2019: 52). Instead, they envelop the space of users' AI production activities by operating at the conjuncture of multiple geographies and production topologies. By applying proprietary computational techniques in cross-divisional ways, lead firms defy common theories of embeddedness in production geographies (Hess 2018). In the next section, I contend that this capability raises sweeping questions about how they consolidate their power.

### *4.4.3 Decentralised AI Production, Centralised AI Provision*

In terms of how lead firms articulate their power in neural production networks, there is a remarkable tension between two seemingly opposed ends of the spectrum. On the one hand, there is the emancipatory narrative of democratisation that promotes the benefits of AI by virtue of providing access to hardware and software resources for everyone. Google, for instance, introduces its Google.ai platform by highlighting that 'AI can meaningfully improve people's lives and the biggest impact will come when everyone can access it' (Google 2022). Microsoft and Amazon also elevate the promise of access as a cornerstone of their marketing with claims like 'AI will empower every person with knowledge and tools to foster innovation' (Microsoft 2019a).

On the other hand, lead firms boast about the powerful clients that use their paid AI-as-as-service offerings. Amazon Science (2020) promotes that 'NASA is working with the Amazon Machine Learning Solutions Lab[35] to use unsupervised learning and anomaly detection to explore the extreme conditions associated with superstorms.' For this, they use 'simultaneous observations of solar wind drivers and responses in the magnetic fields around earth.' But how can we make sense of this contradiction between providing access to everyone *and* cater the needs of high-profile clients?

Building on the combined insights of the previous two sections, I argue that lead firms consolidate their power by decentralising the production of AI and centralising the provision of AI. Proving the relationship between decentralisation and centralisation is of great relevance for disentangling the double discourse of technology giants as lavish enablers of user empowerment and providers of choice for powerful organisations such as the Pentagon or NASA. To develop this argument, I start by introducing Google's open-source framework TensorFlow as a paradigmatic example of AI's simultaneous decentralisation and centralisation. Subsequently, the focus is on the distinction between the production and provision activities of the three lead firms.

In February 2017, Google released a freely accessible version of its TensorFlow framework, which used to be a proprietary software package. TensorFlow contains a wide range of components and modules, from packages for simplifying the creation of artificial neural networks to integrating these models into Google's cloud computing

[35] With respect to the economic relations between NASA and Amazon Web Services, this superstorm project is only the tip of the iceberg. Amazon also provides the cloud computing infrastructure for NASA's Perseverance Rover mission and supports 'NASA JPL in its mission to capture and share mission-critical images and help to answer key questions about the potential for life on Mars' (Amazon Web Services 2021).

infrastructure. TensorFlow acts as a field of action for producing AI systems, enabling and constraining the agency of developers when using it to build their own models. While my focus is not on Tensor Flow's technical details, it is worth stressing Google's rationale for making the software freely accessible in the first place.

The relevant literature on this topic suggests two main explanations for Google's strategy. Dyer-Whiteford et al. (2019: 55) contend that open-source frameworks like TensorFlow actually 'act as 'on-ramps' to the proprietorial infrastructures of large AI companies', luring in potential clients for the generation of rentier relations. Srincek (2020) adds a slightly different aspect, by arguing that free software like TensorFlow is useful for lead firms insofar as it helps to 'build up a community of developers trained in a company's workflow and tools.' Both claims deserve our attention. Discussed in conjunction with excerpts from my case study analysis, they serve to provide clarity about the tension introduced at the beginning of this chapter.

To start with the first explanation, I argue that at the heart of this mode of reasoning is the problem of *interoperability*. The Oxford Dictionary (2021) defines interoperability as 'the ability of computer systems or software to exchange and make use of information.' As discussed concerning user lock-in effects, the production of AI does not only require staggering amounts of computing power but also interoperability between the three production stages of data preparation, model building, and model deployment. As a result of this, lead firms have an economic interest in consolidating their power by preventing users to switch between providers. And despite its open-source nature and the discursive portrayal through Google as a harbinger of access for everyone, TensorFlow remains embedded in Google's proprietary ecosystem. To use a familiar example, TensorFlow allows its users to train their models on tensor processing units (TPUs), with Google prominently advertising the benefits of making use of this paid service. TensorFlow thus provides a direct pipeline to Google's computational resources, in a similar way as the firm's AI-as-a-service offerings.

Seen this way, tracing the link between open-source tools and neural production networks offers a glimpse behind the shiny façade of democratisation imaginaries in order to consider them as integral features of the consolidation of power. Nevertheless, my case study analysis also finds that this consolidation of power is not hegemonic or all-encompassing, given that users have made use of TensorFlow's affordances in ways not intended by Google's designers. A prominent example to underpin this argument is FakeApp, a user-generated software 'that uses Google's TensorFlow machine learning to morph faces in videos' (Robertson 2018). In 2018, the software generated significant media attention it was used to face-swap celebrity faces onto porn performers' bodies as featured in videos that were circulated on Reddit. Crucially, although much scholarship has mushroomed around such 'deep fakes', the role of TensorFlow as the foundational infrastructure for their initial production was neglected. This case exemplifies a broader question: What do the efforts of lead firms to foster the widespread adoption of AI imply for how power operates in neural production networks?

To answer this question, the second above-mentioned explanation as to why lead firms offer free AI software – the aim of building a community of developers – provides useful hints. By circulating their open-source AI tools as widely as possible, lead firms put themselves in a position in which they can influence economic, social, cultural, and educational outcomes in ways that go beyond the provision of a service in exchange for money. As Gershgorn (2018) explains, there are recursive forces at play as well, given that 'people outside the company find and fix [TensorFlow's] bugs, and students are being taught on the software in undergrad and PhD programs, creating a funnel for new talent that already know the company's internal tools.'[36] Technology giants, in other words, insinuate their force of envelopment into fields and domains that transcend the explanatory scope of the notion of *production*. This manifestation of power destabilises a quintessential assumption of global production networks: the idea that lead firms bring a *clearly definable outcome* into being – irrespective of whether that is an infrastructural service, an extracted natural resource, or a physical commodity.

Another factor that exacerbates the difficulty to delineate the power of lead firms is the domain of synthetic data. Technology giants create synthetic data by producing artificial data from scratch or using advanced data manipulation techniques to produce novel training examples (Nikolenko 2021). Crucially, both strategies require large amounts of computing power. Microsoft's huge investments in the San Francisco-based research lab OpenAI provide a relevant example. In June 2019, OpenAI raised a billion dollars from Microsoft for starting an 'exclusive computing partnership to build new Azure AI supercomputing technologies' (Microsoft 2019b). At the heart of the cooperation is the development of one of the world's fastest supercomputers that runs on top of Microsoft's cloud infrastructure: a 'single system with more than 285,000 CPU cores, 10,000 GPUs and 400 gigabits per second of network connectivity for each GPU serve' (Langston 2020). In terms of practical tasks, Wheeler (2020) quotes from a developer conference at which Microsoft announced that the supercomputer could 'examine huge datasets of code from GitHub (which Microsoft acquired in 2018 for $7.5 billion worth of stock) to artificially generate its own code.'

[36] TensorFlow is just one example of broader patterns of what Perrotta et al. (2020) theorise as the emergence of 'platformised education.' Amazon, Google, and Microsoft provide free courses for users to get started in their cloud environments, thereby habituating workers and students into their proprietary hardware and software infrastructures.

In other words, similar to how Google acts as the infrastructural backbone for DeepMind's research projects, Microsoft fulfils this role for OpenAI. A prominent artificial neural network which has been trained on this supercomputer is GPT-3, a language model that uses deep learning to produce human-like text. In 2020, GPT-3 was promoted to be the 'largest and most advanced language model in the world, clocking in at 175 billion parameters' (Scott 2020), exclusively licenced by Microsoft. OpenAI's (2021) text-to-image model DALL-E was also trained on Microsoft's supercomputer. However, the major difference between those Google and Microsoft is that the latter firm is rather transparent about its strategic ambition to monetise the outcomes of this strategic partnership. In a blog entry, Microsoft implies that the aim of its investments in OpenAI is to set the stage for rentier relations (Langston 2020):

> It's also a first step toward making the next generation of very large AI models and the infrastructure needed to train them available as a platform for other organizations and developers to build upon. […] When you're developing a cloud platform for general use, [Microsoft Chief Technical Officer] Kevin Scott said, it's critical to have projects like the OpenAI supercomputing partnership pushing the cutting edge of performance. He compares it to the automotive industry developing high-tech innovations for Formula 1 race cars that eventually find their way into the sedans and sport utility vehicles that people drive every day.

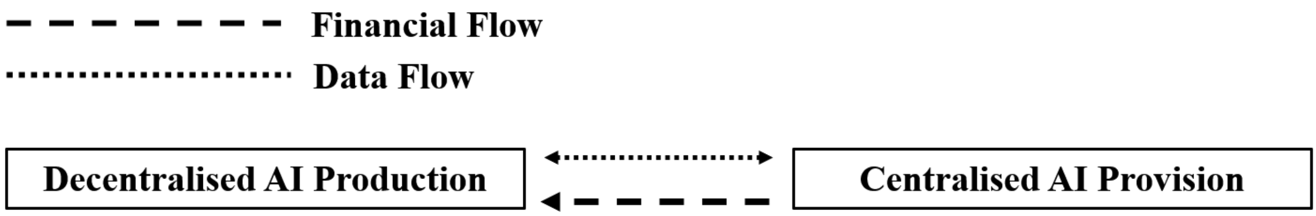
In addition to such economic-geographical considerations, the infrastructural role of supercomputers and proprietary machine learning models such as GPT-3 and DALL-E also poses sweeping questions regarding their cultural ramifications. How can we make sense of the capacity of lead firms in neural production networks to construct synthetic worlds and advanced simulations? How to map the process of production of something that seems to have no beginning – and no end? Chapter 6 will thoroughly engage with these questions. for now, it is important to state that lead firms' modes of consolidating their power are not simply limited to enveloping users. Put differently, the centralisation of tools and computing resources in the hands of a few lead firms is as much a discursive phenomenon as it is a material phenomenon.

Consequently, it is notoriously difficult to delineate the market in which those lead firms actually operate. This analytical vagueness enables Amazon, Google, and Microsoft to downplay their power when they need to. For example, when asked about Amazon's dominance as a provider of cloud computing, the company's UK director of public policy declared during a House of Lords session on internet regulation that 'there is such enormous competition and speed, and there is a difference between dominance and prevalence' (Donnelly 2019). This quote typifies Srnicek's (2020b: 87) argument that 'the inability to theoretically pin down a relevant market has led these companies to routinely argue that they in fact occupy only a small portion of the market.' As a result of this, it is far from straightforward for regulators to find effective ways to push back against the market power of a handful of lead firms.

In neural production networks, this discursive downplaying of power is further complicated by open-source tools such as TensorFlow. When Google's involvement in Project Maven made headline news, a company spokesperson quickly emphasised that 'this specific project is a pilot with the Department of Defence, to provide open source TensorFlow APIs that can assist in object recognition on unclassified data' (Conger 2018). The spokesperson made it sound as if there was no economic incentive for Google to engage in the project, given TensorFlow's open-sourced nature. However, as leaked emails by company executives indicate, 'Google's business development arm expected the military drone artificial intelligence revenue to ramp up from an initial $15 million to an eventual $250 million per year' (Fang 2018) – an opportunity that Amazon and Microsoft were also keen to secure. After protests by Google employees had forced the company to cancel the contract, Microsoft released a statement stressing its support of the Pentagon– and ended up winning a $10 billion cloud computing contract in 2019, thereby outcompeting its primary rival Amazon (Simonite 2019).

Just nine months after Amazon's UK public policy director downplayed the company's dominance, an Amazon US spokesperson said to *Wired:* 'We're surprised about this conclusion [i.e., Microsoft winning the contract]. *Amazon Web Services is the clear leader in cloud computing*, and an assessment purely on the comparative offerings clearly led to a different conclusion' (Simonite 2019, my emphasis). This case unveils two insights. First, the operations of lead firms in neural production networks as AI-as-a-service providers are directly linked to their dominance as cloud computing providers. Second, the ways in which lead firms articulate their own power are entirely dependent on whether it is beneficial to tactically downplay it or not.

Figure 4.10: Links between decentralised AI production and centralised AI provision.




Consequently, their double discourse as providers of open-source software and infrastructural backbones of clients like NASA and the Pentagon appears in a new light. When articulating their infrastructural role, lead firms have a strategic interest in putting a discursive emphasis on the left side of Figure 4.10 (i.e., decentralised production of AI), while downplaying the right side (i.e., centralised provision of AI). Instead of marking two opposed ends of a spectrum, there is a recursive relationship between those practices. For example, without receiving valuable bug fixes from thousands of users due to TensorFlow's open-source nature, Google might not have been in a position to improve its software to the extent that it could eventually underpin Project Maven's compute vision models. As a traditional case of platform-based network effects, if more clients make use of lead firms' paid services, more profit can be used to cross-subsidise their unpaid AI-as-a-service offerings: a virtuous circle.

### 4.5 Conclusion: A Distinct Form of Platformised Industrial Organisation

> Are these platforms skimming rent off capital and labour? Or do they represent a fundamental shift in economics, a new Industrial Revolution? […] Companies such as Alphabet, with a market cap in the neighbourhood of three quarters of a trillion dollars, have claimed to be neutral arbiters and spaces of informational exchange. No one really believes that anymore, but we lack language to grasp the way these platforms collapse profit and the social, culture and capital.

Writing in the *Los Angeles Review of Books*, Weatherby (2018) contends that there is a need for new conceptual vocabulary to account for the role of platforms as paradigm-shifting forms of industrial organisation. In this chapter, I respond to this call for theory building by elucidating the patterns of a distinct manifestation of platformised industrial organisation: the emergence of neural production networks. Drawing on a paradigmatic case study analysis, the findings of this chapter contribute to research at the intersection of infrastructure studies (Plantin et al. 2018), platform studies (van Dijck et al. 2018), and economic geography (Coe & Yang 2022; Grabher 2020).

On the one hand, the chapter points to three major structural similarities between how technology giants enact AI's production geographies. First, unlike lead firms in global production networks, the operations of Amazon, Google, and Microsoft cannot be adequately comprehended by relying on a linear understanding of value creation, enhancement, and capture as categories to operationalise the distribution of value (Hess 2018). Instead, it is necessary to analyse their operations through the lens of always-unfolding rentier relations (Sadowski 2019). The scarcity of computational resources that are designed to optimise the parameters of artificial neural networks, such as

Google's TPUs and Amazon's Inferentia chips, is a key driving force of those rentier relations. As long as a handful of lead firms maintain exclusive ownership and control over such infrastructural technologies, the proliferation of artificial neural networks will amplify, rather than curb, tendencies of industry concentration.

Second, given that at the heart of neural production networks is not the creation of a physical commodity but rather the creation of infrastructural dependencies, a focus on how lead firms are territorially embedded is insufficient. As opposed to commodities whose journeys can be tracked by gazing backwards along the production process, the geographies of AI-as-a-service offerings cannot be methodologically related to discrete places (i.e., clearly identifiable sites of production). Amazon, Google, and Microsoft operate at the conjuncture of multiple geographies and production topologies, thereby integrating their globe-spanning data centre infrastructure with a proprietary ecosystem of platforms and services. Remarkably, there are striking structural similarities in how this integration occurs, illustrated by the role of application programming interfaces (APIs) as central points of data exchange between lead firms and users.

Third, although there seem to be a number of contradictions between discourses of democratisation and commercial interests, those contradictions vanish if we consider that technology giants have an interest in fostering a decentralisation of AI production. The ideological umbrella of decentralisation is a convenient and powerful rhetorical tool for those companies to immunise themselves against regulators. Is it powerful and appealing because it *reflects* a simple fact: the barriers for users to access AI resources (e.g., TensorFlow) are extremely low. But this rhetorical tool also *hides* a more complex fact: the barriers for other firms to compete with Google, Amazon, and Microsoft when it comes to their AI resources are extremely high. In other words, matters of lead firms' material and discursive dominance are closely related.

On the other hand, the chapter also exposes significant differences in how those lead firms operate, especially with regard to their modes of cross-subsidisation. While Google invests both high amounts of money and infrastructure resources in its London-based research and development subsidiary DeepMind, Microsoft prefers a strategic partnership approach with OpenAI. Both lead firms fund headline-generating research projects that showcase the capabilities of AI, such as AlphaGo and GPT-3, and thereby foster the widespread adoption of artificial neural networks. In this respect, Amazon is the outlier, as its research division Amazon Science is primarily focused on informing new products and services, exemplified by AutoML. For future research, an auspicious task is to dig deeper into how those differences play out.

Arguably, each of those structural similarities and differences could form a basis for a self-contained thesis chapter. Given that any case study analysis involves a trade-off between scope and depth (Flyvbjerg 2011), decisions had to be made about what to include and what to exclude from the discussion of findings. In light of the research question framing this chapter (*How do technology giants act as lead firms within AI's geographies of production?)*, I decided in favour of problematising broader patterns of industrial organisation, rather than providing an in-depth analysis of *one* AI-as-a-service offering provided by *one* firm. Although such an analysis would be fruitful, it would have required a different methodological approach (see Chapter 3). What is more, in its aim to provide a strong foundation for answering the remaining two research questions, I argue that the chapter benefits from an expanded analytical scope.

Of course, this decision goes hand in hand with methodological limitations. For example, it could be argued that the explanatory power of this approach remains limited because the actual strategies of lead firms are black-boxed as proprietary trade secrets. In response to this, I would argue that pointing out the *existence* of such black boxes is a relevant empirical finding. Google, for example, praises DeepMind's breakthroughs but does not provide details about how much it charges its subsidiary for training its models using the firm's computational resources. Put differently, the finding that specific pieces of information remain hidden from the public's eye is a strong impetus to theoretically scrutinise why that is the case – and ask follow-up questions about what this strategic obfuscation implies for the efficacy of regulatory responses.

Ultimately, the chapter suggests that the operations of technology giants as lead firms in neural production networks can be taken as an entry point to understanding the economic-geographical processes that underpin AI's asymmetrical political economy (Dyer-Whiteford et al. 2019; Srincek 2020a; Luitse & Denkena 2021; Steinhoff 2022). Much of the research on this topic relies on the hypothesis that patterns of industry consolidation primarily stem from the large-scale extraction and accumulation of user data (Prainsack 2020). This focus on the role of user data and privacy violations is at the heart of regulatory proposals worldwide. But the findings of this chapter imply that the sophistication of their proprietary computational infrastructure is the most important competitive advantage for Amazon, Google, and Microsoft. By leveraging their large-scale computational resources as both a gateway to facilitate company-internal research and a gatekeeping tool to create user lock-ins, lead firms consolidate their dominance in ways that go beyond what can be captured with metrics of market power.

What follows from this argument is that it would be short-sighted to restrict the analysis of regulatory responses by equating industry concentration with market power. By equating economic strength with market power, regulatory debates risk tapping into the pitfalls of corporate lobbying due to the notorious difficulty to delineate the market in which lead firms in neural production networks operate. In the next chapter, I engage with such regulatory considerations in more detail.

# Chapter 5: The Regulation of Neural Production Networks

## 5.1 Introduction: The Dilemma of Digital Sovereignty

> It is time to start thinking of the biggest technology companies as similar to states. These companies exercise a form of sovereignty over a rapidly expanding realm that extends beyond the reach of regulators: digital space. They bring resources to geopolitical competition but face constraints on their power to act.

Ian Bremmer's (2021) analysis of the geopolitics of technology in *Foreign Affairs* is representative of an argument that has gained immense traction within academic and policy circles: the notion that a handful of American and Chinese digital giants have seized control over spheres of society, economics, and national security that used to be the sole purview of the state. Those companies, Bremmer writes, are now more than just corporate behemoths; they have become state-like entities whose territories are neither confined by national borders, nor by legal jurisdictions. And unlike what Bremmer calls 'physical space', the 'digital space' that they are dominating supposedly sits outside the influence sphere of regulators. Leveraging a rhetorical instrument of Cartesian dualism, Bremmer divides the geopolitical field of action into two arenas of sovereignty: a well-known physical space versus a far-flung but omnipresent digital space.

The use of spatial metaphors to articulate the power relations between states and technology giants does not only characterise contemporary geopolitical assessments but also policy agendas with real effects on the geographies of production. In the politics of the European Union, the imaginary of digital sovereignty has become so ubiquitous that it serves as an umbrella term for a wide range of emerging policy initiatives, including the Digital Markets Act and the Artificial Intelligence Act. In this context, the notion of digital sovereignty typically refers to 'Europe's ability to act independently in the digital world' – an objective that includes 'protective mechanisms and offensive tools to foster digital innovation (including in cooperation with non-EU companies)' (EPRS 2020: 1). According to Ursula von der Leyen (2020), President of the European Commission, the domain of artificial intelligence is a key strategic frontier as part of the political quest to achieve European digital sovereignty. In 2020, however, von der Leyen made clear that the Commission's 'aim is not more regulation, but practical safeguards, accountability and the possibility of human intervention in case of danger or disputes.' At first glance, this juxtaposition appears contradictory. How can there be new protective mechanisms *without* more regulation? Or does the term regulation have such a negative connotation that it needs to be hidden under the aura of innovation and sovereignty?

In this chapter, I examine the EU's supranational policy proposals that fall under the umbrella concept of digital sovereignty to understand the positioning of state actors within neural production networks. The research question framing this chapter is: *How does state action shape, and is shaped by, AI's geographies of production?* The chapter answers this question by unearthing the tensions between four overlapping state roles in neural production networks: regulator, facilitator, buyer, and producer. I identify two sets of tensions between those roles in light of their relations to AI's geographies of production. First, the EU's scope of interventions as a regulator is constrained by its role as a facilitator. Second, the EU's shortcomings in acting as a producer exacerbate its dependency as a buyer. Those empirical findings indicate that there are complex reciprocal links between the policy repertoire of state actors and AI's geographies of production. Appreciating the fact that there is an ambiguous multiplicity of state roles is a prerequisite not only for articulating potential points of leverage of state action vis-à-vis technology giants but also for grasping its limitations.

On a theoretical level, I shall unfold this argument by extending the conceptual groundwork of the previous chapters in operationalising AI's geographies of production as neural production networks: that is, *geographically dispersed but computationally enveloped production arrangements powered by artificial neural networks*. In this context, the contribution of this chapter is to expand the scope of this framework by using it to conduct a theoretically-grounded analysis of an important supranational state entity: the EU. This is a valuable contribution to previous research on how state actors are positioned in the geographies of production (Horner 2017; Horner & Alford 2019; Alford & Phillips 2018). This work is aimed at disentangling 'the concrete practices, power dynamics, and organizational forms that give character and structure to cross-border business networks' (Ponte & Sturgeon 2014: 200). However, with a few notable exceptions (Coe & Yang 2022), scholars have not considered the challenges and issues brought about by technology giants as lead firms. As such, a key aim of this chapter is to probe how the nature of state action in neural production networks conforms to, and breaks with, more commonly studied sectors in economic geography.[37]

Empirically, I pursue this objective by problematising the relationship between the operations of Amazon, Google, and Microsoft as lead firms in neural production products (Chapter 4) and the European Commission's Digital Single Market Strategy. In particular, I focus on the analysis of two legislative proposals: the Digital Markets Act and the Artificial Intelligence Act. Those proposals have been analysed by legal scholars and political scientists, typically in relation to the notion of digital sovereignty (Cini & Czulno 2022). But although the idea of digital sovereignty is an inherently spatial metaphor, there is a lack of geographically sensitive discussions on this topic. In this regard, the chapter does not find evidence for an extra-territorial digital space 'that

[37] For example, such sectors include agricultural commodities, textiles, and automobiles (Horner & Alford 2019). More recently, however, a range of interventions have argued to broaden the scope of what counts as a global production network (Howson et al. 2020).

extends beyond the reach of regulators' (Bremmer 2021: 113). Instead of evading the influence sphere of EU regulators, technology giants rather bring about a transformation in how this influence sphere presents itself. Far from making state action obsolete, AI's geographies of production throw into relief how the tensions between state roles as a regulator, facilitator, buyer, and producer matter in profoundly new ways.

To illuminate those ways, the chapter proceeds as follows. First, I assess existing literature on the role and place of state action in global production networks. Given that the majority of this work is grounded in the analysis of physical commodities, I point to the strengths and pitfalls of resorting to this body of literature to situate my enquiry. At the heart of the chapter is the empirical analysis of how the EU's legislative work under the umbrella of digital sovereignty shapes, and is shaped by, the geographies of neural production networks. This section begins by laying out a descriptive overview of the case study and providing a rationale for why it is an important paradigmatic object of analysis. Subsequently, I outline two intertwined sets of tensions between state roles that illustrate what I propose to call the "dilemma of digital sovereignty": *regulator vs. facilitator* and *producer vs. buyer*. The chapter concludes by reflecting on the broader implications of this dilemma beyond the case study of the EU.

## 5.2 State Roles in Global Production Networks

In recent years, a considerable amount of literature has been published on the role of the state in global production networks.[38] This includes theoretical contributions on how to conceptualise the state and its roles (Smith 2014; Horner 2017; Mayer & Phillips 2017)

[38] On a general level, a common denominator of this body of work is a relational understanding of state governance, which appreciates that the repertoire of state actors is enabled and constrained by both a) issues of political contestation *and* b) the economic conditions of global production networks in which they are operating (Werner 2020).

and a wide range of empirical case studies (Lim 2018; Hughes et al. 2017). Despite the mushrooming of this state-oriented research, Werner (2020: 2) notes that, compared to the analysis of corporate power, 'much less attention has been paid to how politics and the state determine these dynamics and, in turn are reshaped by them.' As such, the purpose of this section is to provide an overview of how the existing literature deals with the theoretical and empirical manifestations of state governance to provide a basis for the analysis of its ambivalent role in neural production networks.

In providing a helpful typology (Table 5.1) to schematise the different functions of state actors in global production networks, Horner (2017) differentiates between four state roles: facilitator, regulator, producer, and buyer. These roles, however, are not mutually exclusive; they can be intertwined. As Horner (2017: 6) puts it:

> States may adopt these roles in various combinations as they seek to take control of, or influence, production networks, based on considerations that may go beyond capturing greater economic value. These state roles are shaped by interactions with domestic and foreign firms, business associations, civil society, and even other states and supranational institutions. […] The roles can be interrelated as, for example, many states have played an enhanced enabling role in economic globalization by reducing the earlier regulatory capability.

Table 5.1. State roles in global production networks. Source: Horner 2017.

| Role | Definition | Examples |
|---|---|---|
| Facilitator | Assisting firms in GPNs in relation to the challenges of the global economy | Tax incentives, subsidies, export processing zones, incentives for R&D, implementing and negotiating favourable trade policies, and interstate lobbying |
| Regulator | Measures that limit and restrict the activities of firms within GPNs | State marketing boards, price controls, restrictions on foreign investment, trade policy (tariffs and quotas), patent laws, labour regulation, quality controls, and standards implementation |
| Producer | State-owned firms, which compete for market share with other firms within GPNs | State-owned companies, for example, in oil and mining. |
| Buyer | State purchases output of a firm | Public procurement, for example, of military equipment and pharmaceuticals. |

*Note.* GPN = global production network; R&D = research and development.

In the remainder of this chapter, I argue that while Horner's typology is a useful basis to underpin the analysis of my empirical case material, the latter requires a deeper focus on the frictions between the different state roles. In relation to AI's geographies of production, facilitator, regulator, producer, and buyer roles are not only intertwined but rather in opposition to each other. Before clarifying this point, it is worth taking a step back to outline how those four state roles have been operationalised.

### *5.2.1 Facilitating Markets, Regulating Outcomes*

States, rather than being passive actors, are 'intentional architects' (Mayer & Phillips 2017: 135) in establishing the environments within which global production networks operate. Horner (2017: 7) relates the facilitative role to state-led activities that 'promote, attract, and retain private investment, particularly that which may be footloose and has considerable degree of locational choice as well as support local actors in order to participate in these chains and networks.' As common examples, he mentions trade agreements, intellectual property rights, export processing zones, incentives like tax breaks, and subsidies in key sectors such as pharmaceutical and agricultural industries. According to Mayer and Phillips (2017: 141), an impetus for such facilitative policies is a 'growing tolerance for market concentration and the diminution of competition policy at national and international levels.' In other words, states are more inclined to tolerate higher degrees of industry consolidation as long as some of the value captured by lead firms remains in their territory and they can benefit from it.

In terms of the digital economy, a clear example of the facilitator role can be found in the low corporate tax rate in Ireland, which has attracted more than 800 US companies employing a total of 180,000 people (Lyons 2021), including the European subsidiaries of some of the biggest technology firms. Ireland's tax incentives include generous tax credits to encourage investments in research and development, which is particularly attractive for technology companies. Although it remains controversial if the boosting of employment and the Irish economy or the generation of corporate tax is more important in this context, this example illustrates Horner's spotlight on *footloose* production networks that can, in theory, be grounded in any location. Historically, such policies have also been associated with the aftermath of the global financial crisis post-2008, with the state's role as a global production network facilitator shifting 'from the academic margins to the policy center' (Werner 2020: 3).

Beyond incentivising foreign companies, state facilitation can also refer to the strengthening of domestic lead firms in light of the challenges of the global economy. In their analysis of how the technology giant Tencent – the world's largest video game vendor – is shaped by China's regulatory context, Coe and Yang (2022: 317) argue that the company benefits from a combination between facilitative censorship policies and relatively loose competition regulation. Foreign game developers, for instance, are 'restricted from entering China's market directly and forced to cooperate with local publishers to distribute their games in China' (Coe & Yang 2022: 317). Consequently, Tencent consolidates its market power not only as a *producer* of games but also as a *distributor* of games that were produced elsewhere.[39] This leads us to probe how states may act as regulators in global production networks.

The regulatory role encompasses a state's activities of 'limiting and restricting economic activity within its boundaries' in order to 'protect various societal interests' (Horner 2017: 7). By taking into consideration the measures of states as rule-setters, this role goes beyond perceiving states exclusively as facilitators that shape the geographies

[39] For analysing the positioning of state action in neural production networks, this is as a crucial example. It illustrates the pivotal analytical distinction between *value creation* (as a result of a lead firm acting as a producer) and *value extraction* (as a result of a lead firm acting as a service provider). I problematise this relationship in section 5.4.2.

of production through financial and fiscal incentives. Examples include international trade policies, labour regulations, and environmental standards. Yet importantly, such regulatory regimes and states' enforcement capabilities are unevenly distributed across the world, especially when it comes to ensuring workers' rights and safety. As Mayer and Phillips (2017: 143) rightly put it, 'a great deal of production in contemporary value chains, particularly low-wage, labour-intensive work in agriculture, garments and other sectors takes place beyond the reach of regulatory coverage.' Furthermore, Mayer and Phillips (2017) add that a number of national and supranational regulatory arrangements were privatised, exacerbating the uneven distribution of regulatory leverage.

Although matters of regulatory evasion are also key concerns of the bourgeoning body of interdisciplinary work on platform governance (Gorwa 2019; van Dijck et al. 2018), this research is typically detached from writing on global production networks. I aim to bridge this disconnect since Horner's (2017) typology of state roles is a helpful basis for analysing the state-platform nexus at national and supranational levels. Noting the role of the European Commission as a prototypical entity in platform regulation, van Dijck et al. (2018: 157) state that 'regulation at the supranational level has proven to be most effective with regard to antitrust and privacy protection.' At the same time, they point to two significant hurdles that complicate the efficacy of platform regulation. On the one hand, states lack a nuanced vocabulary to accommodate the strategies by which technology giants aim to consolidate their cross-sectoral dominance, such as algorithmic personalisation and vertical integration (van Dijck et al. 2018: 158). On the other hand, enforcement agencies are ill-equipped for checking that those companies comply with regulatory fixes. Therefore, the complexity of proprietary platform architectures and algorithms as regulatory objects (Seyfert 2018) poses not only political and economic but also veritable *technological* challenges for state actors.

Those hurdles are especially noteworthy since a key concern in the literature relates to 'how the regulator role can be adopted to shape the distribution of rents or gains' (Horner & Alford 2019: 11) within global production networks. Put differently, what is of interest here is how regulatory frameworks influence distributional outcomes and inequalities brought into being by the operations of lead firms. The analysis of the EU's Digital Markets Act and the Artificial Intelligence Act fits neatly into this context, as it responds to Horner and Alford's (2019) call for more research on the relationship between geographies of production and digital policy. Before proceeding, however, we need to consider the relation between producer and buyer roles.

### *5.2.2 Producing Capabilities, Buying Resources*

While the producer role considers the establishment of state-owned enterprises with which states 'take control of productive capacity in key strategic sectors (e.g., security and national resources)' (Horner 2017: 8), the buyer role refers to instances of public procurement from domestic or foreign corporations. Horner and Alford (2019: 12) cite estimates that state-owned businesses make up between 5 and 10 per cent of overall economic activity in the OECD region, with higher proportions in emerging economies. The average proportion of public procurement is between 11 and 14 per cent of the GDP, according to a 2017 study across 89 countries (ibid.: 14). Despite their relevance, there is a dearth of empirical case studies on how producer and buyer roles manifest in global production networks. Werner (2020: 4) provides a plausible reason for this gap, stating that most analyses 'have tended to focus on consumer goods as opposed to strategic sectors such as energy, infrastructure and defence' (i.e., sectors in which state ownership and public procurement practices are more common). Undoubtedly, artificial intelligence is a cross-sectoral issue that is of strategic relevance to governments around the world, illustrated by the mushrooming of national and supranational AI strategies

(Cihon et al. 2020; Smuha 2020). Therefore, it is worth showcasing how scholars have empirically studied the state's buyer and producer roles in other sectors.

In terms of the state-as-buyer role, Hughes et al. (2019) investigate ethical purchasing practices in the UK public sector. They note that legal modifications to UK public procurement laws, which were adjusted in 2014 to account for sustainability and social factors more clearly as part of the contract bidding process, have accelerated the adoption of ethical codes for supply chains in procurement strategies. Nonetheless, they find that practices of ethical sourcing are even less advanced in the UK public sector than they are in consumer goods sectors, given the 'low profile and hidden nature of so many of the materials used in public services, which provides less impetus for the public to trace their origins and biographies' (Hughes et al. 2019: 12). In other words, compared to labour rights violations of well-known companies such as fashion brands, consumer pressure is less of an issue in public sector procurement because most of these contracts naturally remain hidden from the public eye. This structural obfuscation is particularly significant in the context of neural production networks, not least since artificial intelligence is procured not as a physical good that can be traced and localized, but rather as a digitally mediated infrastructure service.

With respect to the state-as-producer role, Lim (2018) presents one of the few production network studies on how state-owned enterprises influence the geographies of production. Empirically, he analyses the acquisition of a Canadian energy company by a Chinese state-owned offshore oil and gas company. According to Lim, China's rationale for acquiring a Western energy company was not simply the creation of economic value but rather the consolidation of geopolitical power by gaining technological know-how and ensuring domestic energy supply. His ability to relate empirical transformations in production arrangements to broader geopolitical considerations underscores the unique analytical strength of the global production network framework for the purposes of this thesis. At the same time, it becomes clear that producer and buyer roles can be closely interrelated, given that 'the state-as-buyer role is more often used to shore up domestic producers' (Werner 2020: 5), including state-owned or subsidised enterprises. In short, there may be less necessity for procurement from *extraterritorial* suppliers if *domestic* producers can provide sufficient supply for a strategically relevant service or product – a key dimension of discourses about European digital sovereignty.

To sum up the analysis of state roles in global production networks, the relevant literature usefully underlines that states do not affect the geographies of production in *one* particular deterministic or unidirectional way. Instead, there is a multiplicity of state roles whose empirical manifestations are highly context-dependent and may vary across sectoral and national contexts. However, although scholars have begun to stress how the frictions and contradictions between these four roles relate to the uneven distributional outcomes (Alford and Phillips 2018), there is a tremendous need for more research in this area, especially with respect to digital policy and artificial intelligence. Building on this conceptual groundwork, the next section begins by offering a descriptive overview of the contextual background of the paradigmatic case study. Subsequently, the chapter examines the EU's Digital Single Market Strategy to understand the role of state actors and state action in neural production networks.

### 5.3 Overview of Case Study: The EU's Digital Single Market Strategy

In July 2014, the newly elected President of the European Commission, Jean-Claude Juncker, presented his political programme to members of the European Parliament in Strasbourg. In this speech, Juncker (2014) put an emphasis on his intention to set in motion 'ambitious legislative steps towards a connected Digital Single Market' as one of his strategic priorities during the first six months of his mandate:

> I believe that we must make much better use of the great opportunities offered by digital technologies, which know no borders. To do so, we will need to have the courage to break down national silos in telecoms regulation, in copyright and data protection legislation, in the management of radio waves and in the application of competition law. […] We can create a fair level playing field where all companies offering their goods or services in the European Union are subject to the same data protection and consumer rules, regardless of where their server is based.

Geographically speaking, Juncker's last sentence is crucial: the fact that a server or a data center might not be located *within* the territorial borders of the EU does not immunise a company from regulatory action *by* the EU. Nearly one year later, Juncker's ambitious vision was moulded into a comprehensive strategy document, which defined the Digital Single Market as one in which 'individuals and businesses can seamlessly access and exercise online activities *under conditions of fair competition*[40], and a high level of consumer and personal data protection, irrespective of their nationality or place of residence (European Commission 2015, my emphasis). In the course of Juncker's presidency between 2014 and 2019, the EU approved 28 legal acts concerning the facilitation of the Digital Single Market (Cini & Czulno 2022).

Against this backdrop, it is important to start with the point that contemporary regulatory developments with respect to AI and platform regulation put forward by the European Commission do not emerge out of a political vacuum. Rather, the European Commission's Digital Markets Act and the Artificial Intelligence Act are embedded in the broader strategy of bringing a Digital Single Market into being. At the same time, it is noteworthy that the history of using the notion of sovereignty to justify the necessity

[40] At first glance, this spotlight on the necessity of establishing fair competition seems to be in contrast with the argument developed by Mayer and Phillips (2017) about the link between competition policy and the state role's as a market facilitator. As the following sections of this chapter demonstrate, this relationship is key for answering my research question.

of establishing such a market goes back much longer than Juncker's speech. Schmitz and Seidl's (2022) analysis of the EU's trade and digital policies usefully pointed me to a 1979 report by the European Commission that is entitled *European Society faced with the Challenge of New Information Technologies: A Community Response*. More than four decades ago, the report discussed the societal impacts of information technologies in the following way (European Commission 1979: 1-2):

> The speed and skill with which these new technologies are developed are critical to the social development of any modern community, to the efficiency and productivity of its industry and services, and, not least, to its position and influence in the world. […] The cost of communicating is being dramatically cut by the introduction into telecommunications of digital, electronic techniques (computing methods and technology) and new transmission techniques (satellites, optical fibres). The cost and availability of artificial intelligence […] enables the power of a computer condensed into a single chip. […] It means that distributed intelligence in terminals, computers, or the family television set can be cheaply linked in European and world-wide networks of great power.

This quote illustrates that geopolitical concerns about the ability of Europe to act independently in the digital world are by no means new. Neither is the appraisal that artificial intelligence is 'ubiquitous, distributed by a network infrastructure, just as electricity' (Dyer-Whiteford et al. 2019: 62). Remarkably, the quote already highlights the integration of AI into globe-spanning networks. What is new and henceforth worth specifying, however, are the regulatory tools that the European Commission developed in order to realise its long-term goal of technological independence by creating a Digital Single Market. While it is beyond the scope of this chapter to provide a comprehensive account of how the EU's policy institutions and decision-making processes operate, the most important aspect to stress here is that the European Commission has a monopoly on preparing, planning, and proposing new legislation. Before proceeding with the case study analysis, the key features of two particular legislative proposals need to be briefly outlined as they provide essential contextual background for the discussion of findings: the Digital Markets Act and the Artificial Intelligence Act.

### *5.3.1 The Digital Markets Act*

The Digital Markets Act aims to introduce 'rules for platforms that act as "gatekeepers" in the digital sector' by 'preventing gatekeepers from imposing unfair conditions on businesses and end users and at ensuring the openness of important digital services' (European Commission 2022). This description entails the most important defining feature of the Digital Markets Act: it only applies to *particular* technology companies, not a whole sector or an industry. Specifically, it only applies to firms that are formally identified as "gatekeepers" with respect to at least one of the Commission's pre-defined 'core-platform services.' Although those services are divided into ten categories, the same company can be identified as a gatekeeper for several core services[41]:

(1) online intermediation services (e.g., Apple Store, Google Play Store)

(2) online search engines (e.g., Google, Bing)

(3) online social networking services (e.g., Facebook, Instagram)

(4) video-sharing platform services (e.g., YouTube)

(5) number-independent interpersonal communication services (e.g., iMessage)

(6) operating systems (e.g., Apple iOS, Android)

(7) cloud computing services (e.g., Amazon Web Services)

[41] In its initial proposal, the European Commission (2020: 1) deploys of the ecosystem metaphor (van Dijck 2020) to describe the operations of gatekeepers, who 'enjoy an entrenched and durable position, often as a result of the creation of conglomerate ecosystems around their core platform services, which reinforces existing entry barriers.'

(8) advertising services (e.g., Google Ads)

(9) web browsers (e.g., Google Chrome)

(10) virtual assistants (e.g., Siri)

Whether or not a company that offers those services falls under the scope of the Digital Markets Act depends on its annual turnover in the European Economic Area (equal to or above €7.5 billion), market capitalisation (at least €75 billion), number of users (more than 45 million monthly active users in the EU), as well as the number of financial years in which those former two criteria were met (at least in each of the last three years). If a company is identified as a gatekeeper, it has to comply with specific legal obligations by implementing changes in its day-to-day business. For example, gatekeepers are not allowed to use the data of business users when they compete with them on their own platform, such as in the prominently discussed case of third-party sellers on Amazon Marketplace (see Chapter 4, section 4.4.1).

For the purposes of this chapter, the integration of cloud computing services into core-platform services and the definition of gatekeepers based on their market power are highly relevant. Indeed, all three lead firms discussed in the previous chapter (Amazon, Google, and Microsoft) would therefore be classified as gatekeepers. Before extending the analytical juxtaposition between the different roles of state actors and the operations of lead firms, the Artificial Intelligence Act also needs to be addressed.

### *5.3.2 The Artificial Intelligence Act*

Other than the Digital Markets Act, the European Commission's Artificial Intelligence Act is not restricted to companies with a particular market capitalisation or number of users. Rather, it applies to all sectors and industries, with the exception of AI systems 'exclusively developed or used for military purposes' (European Commission 2021a:

20). Another key difference between those two legislative proposals is that the Digital Markets Act was already approved by the Council of the European Union at the time of finalising this thesis and is predicted to come into force in 2023, whereas the Artificial Intelligence Act is still at the proposal stage. As a result of this, it is likely that its final legislation text might significantly diverge from the Commission's initial draft proposal, which is subject to negotiations between stakeholders – a rather lengthy procedure. As Heikkilä (2022) reminds us, the EU's General Data Protection Act 'took more than four years to negotiate, and it was six years before it entered into force.'

With that note of caution about analysing proposals that might change in mind, what is the rationale for the proposed Artificial Intelligence Act and what does it entail? As the European Commission (2021b: 1) explains:

> The uptake of AI systems[42] has a strong potential to bring societal benefits, economic growth and enhance EU innovation and global competitiveness. However, in certain cases, the specific characteristics of certain AI systems may create new risks related to user safety and fundamental rights. This leads to legal uncertainty for companies and potentially slower uptake of AI technologies by businesses and citizens, due to the lack of trust. Disparate regulatory responses by national authorities would risk fragmenting the internal market.

At the heart of the proposal is a risk-based approach that categorises AI systems into four levels of product safety-related risk: unacceptable risk, high-risk, limited risk, and minimal risk. While there are no additional legal obligations for providers of AI systems that are classified as posing minimal risk, the Commission proposes to impose

[42] The Commission deploys a broad understanding of AI systems and includes a wide range of approaches and methods, such as machine learning techniques, logic- and knowledge-based approaches and Bayesian estimation methods. Given that the theoretical framework of neural production networks focuses on production arrangements powered by artificial neural networks, the chapter restricts itself to those approaches to AI.

transparency requirements for AI systems that fall in the category of limited risk (e.g., chatbots). Regarding high-risk AI systems, the Commission (2021b) suggests the introduction of mandatory requirements such as human oversight and ensuring the quality of data sets. Systems that are seen as posing unacceptable risk, by contrast, would be banned completely. In the next section, I argue that it is fruitful to relate the Artificial Intelligence Act to the Digital Markets Act for making sense of the ambivalent roles of state actors and state action in neural production networks.

### 5.4. Discussion: State Roles in Neural Production Networks

The analysis of state actors in neural production networks differs from the analysis of state actors in global production networks because it relies on a profoundly different theoretical entry point: a different conceptualisation of the geographies in which state actors and state action are situated. In the case of global production networks, the *territorial expansion* of the space in which production takes place represents the entry point to conceptualise the positioning of state actors vis-à-vis dominant lead firms. By contrast, in the case of neural production networks, the *computational envelopment* of the space in which production takes place represents the entry point to conceptualise the positioning of state actors vis-à-vis dominant lead firms.

As a result of this reconfigured focus, the case study analysis of the EU's Digital Single Market Strategy unveils distinct manifestations of the aforementioned state roles: facilitator, regulator, producer, and buyer (Table 5.2). To demonstrate their theoretical relevance for extrapolating those insights beyond the contextual constraints of the case of the EU, it is worth explicitly juxtaposing those expressions of state action with state roles in the study of commodity-centric production geographies. Here, I am inspired by Horner (2016: 4), who argues that an analytical sensitivity to the reciprocal ways in which both states and production networks mutually shape each other can 'also help demonstrate the limitations, as well as the possibilities, of state and policy agency in economic development.' Nonetheless, I propose that it is necessary to scrutinise and reformulate established ways to delineate those state roles.

Table 5.2. State roles in neural and global production networks.

| | **Global production networks** | **Neural production networks** |
|---|---|---|
| **Empirical focus** | Ability of state actors to intervene in the operations of lead firms that attempt to exploit territorial inequalities<br><br>Territorial expansion of space in which economic production takes place | Ability of state actors to intervene in the operations of lead firms that attempt to create infrastructural dependencies<br><br>Computational envelopment of space in which economic production takes place |
| **Examples of state roles** | Facilitator: Attract foreign investments to capture economic value within territory<br><br>Regulator: Trade policies, environmental standards, and labour protection policies<br><br>Producer: Strategic capacity-building to exert control over physical infrastructure<br><br>Buyer: Ethical public sector procurement of tangible materials and commodities | Facilitator: Stimulate innovation based on perceived economic benefits of AI<br><br>Regulator: Risk-based approach to product safety and ex-ante competition policy<br><br>Producer: Inability of state actors to own and control computational infrastructure<br><br>Buyer: Dependency of state actors on lead firms' hardware and software resources |

The case study analysis of this chapter contributes to the existing geographical literature on the 'state-production network nexus' (Werner 2020: 1) in two ways. First, it addresses the emergence of new regulatory instruments in the domain of digital policy by situating them in relation to AI's production geographies. Second, it argues in favour of overcoming a fixation on commodities as the central objects of analysis to be placed alongside state actors in order to render intelligible their agency and constraints to affect the operations of lead firms. Instead, building on the arguments developed in Chapter 4, the chapter argues that the fact that lead firms do not centralise the *production* of AI, but rather its *provision*, affects how those state roles present themselves. Table 5.2 provides a summary of those impacts, resorting to examples from the earlier sections to allow for a top-level comparison. The purpose of this table is not to encompass all possible forms of state action but rather to point out the major conceptual differences between global production networks and neural production networks.

In the remainder of this chapter, I argue that two interconnected sets of tensions characterise the ways in which the EU's Digital Single Market Strategy aims to shape, and how it is shaped by, AI's geographies of production: facilitator vs. regulator and producer vs. buyer. The discussion begins with the former tension.

### *5.4.1 Tension #1: Facilitator vs. Regulator*

In March 2021, I was in a Zoom meeting with Brando Benifei, an Italian member of the social-democratic Progressive Alliance of Socialists and Democrats (S&D) group in the European Parliament. At the time of this meeting, Benifei served as the EU Parliament's lead rapporteur on the Artificial Intelligence Act[43]. As a lead rapporteur, Benifei's task was to analyse the legislative proposal and draft a report on it, confer with experts in the relevant area and individuals who may be impacted, and, most importantly, 'propose the political line to be followed' (European Parliament 2006). When I asked Benifei about what he perceives as the key hurdle for AI regulation, he emphasised the relevance to consider the economic strength of technology companies that are ought to be regulated. One year later, he specified this argument in an interview (Olivi 2022):

> We are also convinced that we can develop healthy competition and also influence other models in the world by being true to our values; that is to put the human

[43] The meeting was facilitated by the Foundation for European Progressive Studies (FEPS), a Brussels-based think tank that is affiliated with the social-democratic party family in the European Parliament. I mention this informal conversation to provide transparency about what led me to study the tensions between facilitator and regulator roles.

> being at the center, i.e. the human-centric approach. […] We need to put enough investment in, because it cannot be only about regulation. We need to have public and private investment, and we need the regulation to also be helpful to support investment. It is also important that we look at the issue of our digital sovereignty and how we develop our model for AI and the digital space, because that's a part of how we can be a strong Europe in a difficult world like the one we are in now.

Benifei's remarks open up the study of how *facilitative* state policies as part of the EU's Digital Single Market Strategy (e.g., fostering public and private investment) stand in contrast with *regulatory* state policies (e.g., setting up mandatory requirements for digital gatekeepers). Before disentangling this tension, it is worth stressing that my claims are necessarily restricted by the fact that neither the Digital Markets Act nor the Artificial Intelligence has come into force yet. Consequently, my arguments inevitably relate to how AI's geographies of production shape state action, rather than the other way around. But although it is too early to investigate how lead firms organisationally implement new legal obligations 'to ensure fair and open digital markets' (European Commission 2022), their public reactions to those policy proposals provide a number of useful sources of evidence that can underpin theory-building.

A straightforward way to elucidate the impacts of the operations of lead firms in neural production networks on the state-as-regulator role is to argue that without high degrees of industry concentration, there would simply not be the necessity to develop a new regulatory instrument that relies on a 'gatekeeper' categorisation in the first place. Given the criteria set out by the Digital Markets Act, is not a surprise that a number of American commentators not only exhibited plentiful scepticism about those proposals but even interpreted them as digital protectionism. In an op-ed for the *Financial Times*, Charlene Barshefsky (2020), a former United States Trade Representative, laments that the European Commission plans to 'handicap foreign companies and use regulation and funding to promote friendly, compliant local competitors.' Ostensibly, the EU does so by setting up 'discriminatory digital services taxes […] rigged competition laws […] unjustified barriers for foreign AI applications […] massive subsidies for a "European Cloud"' – measures that should be ended in favour of 'greater regulatory co-operation with the US' to push back against Chinese protectionism.'

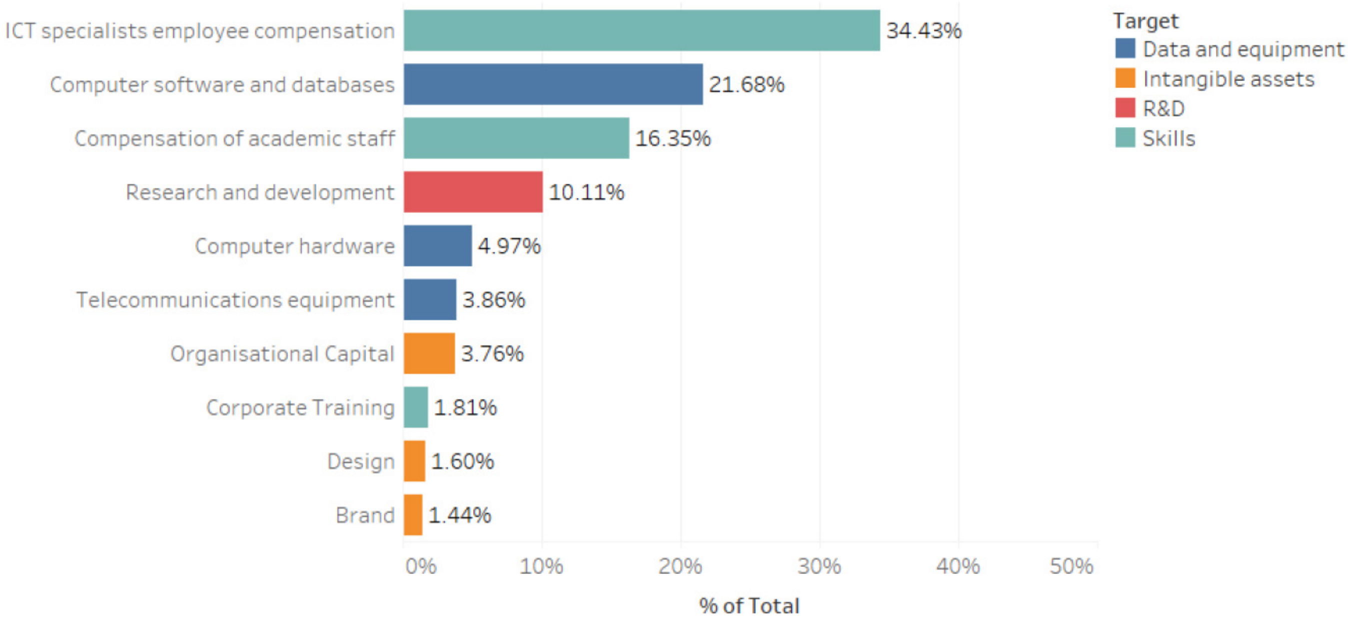
Naturally, American technology giants that are likely to fall under the scope of the Digital Markets Act shared Barshefsky's sentiment (see also Cini & Czulno 2022). In the politics of the EU, it is a common procedure that member states invite views from stakeholders on draft proposals put forward by the European Commission. And given that Google's European headquarters are based in Dublin[44], it took up the invitation by the Irish Government's Department of Enterprise, Trade and Employment to provide a public response. In it, Google (2021: 1) expressed concern that the proposals 'appear to specifically target a handful of companies and make it harder to develop new products to support small businesses in Europe.' In a nutshell, Google's strategic position was that companies are *already* adequately regulated and that there is no need for additional legislation. An Amazon spokesperson similarly voiced 'serious concerns about the Digital Markets Act unfairly targeting Amazon and a few other U.S. companies' (Stolton 2022). Microsoft (2021), by contrast, supported the proposed competition policy, stating that it will ultimately 'benefit the broader ecosystem and provide more opportunity for all participants, including ourselves.'

[44] The case of Google's public response to the Irish government shows the utility of combining insights from the literature on state roles in global production networks (Horner & Alford 2019) with empirical evidence on how lead firms in neural production networks impact on the facilitator and regulator roles of state actors. Neural and global production networks are not mutually exclusive but overlap in intriguing ways. In this case, a nation state acts as a facilitator, while a supranational state actor acts as a regulator. However, the chapter maintains its focus on analytically problematising the tensions between those state roles, rather than providing a comprehensive account of *all* state actors involved.

Against that backdrop, how does the tension between facilitative state action and regulatory tension play out in practice? To answer this question, it is key to consider the concentration of computational resources that are needed to deploy contemporary forms of AI in the hands of a few lead firms (see Chapter 4). As long as those firms own and control the proprietary infrastructure without which it is not possible to adequately fine-tune the parameters of artificial neural networks or train them on large-scale datasets, public and private investments into AI will amplify patterns of industry concentration. Conceptually speaking, this means that the EU's scope as a regulator of lead firms in neural production networks is constrained by its ambition to act as a facilitator of AI-powered innovations. An increase in public investments indirectly undermines the EU's regulatory leverage in ensuring 'healthy competition' (Olivi 2022) because it ends up subsidising dominant infrastructure-as-a-service providers. Without the provision of computational infrastructure, there would be no AI systems that could be classified into different levels of risk by the European Commission.

To unfold this argument, the analysis of EU-level policy documents is a useful methodological tool since it allows for contextualising how public investments in AI are institutionally situated. In this regard, one of the most important sources of evidence is a 2018 report titled "The Coordinated Plan on Artificial Intelligence." In it, the European Commission (2018: 2) lays out its objective to invest at least EUR 1 billion of EU-level funding into AI per year between 2021 and 2027 and 'gradually increase public and private investment in AI to a total of EUR 20 billion per year over the course of this decade.' And indeed, it is estimated that public and private AI investments seem to be approaching that objective. According to the EU's Commission's Joint Research Centre (2021: 3), in 2019, the EU invested 'between EUR 7.9 billion and EUR 9 billion in AI', which corresponds to '40-45% of the annual investment target of EUR 20 billion to be reached by 2030.' As Figure 5.1 indicates, approximately half of those investments were related to the compensation of employees and academic staff, while more than a quarter were used for computer software and software. The public sector accounted for 41 per cent of the EU's overall AI investments, particularly in the form of outlays in education[45] and the adoption of AI technologies in the public sector.

Figure 5.1: EU AI investments by category, 2019. Source: Joint Research Centre 2021.




For making sense of what those proportions mean for the 'distribution of rents or gains' and the state's 'regulatory role in relation to distributional outcomes' (Horner & Alford 2019: 11), a reasonable explanation is that the creation of rentier relations is a

45 As highlighted in Chapter 4, the domain of education is of great strategic relevance for lead firms in neural production networks because the expansion of adopting AI in education fosters the decentralisation of AI production and the centralisation of AI provision. This process does not only go hand in hand with a consolidation of economic power, but also with a sweeping intensification of cultural power. By providing the basic infrastructure on top of which generations of students are habituated into the workings of machine learning, lead firms set the scene for a future influx of high-skilled workers that are already familiar with their proprietary hardware and software configurations.

process 'unfolding over several years, if not decades[46]' (Birch 2020: 13). By developing the notion of rentiership, Birch (2020: 13) theorises 'economic rents and rent-seeking as a social process and practice rather than a distortion of an idealized and naturalized political-economic process or logic (e.g., markets "distorted" by rent-seeking).' While I acknowledge the methodological limitations of analysing publicly available documents and therefore cannot make claims about how much of the value of those investments is captured by lead firms, Birch's (2020) reasoning remains intriguing. If we rely on the assumption that the creation of rentier relations is a decade-long social process, this also means that once infrastructural dependencies are in place, it is unlikely that state actors can counterbalance or even reconfigure them *retrospectively*.

What I mean by this is that expressions of the state-as-a-regulator role vis-à-vis neural production networks are limited in their potential to fundamentally reconfigure distributional outcomes within those production networks. Certainly, the EU's Artificial Intelligence Act and the Digital Markets Act would bring about changes to the day-to-day operations of lead firms. But there are good reasons to question whether they would have significant impacts on the 'distribution of rent or gains' (Horner & Alford 2019: 11) that characterises AI's uneven production geographies. Compared to lead firms in commodity-centric production networks, the distinct ways in which technology giants

[46] In this context, it is worth stressing that the European Commission's Joint Research Centre (2021: 5) defines AI investments as 'expenditures on labour and skills as well as tangible and intangible capital assets incurred by public and private organisations to develop and implement AI to (re-)design business processes in order to create new or improve existing products or services.' The report's authors note that they use a broad understanding of AI investments that includes intangible assets such as employee training 'due to their critical role in the process of AI diffusion' (Joint Research Centre 2021: 5). Normally, those assets are considered as "expenditures" rather than "investments" in in statistics and accounting. I return to this definition when discussing buyer and producer roles.

organisationally operate (e.g., cross-subsidising their entry into new sectors to expand their proprietary ecosystems) enables them not only to swallow up competitors but also to develop better products and services. Overall, those capabilities create an industrial status quo that, infrastructurally speaking, benefits the dominant providers of scarcely distributed resources and services, such as AutoML offerings. But the scarcity of those resources does not lie in their one-and-only existence in space-time, as is the case with scarcely distributed commodities like diamond rings or Rolex watches. Instead, scarcity refers to the proprietary ownership of AI resources and services.

In summary, my argument is that the patterns of industry concentration outlined in Chapter 4 create a context in which the EU's facilitative state policies to seize what it perceives as the economic opportunities of AI inadvertently work against its regulatory policies to establish a level playing field for fair competition. Given the capabilities of lead firms to envelop the space in which the cross-sectoral production of AI takes place, political efforts to facilitate investments in AI inevitably create economic opportunities for Amazon, Google, and Microsoft. Nonetheless, the Digital Markets Act provides an impetus to rethink recent assessments of the declining importance of competition policy as a regulatory tool to shape production geographies (Mayer & Phillips 2017). The case study exemplifies that state actors develop completely new frameworks to account for the historical discontinuities of technology giants as lead firms[47]. In the next section, I

[47] Although it is beyond the scope of this chapter to thoroughly engage with the nuts and bolts of competition policy by tracing its century-long history, the most important feature of the Digital Markets Act is its *ex-ante* approach to regulating technology companies. The term ex-ante means "before the event." This feature affords the European Commission with the political leverage to enforce particular legal obligations for companies that are classified as being gatekeepers without having to make a *retrospective* case that those companies acted in anti-competitive ways in the past (e.g., by filing an antitrust case). It remains to be seen what the long-term impacts of this reconfigured approach will bring about.

extend this argument by considering the tensions between producer and buyer roles with respect to the EU's Digital Single Market Strategy.

### *5.4.2 Tension #2: Producer vs. Buyer*

A fruitful starting point to problematise the relationship between the EU's producer and buyer roles in neural production networks is to return to how its policy proposals related to digital sovereignty were perceived by Charlene Barshefsky, the former United States Trade Representative. As Barshefsky (2020) stresses, a key feature of the EU's alleged shift towards digital protectionism is its objective to prepare 'massive subsidies for a "European Cloud."' More specifically, the European Commission 'proposes to invest €4bn-€6bn in cloud infrastructure to store and process data in Europe and to support European cloud providers' – a proposal that was allegedly developed by 'leading French and German digital companies.' Either intentionally or accidentally, Barshefsky's quote presents two separate initiatives as a single approach. It is necessary to shed a spotlight on what these initiatives are and how they are related to each other.

First, what Barshefsky refers to as the subsidisation of a "European Cloud" can be identified as a declaration to create the "European Alliance on Industrial Data and Cloud" that was signed by 27 member states in October 2020. Echoing the findings of Chapter 4, member states identify a strong tendency of concentration in the global cloud computing industry, stressing that the 'public cloud infrastructure market is converging globally around four large non-European players' (European Commission 2020a: 1). While the declaration does not specify who those players are, it includes a reference to market research data that denotes Amazon, Google, Microsoft, as well as Alibaba as the dominant providers of cloud infrastructure. To push back against the market dominance of foreign firms, the 27 signatories indicate a commitment to fostering 'the emergence of a resilient and competitive European supply for the public and private sector needs of highly trusted, secure, interoperable and energy-efficient cloud infrastructure and services' (European Commission 2020a: 2). Foreign technology giants were even seen as prototypes of how this objective could be achieved. As Peter Altmaier, Germany's Minister for Economic Affairs and Energy at that time and one of the signatories of the declaration, told *Politico* journalists (Heikkilä & Delcker 2020):

> In order to achieve digital sovereignty, we need to start approaching data processing the way major American and Chinese companies – the hyper-scalers – approach it. This is an area where we're far from being equals.

Recapitulating Werner's (2020: 4) point that the empirical manifestations of the state-as-producer and state-as-buyer roles are most prevalent 'in strategic sectors such as energy, infrastructure and defence', this aspiration to replicate the modus operandi of American and Chinese firms is a useful finding. However, it would be too simplistic to argue that the EU's objective is to 'shore up domestic producers' (Werner 2020: 5) by developing public procurement strategies for cloud infrastructure providers. Instead, it is paramount to consider the idiosyncrasies of the EU as a supranational state actor that represents the interests of member states – and that particular member states are more powerful in enforcing their economic interests than others. The European Commission itself does not seek to establish state ownership over computational resources because it does not own companies in the first place. Its member states, by contrast, own shares in a number of companies that directly compete with American and Chinese lead firms in neural production networks. Hence, an adjustment of focus is required.

After the 27 member states declared their willingness to invest in homegrown cloud providers, 27 European CEOs signed a comprehensive report that was addressed to Thierry Breton, European Commissioner for the Internal Market. Breton is one of the leading political architects behind the Digital Markets Act. In the report, an alliance of companies that includes industry and software behemoths such as Siemens, Airbus, and SAP underline the importance of investments – not least because an increase in state subsidies would be beneficial for expanding their *own* computational infrastructure. At the same time, the report clearly indicates the strong links between cloud computing and artificial intelligence and the competitive advantages of technology giants:

> European companies need to achieve scale and provide better functionality – including in areas like cybersecurity – to incentivize users to switch from existing, highly performant, non-EU providers to European alternatives […] The absence of European players capable of competing with global cloud players on software and hardware competencies, such as artificial intelligence, machine learning and silicon technology[48], is accelerating the penetration of these firms into a breadth of European industries, from retail markets to automotive. […] Global players are pushing into the production facilities of the European industry with end-to-end solutions that further increase customer dependency. These initiatives will accentuate their strong market positions (European Commission 2021b: 13-14).

But beyond providing empirical evidence on how dominant lead firms in neural production are perceived by their European competitors, the report is also an impetus to stress the second initiative that Barshefsky's (2020) sweeping critique of contemporary European digital policy touches upon: GAIA-X. When it comes to analysing the tension between producer and buyer roles in the context of AI's geographies of production, the story of GAIA-X is an illuminating example. In line with the purposes of this chapter, I discuss it in a condensed way, rather than providing unnecessary details. In their report to Commissioner Breton, the European CEOs mention the term "GAIA-X" more than 20 times. They introduce GAIA-X as a project 'that has already taken steps to align on common frameworks for federated cloud services' (European Commission 2021b: 9) and point to the importance of building strong synergies with it. Initiated by Germany's

[48] The report cites Google's tensor processing units (TPUs) as an example for this (Chapter 4).

and France's Ministries of Economic Affairs and first presented at a conference in 2019, the initiative's self-proclaimed aim is decisively not to develop a 'new cloud physical infrastructure […] [but] rather a software federation system that can connect several cloud service providers and data owners together' (GAIA-X 2022).

Initially, GAIA-X was met with enthusiasm within both political and business circles: a gateway to reduce an infrastructural dependence on American and Chinese lead firms. Even Angela Merkel, Germany's chancellor at the time, expressed support for efforts that push for European digital sovereignty: 'So many companies have just outsourced all their data to US companies […] the value-added products that come out of that, with the help of artificial intelligence, will create dependencies that I'm not sure are a good thing' (Chazan 2019). Or, as a board member of the partially government-owned telecommunications company Deutsche Telekom[49] put it: 'it's irrelevant whether we are currently already technically able to build up such an infrastructure – there's no question that we must enable ourselves to do it' (Delcker 2019). With that objective in mind, why is it that GAIA-X's vision and mission statement explicitly rules out that the initiative could build its own physical cloud computing infrastructure?

Although there is a variety of ways to answer this question, such as bureaucratic complexity and infighting, I focus on only one of them: the role of corporate lobbying by lead firms in neural production networks. Soon after the initiative was first presented to the public, both Amazon and Microsoft openly challenged the plans. Intriguingly, a Microsoft spokesperson told *The Wall Street Journal* that although digital sovereignty is

[49] Deutsche Telekom was also one the signatories of the report to Commissioner Breton. The German government holds a 31.9 per cent stake in the company. It is estimated that the company's market share in the entire European cloud market was 2% in 2020. Alongside Orange, which is partially owned by the French government, Deutsche Telekom is one of the only cloud providers that exemplify the state-of-producer role.

a legitimate goal, 'in the cloud age, however, we think it is wrong to define sovereignty solely along territorial borders[50]' given that sovereignty 'needs the most powerful cloud solution' (Stupp 2019). But in spite of those reservations, or rather *because* of them, all major technology giants joined the initiative, with Alibaba, and Huawei sponsoring the project's 2021 summit. As a French GAIA-X member emphasised after the summit, all board member companies 'are either a client or a major partner of U.S. cloud giants [...] [and] they have a direct interest to work with these players' (Goujard & Cerulus 2021). Given those infrastructural dependencies, it is not a surprise that the leverage of public-private initiatives such as GAIA-X remains structurally limited.

This latter point is in line with economic-geographical scholarship on state roles in global production networks. As Horner and Alford (2019: 11) argue, a wide range of studies has shown that 'the capacity of public-private governance to achieve equitable distributional gains is fundamentally constrained by the sourcing practices of lead firms and the foundational logic' of production geographies. Undoubtedly, the 'foundational logic' of neural production networks is so skewed toward a concentration of power and control that it is unlikely that public-private modes of governance can reconfigure those uneven distributional outcomes. Herein lies the key argument of this section: the EU's deficiencies to act as a producer exacerbate its dependency as a buyer. Clarifying the meaning of "producer" is important because this term illustrates how neural production networks diverge from other, previously conceptualised forms of cross-border economic interactions. "Provider" is a much more appropriate term than "producer" to indicate state ownership because "producer" actually means "buyer." That is the sophisticated

[50] This quote shows that the discursive metaphor of digital sovereignty is so elastic that it can be completely repurposed and used in direct opposition to its original intention. That is one of the reasons why the notion of sovereignty has been taken up by so many European policy-makers and companies to enforce their strategic interests when it comes to AI.

discursive trick of powerful lead firms: because they decentralise the production of AI and centralise the provision of AI, they foster a coalescence of producer and buyer roles. In order for the EU to become a producer in the economic-geographical understanding of the word, it would have to own and control the means of AI provision.

Instead, as long as those means are in the hands of extraterritorial firms, efforts to accelerate and subsidise AI production result in increased public procurement from them. Although there is a lack of comprehensive statistics on the public procurement of AI given that existing estimates do not specify what it is included in categories such as 'computer hardware' or 'computer software' (Joint Research Centre 2021), the analysis of reports is useful for briefly pointing out *how* procurement takes place. As the High-Level Expert Group on AI, an independent group set up by the European Commission (2019: 18) to provide policy recommendations, notes:

> Today, the most advanced governments are increasingly providing application programming interfaces (APIs) to trusted intermediaries as a way to open up their infrastructure to private sector services and entrepreneurs […]. Through its role as a procurer, the public sector can also make use of public procurement strategies to not only incentivise the development and responsible innovation of AI systems for the public good, but also to promote responsible innovation.

Echoing the prominent discussion of APIs as technical gateways between users and the proprietary ecosystems of lead firms in neural production networks in Chapter 4, it is interesting that governments are explicitly advised to use them for opening up their infrastructure. Such recommendations bring to mind Helmond's (2015) emphasis on the relevance of APIs for 'making web data platform ready.' Empirically tracing and comparing different ways in which governments follow those recommendations is an auspicious task for future research on 'API governance' (van der Vilst et al. 2022). To underpin such an endeavour with a spatially sensitive analytical approach, frameworks emanating from economic geography, notably Horner's (2017) typology of state roles in global production networks, offer an immensely useful source of inspiration.

**5.5 Conclusion: AI's Regulatory Futures**

In this chapter, I analysed how the EU's Digital Single Market Strategy relates to the operations of lead firms in neural production networks. In particular, I focused on two policy proposals that are related to the objective of European digital sovereignty: the Digital Markets Act and the Artificial Intelligence Act. While addressing the research question framing this chapter (*How does state action shape, and is shaped by, AI's geographies of production?)*, I noticed that it is too early to provide an adequate answer with regard to the *first part* of this question: the role of state action in shaping neural production networks. This is due to the fact that the emerging regulatory instruments that I focused on as part of the case study have not come into force yet. Alternative ways to address this first part of the question would have required a rather different methodological approach, such as systematically mapping the institutional factors that enable and constrain the operations of lead firms.

Nonetheless, my approach enabled me to craft conceptual arguments that can be utilised to answer the *second part* of this chapter's research question: the role of neural production networks in shaping state action. Far from rendering state action redundant, patterns of industry concentration brought into being by technology giants reconfigure how the policy repertoire of state actors manifests itself. Specifically, I made the case for an increased analytical sensitivity towards the distinct frictions between the state's role as a regulator, facilitator, buyer, and producer in the context of AI. In juxtaposing empirical findings on the operations of lead firms (Chapter 4) with qualitative evidence on the key contours of the EU's Digital Single Market Strategy, two interconnected sets of tensions where identified that are of relevance beyond the contextual constraints of this case study: First, the EU's political leverage to act as a regulator is limited by its function as a facilitator. Investments in AI go hand in hand with economic opportunities of lead firms to consolidate their infrastructural power, rather than establishing a level playing field for fair competition between different cloud providers. Second, the EU's deficiencies as a producer – that is, an *AI provider* that owns and controls computational resources needed for artificial neural networks – intensify its reliance as a buyer. Due to the fact that neural production networks do not gravitate around the creation of physical commodities but rather the establishment of rentier relations, a new interpretation of the boundary between producer and buyer roles is required.

Those conclusions contribute useful additions to emerging bodies of research at the intersection of AI and platform regulation (Seyfert 2018; Gorwa 2019; van Dijck et al. 2018; Schmitz & Seidl 2022) and economic geography (Coe & Yang 2022; Horner & Alford 2019; Werner 2020). Instead of seeing those fields of work as disconnected debates, this chapter argued in favour of bridging the gap between them by combining insights emanating from both sides. Additionally, my arguments relate to broader cross-disciplinary conversations about the political implications and ramifications of rentier relations (Birch 2020). As Christophers (2020: 474) puts it, 'if rentierism is defined largely by exclusive control of assets, then non-rentierist activity is […] either non-asset-intensive or in which assets are fundamental but not exclusively controlled.' The key question, then, is whether it is possible to organise the computational resources that underpin contemporary forms of AI in such a way that is not exclusively controlled. State ownership, for example, would not represent a magic bullet to solve this problem because it can turn also turn into a form of exclusive control. As the case of GAIA-X lucidly illustrates, even if the grand aspirations of state actors are no less than to 'assert ourselves in the world' (Goujard & Cerulus 2021), as Germany's former Minister for

Economic Affairs put it, materialising those aspirations is a different matter entirely. In short, a lot of conceptual and empirical work remains to be done.

At the same time, it is necessary to acknowledge the limitations of focusing on the EU's Digital Single Market Strategy as a paradigmatic case study to examine the regulation of neural production networks. The first limitation relates to the sheer scale and complexity of the EU as a supranational state actor. As highlighted throughout the chapter, this complexity is not only derived from inter-institutional tensions between different entities such as the European Commission and the European Parliament, but also from an asymmetrical power of nation-states to both assert and successfully push through their strategic interests. If we consider the fact that Germany and France are the two member states with by far the highest shares of investments in AI (Joint Research Centre 2021: 11), it is not surprising that they are key driving forces in attempting to establish a form of digital sovereignty, however that term is defined. In this respect, comparative research on how different member states aim to shape AI's geographies of production would be a productive way to overcome those limitations.

Second, and this is why it is key to interpret the subtitle of this conclusion in the literal sense, a substantial proportion of the policy considerations that I discussed in this chapter remain uncharted territory. No suitable precedent exits that could be studied to anticipate how AI's regulatory futures will unfold. *Other* manifestations of networked production, such as the extraction and commodification of scarce natural resources and their globe-spanning circulation, pose fundamentally different regulatory challenges. But for the first time in history, governments around the world now face the task to reign in neural production networks. However, as the next chapter demonstrates, the operations of technology giants as lead firms and the reactions of state actors are by no means the only forces that determine AI's regulatory futures.

# Chapter 6: Neural Production Networks and Platformised Cultural Production

## 6.1 Introduction: Mona Lisa's Multiplicity

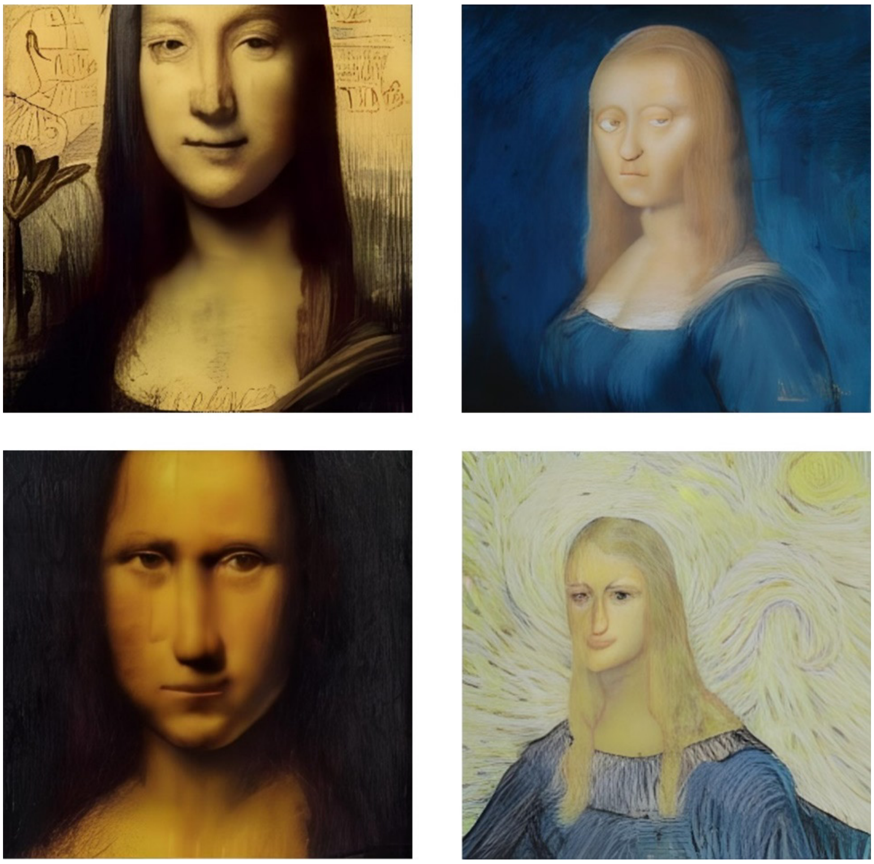
In November 2021, three months before the Russian invasion of Ukraine, Sber – a state-owned Russian banking and financial services company – announced the creation of ruDALL-E, the 'world's first neural network capable of creating images based on text descriptions in Russian' (Sber 2021a). What is noteworthy about the text-to-image model is not only its cost (developers claim the model required 23,000 GPU hours to train it based on 120 million image-text pairs) but that it is open to the public. Unlike its American counterpart DALL-E, a proprietary model presented by the Microsoft-backed research lab OpenAI in early 2021, a 1.3 billion parameters version of ruDALL-E is available under an open-source license. In addition to trying out the model using text queries (Figure 6.1), anyone can download and modify its source code.

In the following months, developers modified ruDALL-E in creative ways. In one instance, the deep neural network was fine-tuned on all official Pokémon images, resulting in a potent and accessible tool for bringing indefinite new Pokémon-inspired creatures into being (Woolf 2022). Another developer simplified the process of fine-tuning so that users can train the model on their own pictures within a few clicks on Google Colab[51] – combined with an appeal to refrain from using their free software to create for-profit non-fungible tokens (NFTs) (Looking Glass 2021). As one artist, who trained the model on bird illustrations from old books in order to create synthetic birds,

[51] Google Colab is a cloud-based machine learning platform. It allows users to run python code in the browser, without having to use additional software. Google Colab also provides free access to computing resources, including graphics processing units (GPUs).

admitted: 'I'm playing around with some silly toys here. I have no idea how this whole thing learns anything' (Solis 2022). In other words, there is no need to comprehend the inner workings of neural networks for training text-to-image models.

Figure 6.1: Images generated by ruDALL-E. Query: Mona Lisa painted by van Gogh.



The example of ruDALL-E illustrates that the creation of synthetic media has become easy, widely accessible, and fast. Synthetic media is an umbrella term for AI-generated images, videos, environments, audio, and text.[52] Historically, the automated production of synthetic media is by no means a new phenomenon. Indeed, modern computing is premised on enabling virtual realities. These realities have become all too

[52] The chapter intentionally uses a broad understanding of synthetic media to cast a wide net of empirical phenomena and highlight their relevance for a theory of hyperproduction. This is a useful strategy to adequately answer the research question framing this chapter.

real for some (Rothman 2018). Deep fakes – the use of face-swapping techniques to manipulate videos or images – are the most apparent controversy over synthetic media within both academic discourse and public debate. But amidst the spotlight on questions of misinformation and geopolitics, literature on the relationship between AI-generated synthetic media and social theory has been underrepresented. Irrespective of whether the synthetic Mona Lisas resemble Vincent van Gogh's style in a meaningful way, the visualisation of this absurd potentiality symbolises the research question of this chapter: *What do AI's geographies of production imply for platformised cultural production?*

The chapter answers this research question by introducing a new term to theorise AI-generated media: hyperproduction. Hyperproduction designates the penetration of cultural life with deep generative models. ruDALL-E is a clear example of such a deep generative model. It is a neural network architecture that was designed to recognise the 'true data distribution of the training set so as to generate new data points with some variations' (Pandey 2018). The 120 million labelled image-text pairs, which include the Mona Lisa and van Gogh's paintings, constitute ruDALL-E's training set. Therefore, it is paramount to begin with the premise that deep generative models, by design, 'present us with a technically and historically predetermined space of visual possibilities' (Offert 2021: 3). Hyperproduction, in other words, never comes from scratch. Far from making way for a Matrix-style simulation that escapes the grip of social and media theory and is somehow detached from the rest of us, the generative worlds of synthetic media remain not only *grounded in* but also *bounded by* reality.

This reality is best defined as rentier capitalism. For Christophers (2020: 1), rentier capitalism refers to a 'system of economic production and reproduction in which incomes are dominated by rents and economic life is dominated by rentiers.' The key idea is that economic actors are now more likely to make profits by monetising their valuable assets instead of selling commodities. An asset is a type of property that can engender revenue streams by 'turning things into resources which generate income without a sale' (Birch 2015: 122). For example, the reason for releasing an open-source version of ruDALL-E is not the empowerment of creatives. Comparable to trial periods for streaming websites, the model is a sample of what is possible at a higher scale. Instead of 1.3 billion parameters, the commercial version of ruDALL-E contains 12 billion parameters. As a result of its increased size, Sber (2021b) claims that this more sophisticated version 'is suitable for creating commercial materials: illustrations for advertising, architectural and industrial design, and vector and stock images.' This, of course, is a marketing pitch. Nonetheless, ruDALL-E exemplifies the transformation of a deep generative model into a proprietary infrastructure offering – a scarce asset whose intended reach goes beyond a singular industry or application domain.

Given the role of such proprietary ownership structures, it becomes evident that hyperproduction is not just an extension of Baudrillard's (1994: 1) idea of hyperreality; that is, the 'generation by models of a real without origin or reality.' Rather than the capital/labour dichotomy that characterised earlier critiques of capitalism, Baudrillard famously argued, simulations, signals, and codes increasingly define social interactions and activities. Capital and labour are only two of a plethora of codes that mediate how people interact with one another and how they conduct their daily lives. In rendering intelligible the role of the media in the second half of the 20th century through the lens of hyperreality, Baudrillard identified a 'precedence of simulation over all that already existed as real' (Hegarty 2004, 60). In the simplest terms, we are unable to distinguish reality from a simulation because simulations already are omnipresent features of reality itself. This argument remains indispensable in an era of AI-generated media. But in the

1990s, Baudrillard could not have foreseen the emergence of deep generative models as algorithmic driving forces of synthetic media production.

In offering a metaphor to capture the cultural ramifications of neural production networks, hyperproduction blurs the line between inputs and outputs as two ends of a spectrum to delineate the production process. In its *Glossary of Statistical Terms*, the OECD (2001) defines economic production as 'an activity carried out under the control and responsibility of an institutional unit that uses inputs of labour, capital, and goods and services to produce outputs of goods or services.' This definition illustrates that the process of production is typically imagined and measured as a linear development of sequential transformation that is defined by inputs and outputs: two easily identifiable variables. But the circular nature of hyperproduction defies this linear paradigm. As the creation of AI-generated media becomes a matter of a few clicks, it is neither feasible nor fruitful to conceptualise their production process as a unidirectional chain of events, just like the production of a physical medium or a commodity.

Juxtaposing two use cases – autonomous vehicles and virtual influencers – the chapter shows that far from reflecting a step-by-step process, hyperproduction is better understood as a circular dynamic that serves to maximise rent extraction. Autonomous vehicles, for example, are now trained in simulated environments powered by video game engines[53]. Virtual influencers, on the other hand, blur the analytical distinction between production inputs and production outputs by design: their success is defined by virtuous social media feedback loops. However, what unites both use cases is that they

[53] A video game engine is a software architecture that comprises required libraries and support tools, thereby partially automating laborious tasks faced by game designers. For example, in 2019, 53 percent of the top 1,000 games in Apple's App Store and Google Play were developed with Unity Engine, including Pokémon Go (Hollister 2020).

exemplify the emergence of new revenue streams facilitated by the confluence of video game engines and deep generative models. Although these use cases are embedded in different industries, they showcase how two companies, Unity Technologies and Epic Games, consolidate a key part of the proprietary infrastructure needed for the industrial use of hyperproduction. The intensification of their economic and cultural power has important consequences across and beyond cultural industries.

To elucidate those consequences, the chapter begins with a brief media history of synthetic production. This contextual background serves to justify the chapter's focus on the combination between deep generative models and video game engines as driving forces of contemporary forms of synthetic media production. The chapter then situates hyperproduction as a distinct logic of rentier capitalism, before proceeding to the case study analysis. The discussion shows that two factors of asset specificity of video game engines are especially pertinent for hyperproduction: real-time physics and 3D graphics, and programmability and modularity. The chapter concludes by pointing out the implications of the burgeoning hyperproductive condition.

### 6.2 A Brief Media History of Synthetic Production[54]

The purpose of this section is to situate the chapter's focus on the confluence between deep generative models and video game engines by considering synthetic production as a longstanding discursive aspiration and material phenomenon. Indeed, synthetic media is as old as computing. Whereas synthetic data is closely tied to the current nature of

---

[54] Section 6.2 was co-authored with Fenwick McKelvey during my stay at Concordia University in Montréal. Its basis was laid in a n article that is under review at *Distinktion: Journal of Social Theory.* I am the first author of this paper and the other sections are my own work. This chapter is a substantially reworked and expanded version of this article. Section 6.2 provides essential historical context to underpin the chapter's arguments.

artificial neural networks and the demand for large-scale training datasets to optimise the parameters of models, the automated production of synthetic media is an enduring aspiration in computer science (Gaboury 2021). Synthetic media is integral to the discourse of AI from the early daemons to Joseph Weizenbaum’s 1964 ELIZA program. Weizenbaum famously warned about the ELIZA effect where humans demonstrate a willingness even a desire to believe in computers, in his case as compassionate listeners. Weizenbaum’s warning about humanity's desire to be deceived is a central premise and anxiety about artificial intelligence. As Natale (2021: 56) puts it:

> [Weizenbaum’s] deep interest, on the contrary, in how the program would be interpreted and “read” by users suggests that he aimed at the creation of an artifact that could produce a specific narrative about computers, AI, and the interaction between humans and machines. ELIZA, in this regard, was an artifact created to prove his point that AI should be understood as an effect of users’ tendency to project identity. In other words, ELIZA was a narrative about conversational programs as much as it was a conversational program itself.

An influential early deceitful media, ELIZA demonstrates constitutive anxiety about computing as much about humanity's ability, if not to say its willingness, to be deceived, and the investment in technologies to match this need. The history of synthetic media should, therefore, rightly be considered a story of a long-standing desire for media to create virtual worlds. It is pivotal to acknowledge the depth of scholarship on this promise in cinema studies and game studies. As Jacob Gaboury (2021: 4-5) writes in the introduction of a much longer history of digital computing, much of the research on computer-generated images is ’almost always framed by discourses of realism or mimesis, or broad narratives of technological development of verisimilitude.’ AI-generated media hyper-delivers on that promise, putting in question the desire for verisimilitude. Gaboury (2021: 5) continues, ‘almost all images we view, make, and interact with on a daily basis are shaped by computation'. Computer graphics are not only becoming more real, but reality is also becoming more computer-generated. We instead see something more of an omega point where computers become more capable of verisimilitude just as the world becomes more computable.

Synthetic media is post-digital. Part of what the post-digital covers, according to Berry and Dieter (2015: 5), is a wide range of issues attached to the entanglements of media life after the digital, including a paradigmatic shift from an earlier moment driven by an almost obsessive fascination and enthusiasm with new media to a broader set of affectations that now includes unease, fatigue, boredom, and even disillusionment. The cultural perception that today's AI-generated media might be a little *too real* is certainly part of the structures of feeling associated with the post-digital. From this perspective, we are living in an already post-digital world defined by decades of laborious world-building through materialising computer modelling and computational imagination. This evolution might be characterised by what Hayles (1999: 8) defines as reflexivity: 'the movement whereby that which has been used to generate a system is made, through a changed perspective, to become part of the system it generates.' In a nutshell, the very possibility of synthetic media being a productive force indicates a particular conjecture in post-digital societies (Berry & Dieter 2015).

Hyperproduction emphasises how circuits of capitalism integrate with these histories of computer graphics and the post-digital era. The very possibility of modern computing in part depends on American military subsidies that looked to the power of computing as new techniques for managing battlefields and wars. Gaboury's historical contribution, among many, lies in situating the beginnings of computer graphics at the University of Utah – an early node on the ARPANET and a major recipient of ARPA funding. The benign teapot, one of the earliest computer models, developed after over a decade of experiments in less visual synthetic media deployed as part of the Vietnam War. As researchers at the University of Utah produced the first computer graphic in 1968, ARPA was working with the CIA under the guidance of Secretary of Defence Robert McNamara to build a “digital hamlet intelligence system.”

The Hamlet Evaluation System (HES) produced monthly statistical reports on the status of ‘population control’ in the countryside, as well as state-of-the-art digital maps showing the status of hamlet control by either the Government of Vietnam or the National Liberation Front. By spring 1967, the HES was fully functional in the field and stayed operational – albeit with several upgrades – until 1973. The HES consumed ‘thousands of man-days of work by countless young army officers’ (Allen 2001: 224), making it one of the largest – and most labour-intensive – geographical information systems of its time (Allen 2001: 417). This virtual Vietnam was an important example of the promised computational imaginary of computing. These early forms of synthetic media demonstrate the investment in computing as a means to better wage war – innovations that were privatised a decade or so later and put in corporate service and partially re-articulated as Silicon Valley innovation (Turner 2006).

The legacy of the US military's funding of synthetic media is directly applicable to this chapter’s focus on how video game engines relate to deep generative models. Game engines are in some way an accidental discovery as early first-person developers found new revenue streams by licensing out their internal engines to game developers. Epic Games, one of the oldest developers of game engines made famous by its success with Fortnite, co-developed this industry by licensing out its Unreal Engine. After the success of Unreal, Epic Games continued to refine its game engine, taking into account the feedback of developers. Unreal’s licensing strategy acted as a breakdown between video games as an entertainment industry to video games becoming more integrated into wider circuits of cultural life. To further this link, the first game to use the Unreal Engine 2 was America's Army, quite literally what Nick Dyer-Whiteford and Greg de Peuter (2009) label as a 'game of empire.' America's Army was a free-to-play online competitive shooter meant as a recruitment tool. That is the moment this chapter wishes to explore: the breakdown when a video game engine acts as a virtual reality that has as its object a kind of post-digital effect, of games becoming soldiers.

Acknowledging the oldness of synthetic media, the chapter sheds a spotlight on the latest formation that automates and streamlines the use of deep generative models across and beyond cultural industries. In coining a new term to grasp the integration of deep generative models into the circuits of media production, the chapter is inspired by Johnson's (2017: 225) argument that 'a theory of media production enables us to understand not just distinct media institutions and industry sectors, but also how those forces figure into larger processes of communication, culture, and meaning creation.' The political economy of communication, in other words, is not only an economic issue, and neither is the rise of hyperproduction. In short, this brief media history of synthetic production indicates the longstanding links between discursive aspirations and material realities. Building on that history, the next section resorts to a particular computational imaginary as the starting point to explore the confluence of deep generative models and video game engines: the idea of an industrial metaverse.

## 6.3 Overview of Case Studies: Unity Technologies and Epic Games

> I'm most excited by the potential of the "industrial metaverse" where the goal doesn't have anything to do with social interaction; rather, it's about *simulating experiences in the virtual world before moving into the physical world*. […] The industrial metaverse can transform the way every physical asset – buildings, planes, robots, cars, etc. – on the planet is created, built, and operated.

In January 2022, Danny Lange, Senior Vice President of Artificial Intelligence and Machine Learning at Unity Technologies, a San Francisco-based video game software development company, shared his vision for an industrial metaverse with the public. Unity is best known for its eponymous game engine, a toolkit used by millions of game designers around the world. Unity's main competitor in the video game engine market is the above-mentioned Epic Games, which created the Unreal Engine and uses it to power its game Fortnite. Initially designed for first-person shooter games, both game engines have become advanced enough to be seen as production tools for film production and graphic design. Both firms generate revenue by renting out access to their proprietary game engines as subscription-based infrastructure offerings. And in recent years, both firms have expanded their provision of engines beyond media industries. Unity's Danny Lange (2022) highlights his work with Boeing in developing 3D simulations for aircraft design. Epic Games spotlights a range of use cases for the Unreal Engine, ranging from architectural design and urban planning to military simulations.

The cross-industrial provision of video game engines and associated corporate imaginaries surrounding a future industrial metaverse neatly illustrate the 'technological expansion and empowerment of rentier capitalism' (Sadowski 2019: 3). Of course, the extraction of rent through infrastructure offerings is by no means limited to video game engines (see Chapter 4). Similar to the provision of AI-as-a-service offerings and cloud computing infrastructure, video game engines were initially created to fulfil company-internal needs.[55] Amazon's tremendously lucrative cloud computing subsidiary AWS,

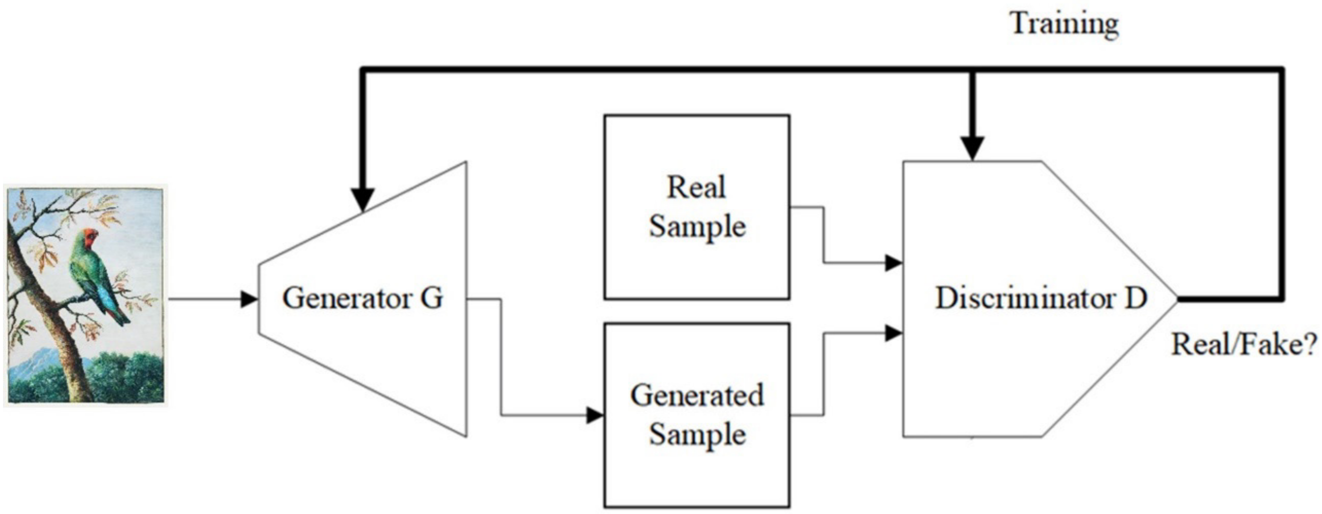
[55] According to Unity's co-founder David Helgason, 'we were going to be a game studio, but when we decided we'd be a technology company back in 2004, we basically wrote on the wall that we wanted democratise game development – to enable small studios with whatever high-end technology they felt was necessary for them' (Cook 2014).

for instance, also sprung out of an internal tool to manage the increasing complexity of the company's own logistical network. In other words, the assetization of video game engines – the process of turning them into an infrastructure offering that is 'owned or controlled, traded, and capitalized as a revenue stream' (Birch & Muniesa 2020: 2) – does not seem to be novel. In this light, one might ask, why does hyperproduction nonetheless constitute a *new* manifestation of rentier capitalism?

### *6.3.1 Deep Generative Models and Video Game Engines*

The answer to this question lies in the computational features of deep generative models as the primary technical driving force of contemporary synthetic media generation. To recapitulate, the purpose of those models is to enable algorithms to recognise patterns in the data they are provided in order to produce new outputs (e.g., images) inferred from the training dataset. While there is a wide range of deep generative models, generative adversarial models (GANs) are their most well-known subtype.

Figure 6.2. Logic of generative adversarial networks. Adapted from Wikimedia (2022).




The central logic of GANs, as introduced by Ian Goodfellow and his colleagues, might be imagined as a productive contest between two artificial neural networks. First, the *generator* network creates artificial outputs from random input, like public domain bird illustrations (Figure 6.2). The second network, called the *discriminator*, is fed both *synthetic* and *genuine* bird illustrations and taught to distinguish between the two. The generator improves if the bogus nature of its outputs is recognised by the discriminator. If the latter is tricked by the generator into believing a synthetic bird illustration to be real, it is less likely to make the same mistake again[56]. Before the discriminator network is finally discarded, 'the interplay of these two networks, both of which want to outplay the respective other, leads to incremental improvements on both sides' (Whittaker et al. 2020: 93). In Baudrillardian (1994: 2) terms, rather than attempting 'to make the real, all of the real, coincide with their models of simulation', GANs attempt to iteratively make their synthetic outputs coincide with their pre-defined models of the real.

To facilitate this iterative process of synthetic media generation in complex use cases, game engines such as Unity and Unreal come into play. The creation of synthetic images from 2D public domain bird illustrations by using a GAN, for example, does not include the rendering of a bird's morphable physical structures. But in other situations, such a 3D modelling might be required – and video game engines are now widespread infrastructural tools to assist with synthesising potent 3D models. In 2016, for example, Wood et al. (2016: 1) presented *UnityEyes*, 'a novel method to rapidly synthesize large amounts of variable eye region images as training data […] that can be used to estimate gaze in difficult in-the-wild scenarios.' Put differently, the aim was to fabricate many realistic images of eyes that can be used to statistically predict where a real person is

[56] In this example, what Figure 6.2 classifies as the *real sample* consists of illustrations rather than images of birds – a playful demonstration that the meaning of *real sample* is context-dependent. Given that we focus on theorizing the role of deep generative models in giving rise to a distinct form of rentier capitalism, we intentionally use a simplified portrayal of GANs. Not all forms of synthetic media are generated by GANs.

looking. The crux lies in the software used by Wood et al. (2016: 5) to generate the 3D eyeball model (Figure 6.3) for rendering one million synthetic eye images:

> We used the Unity 5.24 game engine to render our eyeball and generative eye region model. Our contribution here is a massive speedup in rendering time compared to previous work. This allows us to easily generate datasets several orders of magnitude larger than before – an important factor in successfully training large-scale learning systems […]: a 200x speedup.

Figure 6.3: Eyeball models created with Unity engine. Source: Wood et al. 2016.

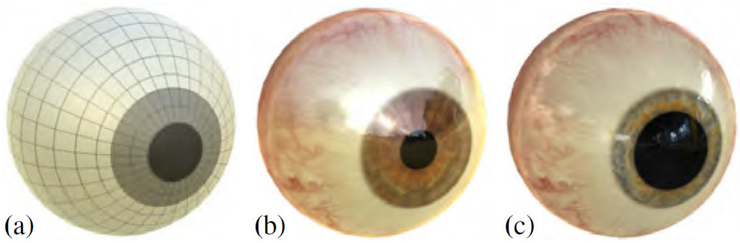



**Figure 2:** *Our eyeball mesh (a) shown rendered with physically-based materials and refraction effects. We model pupillary contraction (b) and dilation (c) as part of the refraction shader.*

In this sense, Wood et al.'s (2016) rationale for using Unity's engine rather than writing their own image rendering software reflects the economics of cloud computing infrastructure more broadly. Most notably, the authors refer to the vast increase of their dataset and the acceleration of rendering time. Such an expansion of resources is also a key reason for the creation of rentier relations in neural production networks, such as by using Amazon's SageMaker platform to simply the development of machine learning models(see Chapter 4). As Srnicek (2021: 39) reminds us, 'cloud computing is desirable for businesses because it enables the rapid expansion of resources, often at levels of technical expertise that are far beyond what the businesses themselves can provide.' But there is a fundamental difference between the operations of technology giants and the providers of video game engines: the dominant lead firms in neural production networks are not necessarily the only driving forces of hyperproduction.

Instead, some of the world's most powerful technology companies and their AI research laboratories rely on the provision of game engines by Unity Technologies and Epic Games as training grounds for machine learning. In 2018, Google's subsidiary DeepMind, for example, signed a deal with Unity Technologies to use its engine 'at a giant scale to train algorithms in physics-realistic environments' (Captain 2018). Yet, both sides refrained from disclosing any details about the licensing costs. Along similar lines, although OpenAI built a simulation on top of the Unity game engine to train a real robotic hand how to solve the Rubik's Cube (Akkaya et al. 2019: 7), details about those rentier relations remain nebulous. In spite of this lack of data, the infrastructural role of game engines evokes Birch and Muniesa's (2020: 6) point that 'assets are often unique, meaning that their value derives from their asset specificity.'

The term asset specificity denotes 'the degree to which the value of a specific or set of tangible or intangible assets is connected to their current use' (Moles & Terry 1997b). A high degree of specificity suggests that it is unlikely that an asset would be used for anything other than what it was initially intended for (e.g., the production of video games). Low asset specificity indicates that an asset has a wide range of potential applications. Game engines are examples of low asset specificity.

If even a behemoth such as Google resorts to the licensing of Unity's game engine, this is a clear sign that Unity's proprietary software architecture cannot be reproduced. Industry-leading game engines, in other words, have become scarce 'infrastructure assets' (Birch & Muniesa 2020: 13) whose reach is not limited to a particular purpose or clearly defined application domain. By infrastructure assets, I mean that their revenue streams underlie synthetic media production across industries. However, there is a relative absence of conceptual work on the assetization of video game engines. Although media scholars have emphasised their non-gaming industry deployment (Freedman 2018; Nicoll & Keough 2019), such empirical use cases have not been analysed through the lens of emerging rentier relations or contextualised as a continuation of a much longer media history of synthetic production.

### 6.4 Discussion: Hyperproduction and Asset Specificity

The remainder of this chapter argues that two factors of game engines' asset specificity are especially pertinent for hyperproduction: real-time physics and 3D graphics, and programmability and modularity. Those factors constitute the low asset specificity of video game engines. Empirically, the chapter substantiates this argument by comparing two use cases: autonomous vehicles and virtual influencers. Juxtaposing the rationale for deploying video game engines across those use cases indicates that hyperproduction is not a unidirectional process that simply transforms inputs into outputs.

Table 6.1: Asset specificity of video game engines in two use cases hyperproduction.

| | **Sim2real: Autonomous vehicles** | **Real2sim: Virtual influencers** |
|---|---|---|
| **Real-time physics and 3D graphics** | Transfer of knowledge and patterns from simulation to reality<br><br>Deliberately design and vary the parameters of 3D simulations to collect synthetic training data<br><br>Randomisation of parameters such as roads and weather conditions | Transfer of knowledge and patterns from reality to simulation<br><br>Simplified 3D rendering on the basis of real-world demographics and psychographics of consumers<br><br>Tailoring of parameters such as facial expression and personality traits |
| **Programmability and modularity** | Integration into R&D, physical manufacturing, global fleet, etc.<br><br>Lock-in effects by creating modular pipelines between services<br><br>Habituation of engineers into the workflows of game engines | Integration into broader marketing campaigns, market research, etc.<br><br>Lock-in effects by restricting use through license agreements<br><br>Habituation of marketers into the workflows of game engines |

Although autonomous vehicles and virtual influencers initially seem to be at odds with each other, they congeal into a useful illustration of the circuitous nature of hyperproduction – a two-directional loop from simulation to reality (*sim2real*) and from reality to simulation (*real2sim*). Sim2real refers to the transfer of knowledge from the synthetic domain to the real domain (e.g., by training a self-driving car in simulation). By contrast, real2sim refers to the transfer of knowledge from the real domain to the synthetic domain (e.g., by creating a virtual influencer based on real-world data). The fact that those directions entail different techniques (Table 6.1) stresses the need to take seriously the licensing strategies of Unity Technologies and Epic Games.

The following two sections begin by explaining the respective factor of game engines' asset specificity on a general level. The sections then discuss its empirical manifestations by analysing patterns and differences in light of the use cases. Finally, the chapter extrapolates those insights beyond those two use cases.

### *6.4.1 Real-Time Physics and 3D Graphics*

A key objective of Unity Technologies and Epic Games has long been the optimisation of their video game engines to render 3D environments in real-time while ensuring the provision of realistic physics and graphics components. As Nicoll and Keogh (2019: 10) put it, game engines can automate 'computational tasks such as rendering, physics, and artificial intelligence, thereby freeing up developers to focus on 'higher-level' aspects of the design process.' This computation of physically accurate transformations of objects is highly relevant for improving a range of machine learning techniques, including deep generative models. To return to the above-mentioned example of *UnityEye*s, merely increasing the size of a dataset of synthetic eye images is not useful if the model trained on these images fails to recognise where a *real person* is looking. Along similar lines, if a robotic hand can only master a Rubik's Cube in a virtual, but fails to do so in physical reality[57], such a result might be less headline-generating.

Nevertheless, demonstrations of the potency of deep generative models, such as OpenAI's robot hand, point to the strategic rationale behind the uptake of game engines as tools for synthetic media generation. In this regard, is worth digging into the details of which computational techniques and logics underlie the use of simulations to fine-tune the parameters of deep generative models. Josh Tobin (2021), a former research scientist at OpenAI and involved in its robot hand work, articulates the advantages of resorting to simulated environments in the following way:

> Unlike robotic data, simulated data is super cheap, basically zero marginal cost. It's very fast, you can run simulators faster than real-time. And it's scalable, right? You can have a simulation running on every corner of your data centre... It's safe, right? So, you can't actually damage something by running a simulation – at least not yet. And you get labels for free: because you design the world, you know where all the objects are. You aren't beholden to real-world probability distributions.

This quote is crucial to grasp the role of game engines as infrastructure assets in hyperproduction. In his lecture on cutting-edge algorithmic techniques to underpin the sim2real transfer, Tobin (2021) refers to the example of training a self-driving car for an encounter with a unicyclist that wears a pink bodysuit – *an edge-case scenario*. There is a lack of training data with respect to this particular road user because the probability of such an encounter in the real world is low. Due to the lack of training data, the car's AI-powered autopilot might not be prepared to react appropriately if this unlikely encounter *does* happen. Using hyper-realistic synthetic worlds brought into being by video game engines, car manufacturers can design and simulate such scenarios and collect synthetic training data for AI systems that supplement their real-world training data.

[57] OpenAI's robot hand solved a Rubik's Cube that required 15 face rotations 60% of the time and one that required 26 rotations 20% of the time (Akkaya *et al.* 2019: 23).

Given those affordances, Unity Technologies and Epic Games are in a powerful position vis-à-vis car manufacturers. When Unity (2018) announced the creation of its automotive division, the firm counted 'eight of the world's top 10 Automotive Original Equipment Manufacturers as customers, helping improve the way they design, build, service and sell automobiles.' Epic Games (2022a) boasts that 'BMW, Ford, and Toyota are already harnessing the power of Unreal Engine to drive efficiencies and open up new creative avenues.' While services include areas such as prototype visualisation and assembly line training, the repurposing of game engines to train the neural networks of autonomous vehicles is the most intriguing phenomenon for this chapter. The higher the demand for such simulations, the more rent can be generated by the two dominant providers of game engine technology. If they remain the proprietors of this technology, an expansion of their market power is a likely scenario.

On a technical level, the progress in transferring machine learning patterns from synthetic to real-world environments has been significant in recent years. Concerning autonomous vehicles, game engines facilitate the combination of two approaches. The first approach is domain randomisation, a 'technique for training models on simulated images that transfer to real images by randomising rendering in the simulator' (Tobin *et al.* 2017: 23). The idea is that if there is sufficient variety in simulation, an AI model may perceive the real world as just another variation of it. OpenAI's robot hand, for example, relied on this approach. The second one is procedural generation, which we might imagine as the algorithmic, rather than manual, creation of content like roads and trees. Tesla (2021), for example, never publicly disclosed that it relies on the provision of game engines by Epic Games or Unity Technologies for its endeavour to become the leading manufacturer of autonomous vehicles. Nonetheless, its demonstration of recent work with simulations at the firm's AI day provides useful hints:

> Most of the data that we use to train [deep neural networks] is created procedurally using algorithms as opposed to artists making these simulation scenarios manually. The interactions between parameters such as curvature, varying trees, cones, poles, and cars with different velocities produce an endless stream of data […] We want to recreate any failures that happen to the autopilot in simulation so that we can hold autopilot to the same bar from then on. […] We have used 371 million simulated images with 480 million labels to train the neural networks of the car.

While the affordance of game engines' physics and graphics components to train autonomous vehicles lies in their power to render procedurally generated environments to simulate varying scenarios – a tool of Brute Force Creativity, so to say – their role in the fabrication of virtual influencers is rather different. Although the history of virtual influencers goes back to Japanese cyberpunk in the 1980s, today's computer-generated characters are inevitably embedded in the attention economy of social media platforms, primarily Instagram, YouTube and TikTok. Among other things, virtual influencers act as ambassadors for fashion brands and announce brand deals. A recent industry survey found that 58% of American survey respondents follow at least one virtual influencer, with 35% saying that they bought a product or service promoted by one (The Influencer Marketing Factory 2022). As a counterpoint to autonomous vehicles, virtual influencers are representative of the *other* direction of how hyperproduction operates: the transfer of knowledge from the real domain to the synthetic domain.

Consequently, the rationale for an uptake of game engines to render them is not the *randomisation* of parameters but instead the *tailoring* of parameters such as facial expressions, clothing, and gestures to consumer demographics and psychographics. In other words, the point is not to collect synthetic data in a randomised simulation but to use real-world data to create synthetic personalities that do not have to be realistic at all. Their value is defined by follower count and lead generation. In 2022, Lil Miquela, at the time the world's most successful virtual influencer, for example, had more than 3 million followers on Instagram and more than 3.5 million followers on TikTok. In 2020, Lil Miquela earned an estimated $11.7 million for her creators, a Los Angeles-based studio named Brud (Ong 2020). It is no secret that there are economic links between Brud and Epic Games. The music video of Lil Miquela's debut song, *Hard Feelings*, for example, was produced 'using Unreal Engine, while virtual cameras were used to set up shots that had a grounding in reality but that were also impossible in the real world' (Webster 2020). In the video, Lil Miquela performs acrobatic dance moves on the back of a lorry. After being chased by police cars, she ends up relaxing on the truck while it slowly disappears on the horizon amidst an endless field of flowers.

Virtual influencers like Lil Miquela illustrate that game engines are as integrated into the cultural life of hyperproduction as they are with its technical workings. The capability of game engines to fabricate hyper-realistic 3D worlds is as much a material force for computation (e.g., training neural networks) as it represents a cultural force for entertainment (e.g., rendering engaging music videos). How exactly does this circuit of transfer from the real domain to the synthetic domain and vice versa work in practice? To answer this question, we need to consider the programmability and modularity of video game engines as decisive factors of their low asset specificity.

### *6.4.2 Programmability and Modularity*

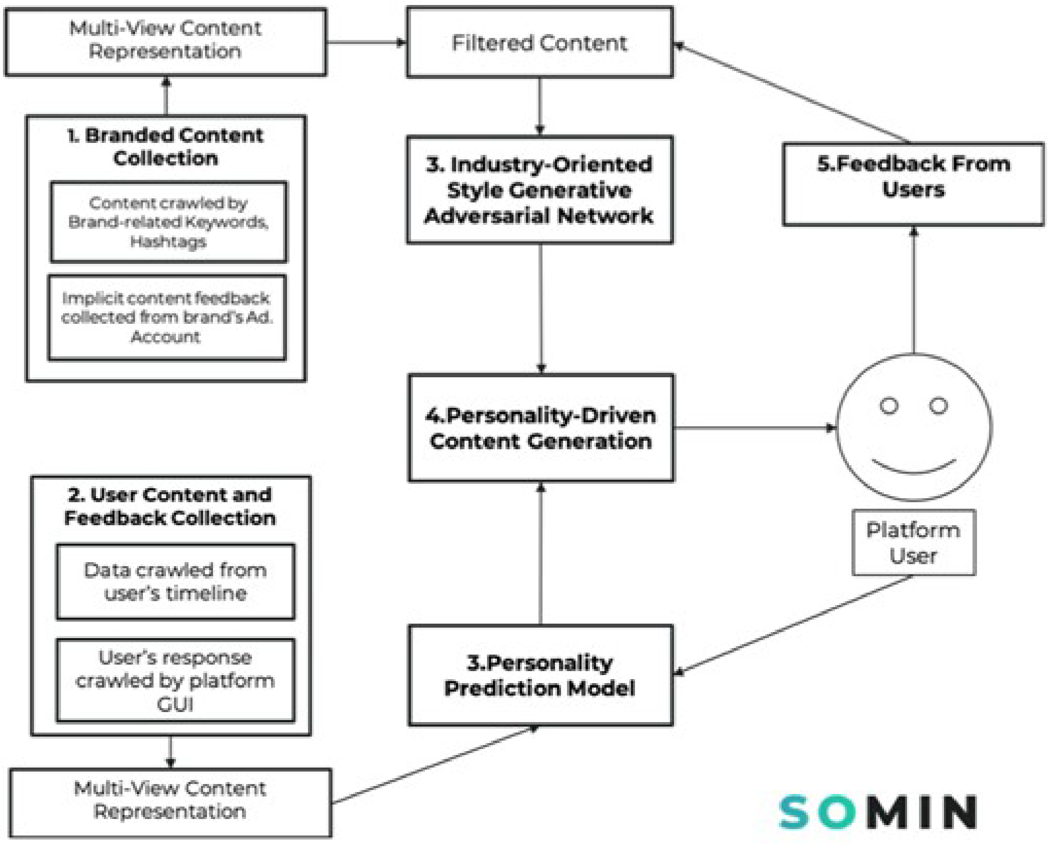
The term programmability refers to the fact that video game engines are 'capable of being programmed' (Chun 2008: 224) and can be used to create indefinite new digital objects, from synthetic eye images to a trailerless lorry traversing a red flower field. The term modularity means that the architecture of game engines is composed of two parts. First, there is a core with low variability (e.g., real-time 3D graphics). Second, there is a periphery with higher variability (e.g., sector-specific tools). Modularity, in other words, is a regime of high-level programming (McKelvey 2011; Russell 2012).

For example, Unity Technologies and Epic Games offer tools for architectural design, whose functionality is different from tools for military simulations. But some functions, such as real-time physics, apply across the board. As Nicoll and Keogh write (2019: 11), game engines are not 'discrete, unchanging, or totally black-boxed objects, but rather modular toolsets that can be customized to facilitate a variety of different design methodologies and project types.' This interplay between programmable and modular features transforms game engines into valuable infrastructure assets.

Naturally, much of the information about how those features manifest in the daily business of renting out game engines is hermetically sealed by non-disclosure agreements and trade secrets. For example, the exact amount of how much Epic Games benefits from Lil Miquela's viral success is unlikely to come to light. But this does not mean that we cannot attempt to comprehend the empirical function of programmability and modularity. Compared with the prominence of Lil Miquela, the Instagram follower count of Maya – a virtual brand ambassador for Puma in the Southeast Asia region – is rather low. But unlike Lil Miquela's creators Brud, Maya's originators are surprisingly transparent about their strategy in extracting and analysing consumer data to fabricate her. Maya is the result of the combined effort of the work of multiple firms coordinated by UM Studios x Ensemble Worldwide, an agency based in Kuala Lumpur (Lim 2020). One of the companies that co-created Maya is the Singapore-based SoMin AI. On its website, SoMin AI does not only showcase its products to 'supercharge and automate your digital ad performance' but also technical demonstrations of what its data-driven virtual influencer production pipeline actually looks like.

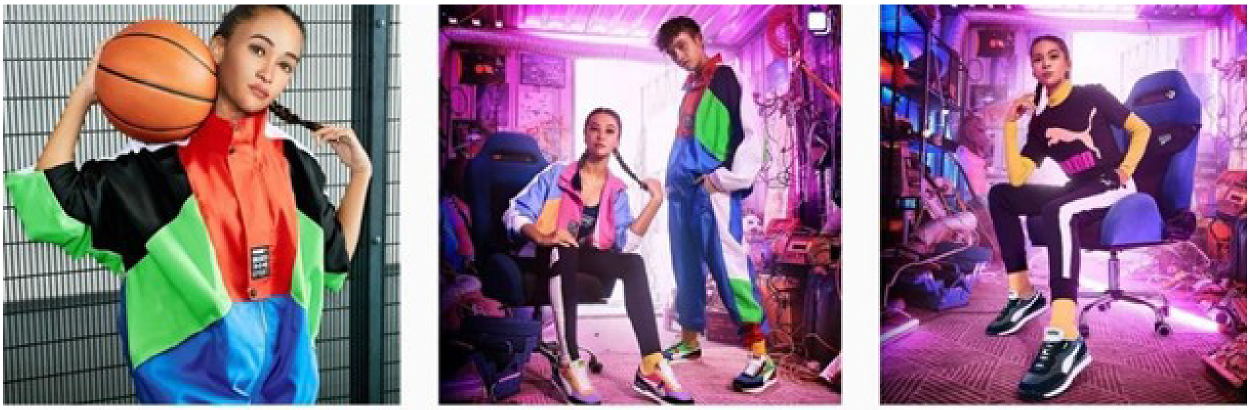
SoMin AI 'combines deep multi-view personality profiling framework and style generative adversarial networks facilitating the automatic creation of content that appeals to different human personality types' (Farseev *et al.* 2021: 890). In other words, the company uses a person's social media content to infer their personality type and then creates tailored social media content based on the preferences of people with the same personality type. It utilises the Myers-Briggs Type Indicator, an instrument that was developed in the 1940s. These days, it is seen by most personality psychology experts as 'little more than an elaborate Chinese fortune cookie' (Hogan 2006: 28). SoMin AI eventually combines this automated personality type inference with retrieving the best-performing brand content among users with a similar Myers-Briggs personality type to train a deep generative model for producing content (Figure 6.4).

Figure 6.4. Real2sim influencer production pipeline. Source: Farseev *et al.* 2021.




For example, in the case of Maya, the company 'mapped millions of faces in SEA from multiple online sources', including Instagram, and used this data to render 'several face versions, the first building block to creating a virtual person' (Lim 2020). Subsequently, other companies built on this groundwork and fabricated Maya's 3D face model (Figure 6.5) and refined her personality traits, which indicate a robotic passion for Puma products – the epitome of the post-digital (Berry & Dieter 2015).

Figure 6.5. Maya, a virtual influencer for Puma in the SEA region. Source: Instagram.



Given her roots in the extraction of real-world user data, Maya is an apt example of the real2sim transfer in hyperproduction. Compared to other 3D animation software, Unity Technologies and Epic Games do not only promise to improve the rendering of virtual influencers like Maya but also the ease of fine-tuning details. In February 2021, Epic Games (2021) announced MetaHuman Creator, a browser-based app that enables 'creators of 3D content to slash the time it takes to build digital humans from weeks or months to less than an hour, while maintaining the highest level of quality.' Similar to Google Colab, the platform used by artists to modify ruDALL-E, MetaHuman Creator runs in the cloud, so that its computational processes take place in data centres. After having created a character in the browser, users can then export it to Unreal Engine for animation and motion graphics. The crux is that characters developed with MetaHuman Creator are licenced for use *only* with Unreal Engine-based products.

Although the tool is free to use, it illustrates the consolidation of a proprietary ecosystem. As Birch and Muniesa (2020: 24) put it, 'as content, in whatever form, has become almost costless to reproduce, companies and individuals have turned to IPRs [i.e., intellectual property rights] to secure their profits.' For asset holders in the era of hyperproduction, the modular structure of video game engines is perfectly suited to prevent users from switching providers as they allow for the technical integration of trademarks. In the license agreement for MetaHuman Creator, Epic Games (2022b) legally ensures that users are not allowed to ‘remove, disable, circumvent, or modify any proprietary notice or label’ included in the 3D characters. If the virtual influencer industry expands, game engine providers may thus siphon off profits from the likes of Lil Miquela (quite literally) as long as they underpinned their *initial* creation.

The case of autonomous vehicles underscores that the role of programmability and modularity is not limited to intangible outcomes. As tools to facilitate the sim2real transfer, car manufacturers leverage game engines in the pursuit of improved physical and tangible outcomes. Unity Technologies (2021: 11-13), for example, showcases its provision of game engine tools to Volkswagen Group of America. Although it is vital to be aware that such reports are designed to promote the product in question (i.e., Unity’s game engine), I hold that they still entail insightful evidence:

> We use image and ground-truth data generation in Unity to train neural networks for the implementation of autonomous driving components such as sensors, perception, prediction, and driving. […] For some projects with limited scope, we use synthetic data as a cheaper data-generation option, whereas projects with broader goals take advantage of a mixture of real-world and synthetic data. […] Unity’s vast variety of assets makes it possible to design desired scenarios.

In this example, Unity’s game engine is embedded in the planetary-scale value chains of a multinational car manufacturer, including its R&D divisions and real-world fleet data. Once a company like Volkswagen successfully uses Unity products in terms of a domain such as synthetic data generation, Unity can then set up modular pipelines between its services. Unity Perception, for example, is a software development kit built on top of Unity’s 3D rendering engine that aims to radically simplify the combination of synthetic data generation and domain randomisation. While Unity Perception is free to use, Unity's developers explicitly stress its 'easy-to-use built-in support for running dataset generation jobs in the cloud using Unity Simulation, making it possible to generate millions of annotated images' (Borkman et al. 2021: 2). Due to the scale of clients in the automotive sector and their astronomic amounts of training data, providers of game engine technology have an interest in being the *only* provider of such services. At the same time, the latter benefit from the collection of data that can be used to refine their offerings and tailor them to the needs of clients. Data is 'used to grow the value of assets' (Sadowski 2019: 6) – in this case, digital infrastructure assets that lay bets on the computational imaginary of an industrial metaverse (Lange 2022).

Another consequence of the programmability and modularity of game engines is the cultural habituation of firms in non-gaming industries into their modus operandi. In this context, experience with game engines has even become a hiring criterium, both within the automotive sector and the virtual influencer industry. For example, in a job advertisement to join the company as an Autopilot Simulation Engineer, Tesla (2022) adds 'experience working with Unreal Engine 4' as the desired requirement. Along similar lines, Lil Miquela's creator studio Brud (2020) looked for candidates that would 'work on character-driven projects and stories with the real-time production team in the Unreal Engine.' The document analysis of this chapter underlines that

> the industrial migrations of game engines (across entertainment forms, 3D imaging, prototyping and manufacturing, scientific visualisation, architectural rendering, cross-platform mixed reality, and within the military) suggest they have broad power for organising the cultural field (Freedman 2018: 237).

To migrate into non-gaming industries, Unity Technologies and Epic Games use an economic strategy that they have in common with technology giants like Alphabet and Amazon: intrafirm cross-subsidisation (see Chapter 4). This means that both firms strategically accept (temporary) losses in certain domains (e.g., high initial costs for setting up a tool such as MetaHumanCreator) in light of their skyrocketing profit margins in other business domains (e.g., micro-transactions in Fortnite). Such profits also facilitate the acquisition of start-ups. According to Crunchbase (2022), Epic Games has made a total of 18 acquisitions, including a Serbian company that helped to build MetaHumanCreator. Unity Technologies has acquired 25 companies, most recently a Canadian company specialising in the production of digital humans. Although the analysis of how cross-subsidisation and such mergers shape the trajectories of synthetic media and how such dynamics affect the political economy of platforms (van Dijck et al. 2018) is a task for future research, we can already see signs that economic power and cultural power go hand in hand in an era of hyperproduction.

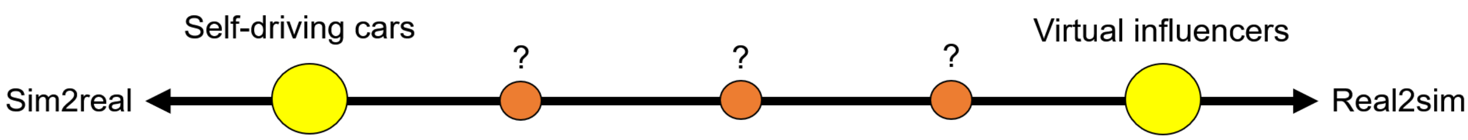
### 6.5 Conclusion: The Hyperproductive Condition

This chapter argued that the penetration of cultural life with deep generative models is neither a temporary nor a coincidental phenomenon. We witness a transformation of the nature of platformised cultural production. Seen this way, the corporate imaginary of an 'industrial metaverse' (Lange 2022) is not a proposal to re-write reality, but rather a discursive attempt to push more production into a rentier relationship with proprietary reality engines. What is at the stake here is not simply the dissolution of reality and simulation in a philosophical sense but the intensification of rentier relations across cultural life. Hyperproduction, in this sense, can be seen as describing an 'increased or excessive production' (Collins Dictionary 2022) of dependencies between owners of infrastructure assets and those who rely on them on a day-to-day basis.

As a result of this, a primary conclusion is that actors that control the necessary hardware and software to deploy hyperproduction, including Unity Technologies and Epic Games, are likely to expand their economic and cultural power as more industries will come to adopt deep generative models in their daily work. This conclusion entails a number of important empirical, political, and theoretical implications.

On an empirical level, the main conclusion is that hyperproduction can serve as an impetus for a wide range of future empirical case studies. This chapter has focused on a certain facet of this idea: the emergence of video game engines as infrastructure assets. The juxtaposition of autonomous vehicles and virtual influencers showed that although game engines are used in very different ways, two factors of asset specificity can be clearly identified: real-time physics and 3D graphics, and programmability and modularity. While the point of hyperproduction with respect to autonomous vehicles is to *randomise* the parameters of synthetic worlds as a way to optimise artificial neural networks for their real-world use in cars, the case of virtual influencers neatly illustrates the logical inversion of this process: the *tailoring* of influencers' synthetic parameters based on the real-world social media data of users. Consequently, these cases represent prototypical examples of two opposed directions of hyperproduction: that is, simulation to reality (sim2real) and reality to simulation (real2sim).

Figure 6.6: The empirical spectrum of hyperproduction.




Against that backdrop, it is also conceivable that other use cases are positioned in between those paradigmatic examples. Although it is paramount to be wary of the sensationalistic rhetoric of industry actors aimed at attracting venture capital, it is not an exaggeration to anticipate the emergence of creative manifestations of hyperproduction in the years to come. In other words, the importance of synthetic environments to fine-tune the parameters of artificial neural networks is likely to increase. Gartner (2021b), a market research firm, predicts that by 2024, 60% of the data used for the development of AI will be synthetically generated. This means that in theory, firms can bypass the reliance on real-world data and circumvent data privacy regulations, such as the EU's General Data Protection Regulation (GDPR). As the concluding chapter of this thesis argues, the possibility of such modes of regulatory circumvention raises intriguing questions about the future governance structures of AI-generated media.

On a theoretical level, far from creating a multitude of simulated worlds that we cannot grasp with the toolkit of social theory, synthetic media remain firmly grounded in, and bounded by, reality – not only in the sense of training data (Offert 2021) but also in the sense of political-economic realities. As a cultural phenomenon, hyperproduction calls into question established modes of thinking about what AI-generated media is and how it can (or should) be governed. For example, if deep fakes and geopolitical issues of misinformation remain the central point of reference that shapes our thinking of how AI impacts media production, the types of research questions and arguments that follow from this entry point would remain restricted by it. By contrast, a more imaginative and creative analytical repertoire to make sense of AI-generated media helps go beyond an excessive focus on *media content* (i.e., types of misinformation) in favour of a nuanced structural analysis of changes in *media production*.

This expanded perspective opens up fruitful new ways to analyse and theorise the cultural ramifications of AI's geographies of production. While much attention has been given to the shipping container as both a physical and digital object that enables supply chain capitalism (Klose 2015; Tsing 2009), this chapter has unearthed another kind of imbrication of capitalism and worlding. Not only a continuation of a political economy over what companies control global operating systems (McKelvey 2018) but an attempt to shift the terrain of media production to reality itself. Hyperproduction is also a possibility for production building on Rossiter's (2014) idea of 'logistical worlds.' Hyperproduction refers to the attempts of firms to create meta-logistical worlds upon a subscription basis to a new generation of downstream producers, competing on a level planning field between human and synthetic talent. In the next chapter, I expand those ideas by relating hyperproduction to the other forces of neural production. For example, instead of existing in an economic or political vacuum, hyperproduction also relies on computational resources and is affected by regulatory frameworks.

In appreciation of the circular nature of hyperproduction, it is worth returning to the chapter's beginning as it now comes to an end. As more digital content is produced on top of already-hyperproduced content (e.g., the fake Mona Lisas), a consequence of this development might be the homogenisation and convergence of media production overall. The point of the Mona Lisa is that there is only *one*, not infinite AI-generated imitations. But if synthetic outputs soon become real enough to be seen as the training grounds for deep generative models, models that produce better synthetic media, that in turn transform into the new training grounds, hyperproduction would trigger an endless spiral that recycles archived content. While the cultural ramifications of such a spiral of synthetic media would not be dramatic in hermetically sealed environments such as the training of autonomous vehicles, they could potentially have detrimental effects on the deliberative and democratic conventions of the mediated public sphere.

With respect to this latter point, Baudrillard's (1994) work on hyperreality is helpful to problematise the role of media industries and simulations in shifting social, economic, and cultural life. But unlike hyperreality, hyperproduction is not meant to be a superstructure hovering *above* the capitalist system – it grasps *how* capitalist systems co-constitute reality. Although we cannot foresee the future worlds of synthetic media, we can strive to develop a theoretical repertoire to imagine them.

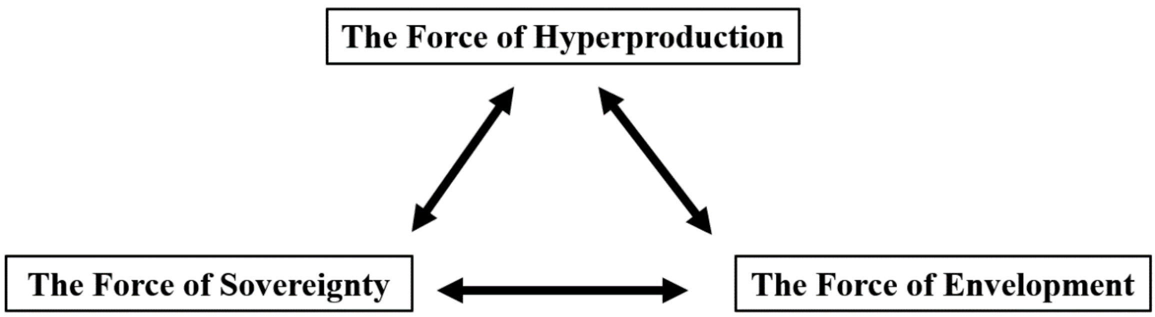
# Chapter 7: Conclusion: The Forces of Neural Production

## 7.1 Introduction: A Triangle of Tensions

> AI systems are built to see and intervene in the world in ways that primarily benefit the states, institutions, and corporations that they serve. In this sense, AI systems are expressions of power that emerge from wider economic and political forces, created to increase profits and centralize control for those who wield them.

Thus begins Crawford's (2021: 211) conclusion of her book, *Atlas of AI*. In this thesis, I introduce an original theoretical framework that helps to systematically analyse those 'wider economic and political forces' from which AI systems emerge: the idea of neural production networks. But in contrast to Crawford's terminology, I do not wish to restrict my focus to the forces *that enact* AI's geographies of production, like technology firms and their powerful computational infrastructure. Beyond that, I also elucidate the forces that are *enacted by* AI's geographies of production, such as the development of new regulatory frameworks and the sophistication of AI-generated media.

Figure 7.1: The forces of neural production.




The purpose of this concluding chapter is to wrap up the thesis by encapsulating the relations and tensions between three of those mutually conditioning forces: the force of envelopment, the force of sovereignty, and the force of hyperproduction (Figure 7.1). While each of the previous three chapters focused on answering research questions that foreground different facets of the central research problem, it is paramount to conclude by synthesising these insights. As MacQuarrie (2010: 442) writes, a feature of holistic case studies is that 'the various bits of empirical materials that comprise the case[s] cannot be interpreted without allowing for the synergy of the whole.' But crucially, as explained in Chapter 3, this does not imply that I wish to argue in favour of a static or fixed 'transcendent unity' (Saldanha 2003: 427) between those forces. Instead, I use the metaphor of a triangle of tensions to illustrate that those forces are *fluid* rather than *pre-defined* categories. As such, the dynamic ways in which they relate to each other are in a continuous state of flux. New forces may emerge, and existing ones may transduce into something else. Nonetheless, I hold that we can know more about those forces if we see them as an aggregated movement. Holism does not mean statism.

Because of the insinuation of AI into countless economic, political and cultural interactions and transactions in every corner of the planet, neural production networks already infiltrate infrastructural life. But beyond their empirical manifestations studied in this thesis, expounding a new understanding of those distinctly neural geographies of production also motivates a range of questions and directions for future research. This is the starting point, not the end, of making sense of the forces of neural production. Much more is there to investigate, scrutinise, and discover if we take seriously the argument of this thesis that established ways to imagine border-crossing economic networks cannot account for the historical discontinuities of a world shaped by artificial neural networks, deep generative models, and synthetic environments. The omnipresence of such models in enabling and constraining our perception and experience of time and space is not a hypothetical future scenario, but a realistic assessment of the here and now. Therefore, opening out a new conceptual space to probe how the forces of neural production are related to each other is not a futile endeavour, but an analytical necessity – an impetus for tracing connections, finding patterns, and imagining alternatives.

To unfold this argument, this chapter combines a summary of key findings and insights from previous chapters with notes about fruitful avenues for future research that builds on the conceptual groundwork of this thesis. However, my goal is not to simply enumerate all potential ways in which the idea of neural production networks could be taken up but rather to clarify productive intersections with other strands of enquiry. Far from situating the arguments developed in this thesis within one particular, hermetically sealed disciplinary tradition, method, or conversation, I explicitly wish to think across and beyond such socially constructed boundaries. As those forces of neural production do not have a will of their own that separates them from the actors that bring them into being, they do not float within an elevated space that is closed off from other debates, ideas, and arguments. Rather than teleological orthodoxy, I encourage reinterpretation, repurposing, reconfiguration, and, last but not least, rebuttal.

The chapter begins by problematising the tensions between hyperproduction and envelopment. Because lead firms in neural production networks, by definition, envelop the space within which production takes place, they also envelop the space within which hyperproduction takes place. Given the overlaps between proprietors of envelopment and proprietors of hyperproduction, new power struggles are on the horizon. Extending those considerations, I analyse the tensions between hyperproduction and sovereignty, arguing that the governance of synthetic media is a paramount matter of concern, both analytically and practically. The third tension is to be found between sovereignty and envelopment. As long as the means to deploy cutting-edge artificial neural networks remain in the hands of just a handful of technology firms, policy attempts to regain sovereignty over neural production networks will be constrained in their regulatory leverage. Ultimately, the chapter returns to Carnegie's *The Gospel of Wealth*, ending this thesis by reconsidering the analogy with which it started.

### *7.1.1 Envelopment vs. Hyperproduction*

Although its first version was released in July 2021, the text-to-image model DALL-E mini only became a viral internet phenomenon in June 2022. At that time, AI-generated images originating from DALL-E mini were so omnipresent on Twitter that it was hard to avoid them. *Weird DALL-E Mini Generations*, a Twitter account showcasing text-to-image outputs, attracted hundreds of thousands of followers within a few days. More or less imaginative text queries like "Winnie the Pooh as Golden Buddha Statue" or "Bart Simpson Depicted in Egyptian Art" yielded impressive results, both in terms of their visual outputs and the number of likes that those outputs received on Twitter. However, what remained absent from the enthusiasm and excitement that naturally coincides with any new and trending internet phenomenon was the way in which those mind-boggling images were computationally generated in the first place.

Even though the process of text-to-image creation only takes a few clicks thanks to a simple user interface, images generated by DALL-E mini are not constructed in a void. Like other manifestations of neural production, DALL-E mini is a direct result of computational envelopment. Lead firms in neural production networks figuratively sit beneath the simultaneous *sophistication* and *simplification* of synthetic media. Just because DeepMind and OpenAI are dependent on the infrastructural provision of game engine technology by Unity Technologies and Epic Games (see Chapter 6) this does not mean that the force of envelopment loses traction or becomes obsolete. However, what this partial dependency *does* mean is that a likely future scenario will involve clashes and power struggles between proprietors of computational envelopment and proprietors of the means to foster and accelerate hyperproduction[58].

[58] This is not to suggest that those two categories of proprietors are mutually exclusive or that there are no power struggles or clashes amongst proprietors of the same type.

To clarify this point, let us consider the infrastructure[59] that underpins DALL-E mini. Work on the model began in June 2021, as part of a community event for machine learning experts. As one of the model's developers explains in a blog post: 'Google sponsored the community event and allowed us to use a TPU v3-8 for training. After the event, they provided us with free access to several TPU instances to keep working on the project' (Cuenca 2022). In other words, Google's proprietary hardware did not only facilitate the initial fine-tuning of DALL-E mini's neural parameters but also its day-to-day fabrication of tens of thousands of synthetic images. As Cuenca (2022) specifies, 'we offload all requests to our backend infrastructure. This allows us to take advantage of the TPUs Google lent us.' Organisationally, Google sets up partnerships like that through its "TPU Research Cloud" programme, whereby researchers have to submit their ideas for being considered for free access to computing resources.

This brief vignette illustrates that there is no artificial line between envelopment and hyperproduction. Instead, those forces of neural production are inextricably related. In this thesis, I argue that computational envelopment is the key mechanism by which technology giants such as Google enact AI's geographies of production. Theoretically, my analysis of envelopment draws on the premise that space should not be defined as being restricted to discrete, fixed, or stable horizontal territories that can be mapped on a

[59] Taking seriously Larkin's (2013: 330) argument about the role of 'categorizing moments' in the conceptualisation of infrastructures requires transparency about my choice to prioritise the study of computational resources over the study of training datasets. A compelling case could be made for the key infrastructural role of training data in hyperproduction. DALL-E mini, for example, was trained on about 15 million labelled images. A set of intriguing research questions for future studies relates to the spatialities those training datasets shape future trajectories of hyperproduction. As Graham and Dittus (2022: 161) put it, 'machine learning […] learns by looking in the past. And within that past are contours, biases, and geographies […] that will continue to shape even the most frontier technologies.'

canvas. Rather, speaking of envelopment accentuates the deeply *processual* character of the space within which neural production takes place. The key question, then, as Kitchin (2009: 272) puts it, is not 'what space is' but rather 'what space becomes.' Inspired by this ontogenetic and relational understanding of space, Chapter 4 identifies a number of patterns of how lead firms give aim to give it 'its form, function, and meaning through practice' (Kitchin 2009: 272). Drawing on global production network scholarship, the chapter interprets "practice" as "modus operandi as a lead firm." Its main finding is that, unlike lead firms that centralise the production of a physical commodity, lead firms in neural production networks decentralise AI production and centralise AI provision, by virtue of controlling and optimising proprietary hardware and software.

Against this backdrop, it is absolutely in line with the strategic rationale of lead firms to extensively subsidise the production of synthetic media by sponsoring and supporting developer communities, such as in the case of DALL-E mini. As Google benefits from the widespread adoption of its tools and services, there could hardly be a better viral marketing campaign for its proprietary TPUs. Even though Google's infrastructural role largely remains absent from media reports on this topic, it does not remain absent from developer communities and start-ups that perceive the domain of synthetic media as a growing and lucrative market. Undoubtedly, models like DALL-E mini showcase the creativity of developers. But they are also demonstrations of the potency of Google's infrastructure-as-a-service offerings.

Although Chapter 6 – the one that addresses the force of hyperproduction – does not explicitly resort to economic-geographical concepts to underpin its claims, it shares similarities with the processual (or ontogenetic) approach of Chapter 4. At the centre of analysis in Chapter 6 is not the provision of computational infrastructure, but rather the penetration of cultural life with deep generative models – a process that I theorise as hyperproduction. Crucially, the similarity between those chapters lies not only in the fact that they both relate to different facets of neural production networks but also in that my actual focus is not on what the cultural ramifications of AI-generated media *are* (i.e., in a static sense) but rather what they *become.* In capturing this approach by using a circular metaphor that gravitates between simulation and reality, Chapter 6 stakes out two ends of the wide empirical spectrum of hyperproduction. One end of this spectrum, exemplified by the case of virtual influencers, is indicative of a cultural transfer from the real domain to the synthetic domain. Virtual influencers are fabricated on the basis of real-world social media data. By contrast, the other end of this spectrum, illustrated by the case of autonomous vehicles, shows a cultural transfer from the synthetic domain to the real domain. This is because video game engines are being deployed to optimise a car's artificial neural networks in synthetic environments. Chapter 6 can be seen as an allegory of inverted cultural flows between simulation and reality.[60]

But when it comes to the clashes between envelopment and hyperproduction, so much more is there to be studied. And so much more will unfold. For example, what happens if the Microsoft-backed research lab OpenAI starts to licence out its extremely powerful DALL-E model to clients? Will it be adopted by media organisations around

[60] Of course, I acknowledge that there is a stylistic rationale behind this argument. The boundary between simulation and reality is not fixed or clearly defined, as Baudrillard (1994) lucidly reminds us. Even when it comes to technical dimensions, it is not possible to state whether a transfer happens from simulation to reality or from reality to simulation. When it comes autonomous vehicles, for example, there is a constant back-and-forth between data that is generated in synthetic environments and data that is generated in on physical streets, with real-world encounters. The same back-and-forth argument applies to the virtuous feedback loops of virtual influencers. Nevertheless, I hold that there is an immense value in crafting Chapter 6 in this way: not as an example of a Cartesian dualism, but rather as a strategy to delineate a flexible and dynamic spectrum of empirical possibilities.

the world as, a convenient way to automate the production of headline images on news sites? As a result of this, will the currently dominating proprietors of computational envelopment also become dominant proprietors of hyperproduction – without having to rely on smaller firms for integrating deep generative models into platformised cultural production? Will Epic Games, which is partially by the Chinese digital giant Tencent, outcompete Unity Technologies and eventually become a monopolist[61] on game engine technology? Those questions just represent some examples to demonstrate the utility of conceptualising neural production networks in such a way that the idea integrates with a number of economic, political, and cultural developments.

With respect to future research in this area, I see a particularly strong potential for integrating political-economic approaches to the study of platforms (van Dijck 2020; Helmond 2015) with geographical thought on rentiership (Birch 2020) in order to trace the evolution of AI-generated media. Such a cross-disciplinary approach would be well-placed to disentangle the tensions between discursive aspirations and techno-optimistic imaginaries (e.g., the metaverse), and material realities of uneven economic outcomes. For example, a concrete object of study could be OpenAI's large-scale language model GPT-3, which the company offers on a subscription basis to users, similar to how lead firms provide access to computational resources. However, the difference from what I described in Chapter 4 is that the provision of such pre-trained deep generative models is still in its early stages. Although it is difficult to predict whether models like GPT-3 or DALL-E will find their way into the daily business of leading media organisations, mapping this emerging ecosystem is an auspicious task.

[61] Srincek (2016: 76) argues that, because platform-powered businesses rely on similar modes of cross-subsidisation and network effects, competition between those firms 'drives them to enclose themselves increasingly.' Indeed, such a situation would be likely outcome of the 'great platform wars' (Srincek 2016: 60) between game engine providers.

In the next section, I maintain this spotlight on thinking through the empirical and policy implications of the force of hyperproduction by putting it into conversation with the force of sovereignty: the second key tension of this chapter.

### *7.1.2 Hyperproduction vs. Sovereignty*

Without a doubt, sovereignty is a highly ambiguous term. As Chapter 5 demonstrates, it can be used to serve all kinds of interests: those of German policy-makers that aim to strengthen domestic firms in light of the challenges posed by foreign cloud providers; of said providers that reinterpret it in ways so that it fits into their lobbying strategy; and of European bureaucrats that perceive it is a one-size-fits-all buzzword for all digital policy debates. Precisely that is why I see it as a force in its own right.

The term's discursive ambiguity is not so much a weakness but rather a strength. As Coleman (2009: 255) aptly puts it, 'sovereignty is literally about the setting of limits to social practice […] via epistemological appeals about how best to know and act in the world.' Therefore, she argues, we must not necessarily find and use a singular definition for the term. The key task is to 'understand how, for what ends, and in what contexts the term is used and expounded upon, as well as attend to the work done by it' (Coleman 2009: 255). Inspired by this suggestion, Chapter 5 provides a starting point to provide such an understanding in the context of neural production networks. Specifically, the chapter examines the policy discourse around European digital sovereignty in order to theorise the positioning of state actors and state action within AI's geographies of production. However, there are also intersections between the force of sovereignty and the force of hyperproduction. I outline them in this section.

The most visible friction between those forces can be interpreted as the much-discussed phenomenon of deep fakes – an empirical manifestation of a vaguely defined semantic superstructure of "misinformation." Not unlike the idea of "sovereignty", the term “misinformation” can be used for all kinds of purposes, by all kinds of actors. For example, in the wake of the Russian invasion of Ukraine, Russian state media labelled Western media outlets as propagators of misinformation, while the same term was used by Western media to describe the day-to-day operations of Russian state media. But as Chapter 6 shows, the underlying technology of deep fakes – deep generative models – is by no means restricted to the cultural practice of manipulating videos. Deep fakes are merely the tip of a much bigger, and largely unexplored, iceberg: the integration of deep generative models into the ‘platform society’ (van Dijck et al. 2018).

Consequently, in addition to the challenges of regulators posed by lead firms in neural production networks, debates surrounding digital sovereignty will gradually have to get to grips with the crucial task of synthetic media governance. By this, I mean the regulatory, institutional, and organisational checks and balances that will be put in place to ensure that the spiral of hyperproduction does not spin out of democratic oversight and control. This task involves analytical, political, and technical challenges. On the analytical level, there is an urgent need for more empirical evidence on the different manifestations of synthetic media and the ways in which deep generative models are deployed to create content, deceive users, or facilitate digital advertising. Mapping, understanding, and explaining the proliferation of those models into more and more spheres of the digital economy is not something that can be done within the scope of a singular thesis or research project – it is a collective effort that requires interdisciplinary collaboration. What I am describing here is not a speculative scenario. Firms around the world are already relying on synthetic training data to supplement real-world data and fine-tune the parameters of their artificial neural networks, circumventing and bypassing existing regulatory frameworks such as the EU’s General Data Protection Regulation (GDPR). This capability coincides with two political challenges.

First, I agree with Steinhoff (2022) in that it is pivotal to go beyond regulatory agendas that are purely derived from *privacy-oriented* concepts[62], like Zuboff's (2019) influential notion of surveillance capitalism. At its core, Zuboff's argument relies on the key premise that the extraction, accumulation, and analysis of real-world user data is the most important competitive advantage for technology companies. However, as Chapter 6 argues, the ability of Unity Technologies and Epic Games to dominate the worldwide market for video game engine technology by setting up licencing strategies that result in a quasi-oligopolistic industry structure calls into question an exclusive focus on privacy and data. If more training data stems from synthetic sources, a focus on user data as a regulatory point of leverage is likely to be less impactful (see also Srincek 2020a). In any case, there is an urgent need for more empirical evidence to grasp and problematise the potentially anti-competitive nature of rentiership (Birch 2020) when it comes to the provision of infrastructure assets such as video game engines. But if synthetic media is purely presented as a problem of misinformation within regulatory debates, such hidden oligopolistic tendencies would remain unexplored and undiscovered.

Second, future regulatory frameworks on digital sovereignty must consider the fact that the proliferation of AI-generated media fundamentally transcends cultural and media industries. At the time of finishing this thesis, most legislative frameworks and emerging policy proposals do not adequately account for the cross-sectoral role of synthetic media. The European Commission's Artificial Intelligence Act (featured in Chapter 5), for example, includes a provision to disclose the artificial nature of media content that resembles 'existing persons, objects, places or other entities or events' and

[62] Admittedly, this is not so much an "no, but" argument but rather a "yes, and" argument. Of course, data privacy regulations are highly important. However, there is a danger in using the extraction of user data as the one-and-only explanation of uneven economic outcomes in the context of AI – and as the key mechanism to regain digital sovereignty.

would falsely appear to a person to be authentic' (Vaele & Borgesius 2021: 108). But this disclosure obligation legally falls on the *users*, not the *providers*, of synthetic media software. In other words, companies that provide the infrastructural tools without which certain forms of synthetic media would not exist in the first place would largely remain unregulated due to how the initial proposal of the EU's Artificial Intelligence Act was drafted. A side effect of the EU's justified focus on technology giants – demonstrated by the Digital Markets Act – might be a lack of legal accountability for smaller firms, such as Unity Technologies of Epic Games. If everyone else can hide in the shadows of technology giants, then that is not a desirable policy outcome. A spotlight on technology giants should not go hand in hand with a neglect of less visible issues.

On the other hand, there are already examples of what *provider-centric* modes of synthetic media governance could look like. In 2022, the Cyberspace Administration of China unveiled a draft bill that aims to regulate deep synthesis services that are defined as 'technologies using generative sequencing algorithms to make text, images, audio, video, virtual scenes, or other information' (China Law Translate 2022). While it needs to be noted that this is not an example of *democratic* oversight, it still remains a useful example of oversight. What the draft bill shows is that, without some sort of conceptual abstraction[63], it is not possible to transform the elusive phenomenon of synthetic media

[63] As McCormack (2012: 722) describes it, 'abstraction makes complex materials available for manipulation and management.' Therefore, it is important to understand 'way in which the abstraction articulates the relation between complex materials and institutions designed to regulate these materials' (McCormack 2012: 722). This argument resonates with Law's (2004: 141) claim that the key issue of method 'becomes how to make things different, and what to make.' In probing the relationship between the force of hyperproduction and the force of sovereignty, I connect an abstraction that renders intelligible a set of complex materials to an abstraction that renders intelligible a set of complex institutions designed (but not always adequately prepared) to regulate these materials.

into a regulatable set of practices, processes, and relations. In this case, the Cyberspace Administration uses an elaborate definition of deep synthesis services that takes into account the crucial technological role of generative sequencing algorithms – a different term for deep generative models. The algorithm becomes a 'regulatory object' (Seyfert 2018) in its own right, rather than remaining a black-boxed entity. By establishing the legal category of 'deep synthesis service providers' (China Law Translate 2022), the proposed draft bill exhibits a neat infrastructural sensitivity.

Against this backdrop, I hope that the notion of hyperproduction to designate the penetration of cultural life with deep generative models will be a helpful reference point for scholars, regulators, activists, and other actors engaged in the emerging domain of synthetic media governance. Ideas emanating from fields such as media theory have a place in these debates, illustrated by the uptake of the platform ecosystem metaphor in the Digital Markets Act (see Chapter 5). In this sense, I see conceptual abstraction as a prerequisite for effective regulation – not as the *only* requirement, but an important one. Without adequately representing the complexity of AI-generated media (i.e., by making it intelligible to a broader audience), its cross-sectoral role and implications will not find the appreciation that it requires, within both academic and policy circles. Because of its conceptual elasticity, the idea of neural production networks offers a useful framework to link those debates to broader conversations around the regulation of AI.

Beyond analytical and political challenges that go hand in hand with grasping the multifaceted relationship between sovereignty and hyperproduction, there are also technical challenges that complicate the transformation of AI-generated media into a 'regulatory object' (Seyfert 2018) – or, rather, a *regulatable set of objects*. Given that deep generative models continuously gain sophistication and complexity, it becomes difficult to identify synthetic media in the first place. If there are no reproducible ways to separate synthetic media content from real content, how should regulatory agencies go about enforcing new legal obligations? Developing policy proposals is one thing, but actually enforcing them brings about veritable practical challenges. But I do not wish to imply that those issues cannot be solved. For example, researchers were able to track the identities of real people whose faces were used as part of training datasets to optimise the popular *This Person Does Not Exist* model – a website that creates a hyper-realistic synthetic face each time it is opened in the browser. As Knight (2020) explains, 'the fake faces can effectively unmask the real faces the GAN [i.e., generative adversarial network] was trained on, making it possible to expose the identity of those individuals.' As is the case with many technological phenomena, such a capability to relate synthetic faces to real faces can be effective for both regulatory purposes (i.e., to identify whether an image is real or not) *and* subversive purposes[64] (i.e., data breaches).

To summarise, the key tension between the force of sovereignty and the force of hyperproduction is that the scope of the former will have to account for the reach of the latter. Due to the ambiguity of "sovereignty", this might not be a discursive challenge. But it will be a challenge on the regulatory level. How can we reimagine the design and governance structures of synthetic media in ways that prioritise human rights, social justice, fair competition, and democratic values? Future research would be well-advised to explore this sweeping question, juxtaposing rich empirical evidence on the actually-existing economics and cultural practices of hyperproduction with regulatory debates that are supposed to provide oversight for AI-generated media. Such empirical enquiries should go beyond an analysis of the status quo and also envision alternative trajectories

[64] In this regard, the regulatory relation between hyperproduction and sovereignty could also be explained through the lens of deep fakes, whereby a similar double logic applies. But due to the fact that I argue in favour of going beyond an obsessive focus on misinformation as the entry point for AI-generated media, the synthetic face example is more suitable.

of how deep generative models could be used and governed, imagining and conceiving multiple futures of how synthetic modes of production could serve the common good, rather than undermining trust and solidarity. Even if those futures will not materialise, they could act as a normative benchmark of what *should* be done.

### *7.1.3 Sovereignty vs. Envelopment*

Chapter 5 develops the argument that there is a tension between the force of sovereignty and the force of envelopment through the analytical prism of conflicting state roles in neural production networks. On the one hand, the chapter shows that the EU's political leverage as a regulator is constrained by its ambition to act as a facilitator. On the other hand, it points out that the EU's shortcomings in acting as a producer serve to amplify its dependency as a buyer. Although the chapter uses this structure to convey its key arguments, a note of caution is required that those state roles should not be interpreted as natural laws or *a priori* categories. In policy practice, of course, there is no room for classifying regulatory practices or initiatives into already-labelled roles. As such, this section aims to extrapolate the claims of Chapter 5 beyond those state roles and beyond the case of the EU. Such an expanded focus highlights the tensions between sovereignty and envelopment on a higher level of abstraction and provides an impetus to sketch out avenues for future research ideas that also may resort to different methods.

Democratic discourse, according to the political theorist Chantal Mouffe (2000), by definition, requires antagonisms: A "we" needs a "them" to justify its existence. For Mouffe (2000: 69), that 'means establishing a frontier, defining an "enemy."' In crafting this argument, Mouffe builds on Derrida's notion of *différance* – the idea that meaning-making is constituted by oppositions. Arguably, similar discursive dynamics are at play when different actors leverage the idea of digital sovereignty as an ideological umbrella to achieve their interests.[65] For example, in her *Financial Times* commentary, the Former US Trade Representative notes (Barshefsky 2020):

> Demonising US technology companies hinders efforts to address the foremost challenge for both sides with respect to the digital economy: China. Chinese protectionism – which fuses state and Communist party control, and creates subsidies and intellectual property theft on an unparalleled scale – poses an existential threat to a vibrant digital economy. For example, China is pressing for a new centrally controlled internet, which the US and EU oppose.

The metaphor of digital sovereignty, in this sense, presents itself as a proxy for geopolitical interests that go beyond the question of who owns computational resources that are needed for contemporary forms of artificial intelligence. It becomes evident that neural production networks intersect with those geopolitical considerations as much as they intersect with national and supranational policy agendas. Frictions between public and commercial values coincide with a reconfiguration of state roles (e.g., due to the semantic confusion between "producer" and "buyer" roles), but they also affect the relations *between* states. Given the constraints of Chapter 5, it was not feasible to probe those geopolitical relations and economic links in a more granular way, but there is abundant room for future investigations with a focus on how inter-state transactions shape, and are shaped by, neural production networks.

Latin America, for instance, has become a site for power struggles between American and Chinese lead firms, 'who are in a war of position to dominate the cloud and AI market, attempting to become key partners that help private and government actors alike to spur digitalization' (Vila Seoane 2021: 104). Large-scale investments in physical infrastructures, such as data centres, are now key features of the playbook of

[65] In other words, without the industry dominance of American and Chinese lead firms, there would not be a need to regain any form of digital sovereignty in the first place.

proprietors of envelopment. Putting such transnational expansion strategies of lead firms into conversation with the expectations, practices, and expertise of those who are affected by them would afford the chance to give more prominence to questions that prioritise lived experience vis-à-vis conceptual abstraction. This methodological entry point may foreground different forces, relations, and problems than the methodological entry point of this thesis. This, however, would not take anything away from the utility of carving out a new conceptual space for neural production networks.

In this regard, the deployment of imaginative understandings of space to grasp the uneven geographies of neural production networks is not only conceivable but also necessary. Infrastructural dependencies are not limited to the provision of computational resources. Peck (2017: 55), for example, proposes that the global outsourcing complex (i.e., business process outsourcing/offshoring) also has a 'bridging or "infrastructural" role, as a facilitator of methods and mechanisms of cross-border restructuring that were otherwise unrealizable.' Peck relies on a relational understanding of space[66] that enables him to trace how the dominance of some actors entrenches the dependence of others – a way of making sense of space that has not lost any of its relevance in an age of digital platform companies and planetary-scale computational infrastructure.

For example, Facebook only appears as a seamless flow of curated information because outsourced and poorly paid content moderators who toil under psychologically extremely demanding conditions have to check and remove harmful content. The same

[66] Specifically, Peck (2017) refers to Massey's (1991: 28) conceptualisation of a global sense of place. From this perspective, space can be imagined as the 'articulated moments in networks of social relations and understandings, but where a large proportion of those relations, experiences and understandings are constructed on a far larger scale than what we happen to define for the moment as the place itself, whether that be a street, or a region or even a continent.' A relational understanding of space and a processual understanding of space are not mutually exclusive; those perspectives complement each other.

applies to TikTok, Instagram, and Twitter. Although the focus of this thesis is not on such issues of labour exploitation, I do not wish to invoke the impression that collective labour is *not* a force of neural production. Far from it. Annotating large-scale training datasets, maintaining and fixing hardware, running data center facilities, and writing code – all those things require the accumulated work of people in different parts of the world. Far from making outsourced labour obsolete, neural production networks give rise to new tasks, new divisions of labour – and, importantly, new ways of making it look like this labour does not exist or that it will disappear soon.[67]

But while considerations of the territorial expansion of the space within which production takes place are highly relevant and need to be studied alongside relational inequalities and inequities, this thesis argues that it is important to move *beyond* them. Neural geographies of production operate according to a different logic than previously conceptualised production arrangements. Their rise coincides with distinct patterns of industry concentration. Their regulation poses completely new challenges for policy-makers. Their ramifications affect cultural life in contingent and unpredictable ways, facilitating viral internet phenomena and the repurposing of media objects such as game engines. But in addition to my objects of analysis, neural geographies of production also create *entirely different spaces*. Those spaces will have a profound impact on how the tensions between envelopment and sovereignty play out for decades to come.

One such space is the 'deep latent space' (Offert 2021) of generative models like DALL-E and GPT-3. Latent space can be imagined as a 'representation of compressed

[67] A useful way to integrate neural production networks and digital labour scholarship could be to compare and contrast how modes of labour agency operate within those networks. For example, are Google's outsourced data annotation workers more likely to improve their working conditions than Amazon's independent contractors? Focusing on the role of the lead firm is crucial because it has the power to organise things differently.

data in which similar data points are closer together in space' (Tiu 2020). For example, a text-to-image model is able to generate visualisations of text queries such as "Mona Lisa painted by van Gogh" (Chapter 6) because it first decomposes this natural language input into its constitutive elements: "Mona Lisa" and "painted by van Gogh." It then relates those elements to structural similarities that it identified in the data it was trained on (which includes both the Mona Lisa as well as van Gogh's paintings). As Tiu (2020) explains, 'these similarities, usually imperceptible or obscured in higher-dimensional space, can be discovered once our data has been represented in the latent space.' After having related the constitutive elements of text queries to representations of compressed training data, the model reconstructs the latent space representation into a 2D image that represents a morph between the Mona Lisa and van Gogh's paintings – not in the sense of an artistic morph, but in the sense of how the model *interpreted* the relation between data points within latent space. Every time this process takes place, a different outcome is generated. Paradoxically, every synthetic Mona Lisa is unique.[68]

There is a direct line between neural production networks and the emergence of those distinctly neural spaces, representations, and geographies. I advocate for openness towards these spaces and their creative exploration. The metaphor of envelopment helps scrutinise the uneven power relations that bring latent space into being. The metaphor of sovereignty helps us understand the efficacy of regulatory frameworks that are supposed to govern and co-constitute neural spaces of production. But the fundamental question is

[68] By stressing this, I do not wish to contradict the conclusion of Chapter 6 that there might be a homogenization of platformised cultural production. The computational circumstance that there is, in theory, an endless number of possibilities to create relations and links between data points in ever-more sophisticated forms of latent space is what makes models such as DALL-E and GPT-3 powerful in the first place. A homogenization of cultural production can co-exist with a heterogenization of synthetic media production.

whether the scope of the force of sovereignty will find its way into latent space as well. Technological progress in the domain of deep generative models is occurring so rapidly that regulators cannot keep up. By design, regulation is a *reaction* to things, events, and patterns that happened in the past. However, as I stress in Chapter 5, the spatiotemporal reconfiguration of new legislative instruments like the Digital Markets Act into *ex-ante* regulations (e.g., setting up a legal category for gatekeepers) reflects a potential way to set the conditions for future developments. Even if the consolidation of market power in neural production networks in the hands of a few American and Chinese lead firms can neither be fully forestalled nor reversed, its expansion can be anticipated.

In other words, the force of envelopment will take different forms and shapes in the years to come, and it is important to recognise those practices as what they are: ways to create and sustain infrastructural dependencies. For example, Chapter 4 uses the case of AutoML techniques to render intelligible the force of envelopment. But in the future, lead firms will continue to optimise and simplify those techniques so that they reach a much broader user base than they do now. Analytically speaking, the most important task for future research is to maintain an 'awareness of essential things [that] so quickly fades into "beaten paths of impercipience"' (Peters 2015: 31). Google's search engine, for example, has become such a taken-for-granted infrastructure of everyday life that I initially did not consider it to be an empirical case of computational envelopment when crafting the arguments and the case study design of Chapter 4. But, not unlike DALL-E mini or DeepMind's AlphaGo, Google's search engine is also underpinned by the firm's proprietary tensor processing units (TPUs). Adapting the arguments of this thesis to the mundane realities of infrastructural life is a fruitful endeavour – one that will continue to require an imaginative understanding of how the forces of neural production relate to each other and to the spaces that they enact and transduce.

### 7.3 Epilogue: Carnegie's Legacy

Only three years after Carnegie published *The Gospel of Wealth*, his main steel plant in Homestead, Pennsylvania, became the scene of one of the bloodiest labour struggles in U.S. history. Although Carnegie publicly condemned practices of strike-breaking, the Carnegie Steel Company hired 300 Pinkerton agents to violently put down the strike, with riots resulting in the deaths of seven workers and three security guards. It is not an exaggeration to say that there were stark discrepancies between Carnegie's essay and the ways in which his steel company treated its workers, whose labour made possible the surplus capital for his philanthropic activities in the first place.

When thinking about an appropriate starting point for my thesis, I initially chose a historical analogy to *The Gospel of Wealth* for two reasons. First, to foreground the question: Why is it that we tend to think of individuals or innovative technologies as guarantors of social progress, instead of collective human labour? One potential answer to this question, I considered, is to be found in the fact that powerful economic actors put a lot of discursive and ideological work into eradicating the central role of collective work, foregrounding *their* skills. Second, I aimed to problematise the systemic opacity of capitalist activities, which makes it difficult to grasp production arrangements – a phenomenon that did not start with Carnegie and that will not end with the networks of outsourced human labour powering today's AI systems.

But in 2022, three years after those initial considerations, the *Gospel of Wealth* appears in a different light to me: it is a prototypical example of how corporate actors insinuate their economic force into the fabric of political and cultural life – the fabric of infrastructural life. While Carnegie's exploitation of workers may be locatable in time and space (at least if we interpret exploitation in such terms), his longstanding impacts on infrastructural life are *not* locatable in distinct time and space. Herein lies Carnegie's legacy as a historical analogy for the purposes of this thesis: his power in influencing how wealth, societal progress, and the distribution of resources are discursively framed, perceived, and acted upon, decades and centuries after his death.

In using this historical analogy, I do not wish to imply that economic, political, and cultural developments can be explained by singling out particular individuals such as Carnegie. Instead, I deploy it to point out the importance of the historical context in which the creation and circulation of new concepts and ideas take place. Carnegie's distinct modes of industrial organisation, and those of other business tycoons around the world, prompted a generation of critical thinkers to conceive and craft new abstractions and metaphors to encapsulate the deeply uneven power relations that characterise them. But abstractions emanating from this period, such as the Marxian concept of productive forces, also carry the biases of their creators. As Graeber (2018: 232) argued, 'in fact, there was never a time most workers worked in factories.' Although I explicitly do not focus on questions of work and labour in this thesis, I emphasise this point because it illustrates that theory-building does not occur in a historical vacuum. New concepts cannot (and should not) pretend that they are disconnected from the events, processes, and developments alongside which they emerge.

In this thesis, I open out a new conceptual space at the nexus of platform studies and economic geography. The idiosyncrasies of neural production networks justify a new understanding of geographies of production – not in the static sense of what they *are* but rather in the ontogenetic sense of what they *become*. To grasp this continually-unfolding process of becoming, it is useful to elucidate the frictions and contradictions between the forces of neural production. For example, the force of sovereignty *requires* the force of envelopment. I aimed to demonstrate that the synthesis of those conceptual links does not result in a totalising account of neural production networks. At the same time, I wish to acknowledge that the arguments of the thesis carry a range of biases. First, as I resort to the analysis of paradigmatic case studies, my approach to research exhibits a tendency to prioritise scope over depth (see Chapter 3). My empirical material was selected and analysed based on the grounds that it can 'play a supportive role, and it facilitates our understanding of something else' (Stake 2005: 445). Second, my claims are constrained by the deficiencies of document analysis that the primary data gathering technique. Third, due to non-disclosure agreements and trade secrets, specific pieces of information will continue to remain concealed by corporate walled gardens.

But those biases have an upside. By definition, a thesis is a complex whole that is supposed to be greater than the sum of its chapters. The most important contribution of this thesis is its extrapolation of otherwise-unelaborated similarities, patterns, and processes associated with the infrastructural geography of artificial intelligence into a new cross-disciplinary framework. A framework that poses new research questions. A framework that explores and combines economic, political, and cultural matters. And a framework that is open for reconfiguration and reappropriation. But while a thesis ends at some point, curiosity does not end. Curiosity is impulsive, affective, and diffuses into seemingly disconnected directions. Ideas about how those directions relate to each other often present themselves in hindsight, not prior to their movement.

Neural production networks are already diffusing into all manners of seemingly disconnected directions. As they come to infiltrate ever-more economic, political, and cultural spheres, locating them in time and space will be challenging. Like Carnegie's ideological infusions of public consciousness, they will shape things and processes that we can neither anticipate nor prevent. But just because their diffusion is unpredictable, this does not mean that it cannot be theorised. Studying their movement will require an understanding of the forces that transform them into something else.

**Bibliography**

Alford, M., & Phillips, N. (2018). The political economy of state governance in global production networks: Change, crisis and contestation in the South African fruit sector. *Review of International Political Economy*, *25*(1), 98–121.

Allen, G. W. (2001). *None So Blind: A Personal Failure Account of the Intelligence in Vietnam*. Ivan R. Dee Publishing.

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., ... & Asari, V. K. (2018). The history began from alexnet: A comprehensive survey on deep learning approaches. arXiv preprint https://arxiv.org/abs/1803.01164

AI for Good (2022). About. https://aiforgood.itu.int/about-ai-for-good/

Amoore, L. (2018). Cloud geographies: computing, data, Sovereignty. *Progress in Human Geography*, 42(1), 4-24.

———. (2021). The deep border. *Political Geography*, 102547.

Ash, J. (2015). *The interface envelope: Gaming, technology, power*. Bloomsbury.

Bair, J. & Werner, M. (2011). Commodity chains and the uneven geographies of global capitalism: a disarticulations perspective. *Environment and Planning A*, 43(5): 988–997.

Bair, J. (2005). Global capitalism and commodity chains: looking back, going forward. *Competition & Change*, 9(2), 153-180.

Baudrillard, J. (1994). *Simulacra and simulation*. University of Michigan Press.

Bennett, J. (2010). *Vibrant matter: A political ecology of things*. Duke University Press.

Bernevega, A., & Gekker, A. (2022). The Industry of Landlords: Exploring the Assetization of the Triple-A Game. *Games and Culture*, *17*(1), 47–69.

Berry, D. M., & Dieter, M. (2015). *Postdigital Aesthetics*. Palgrave Macmillan.

Birch, K. (2015). *We Have Bever been Neoliberal: A Manifesto for a Doomed Youth*. Zero Books.

———. (2020). Technoscience rent: Toward a theory of rentiership for technoscientific capitalism. *Science, Technology, & Human Values*, *45*(1), 3-33.

Birch, K., & Muniesa, F. (2020). *Assetization: Turning Things into Assets in Technoscientific Capitalism*. MIT Press.

Birch, K., & Cochrane, D. T. (2022). Big Tech: Four emerging forms of digital rentiership. *Science as Culture*, *31*(1), 44-58.

Blanke, T., & Pybus, J. (2020). The Material Conditions of Platforms: Monopolization Through Decentralization. *Social Media+ Society*, 6(4), 2056305120971632.

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2009). Toward information infrastructure studies: Ways of knowing in a networked environment. In J. Hunsinger, L. Klastrup, and M. Allen (eds.), *International Handbook of Internet Research*. Springer.

Bowen, G.A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27-40.

Bratton, B. H. (2016). *The Stack: On Software and Sovereignty*. MIT Press.

Bridge, G. (2008). Global production networks and the extractive sector: governing resource-based development. *Journal of Economic Geography*, 8(3), 389-419.

Broad, T. (2019). Topological Visualisation of a Convolutional Neural Network. terencebroad.com/nvnvis.html.

Brodie, P. (2020). Climate Extraction and Supply Chains of Data*. Media, Culture & Society*, 42(7-8), 1095-1114.

Broussard, M. (2018). *Artificial Unintelligence: How Computers Misunderstand the World*. MIT Press.

Bryman, A. (2012). *Social Research Methods.* Oxford University Press.
Bucher, T. (2018). *If... Then: Algorithmic Power and Politics*. Oxford University Press.
Canales, J. (2020). *Bedeviled: A Shadow History of Demons in Science*. Princeton University Press.
Cambridge Dictionary (2022a). Force. https://dictionary.cambridge.org/dictionary/english/force
———. (2022b). Infrastructure. https://dictionary.cambridge.org/dictionary/english/infrastructure
Castree, N. (2001). Commodity fetishism, geographical imaginations and imaginative geographies. *Environment and Planning A*, *33*(9), 1519-1525.
Carnegie, A. (1889). Wealth. *North American Review*, 148(391), 653-665.
Christophers, B. (2020). *Rentier Capitalism: Who Owns the Economy, and Who Pays for It*. London: Verso.
Chun, W. (2008). Programmability. In M. Fuller (Ed.), *Software Studies: A Lexicon.* MIT Press.
Cihon, P., Maas, M. M., & Kemp, L. (2020). Fragmentation and the Future: Investigating Architectures for International AI Governance. *Global Policy*, *11*(5), 545–556.
Clegg, S. (2015). Power. *Oxford Bibliographies (Sociology)*, 29.06.2015.
Coe, N. M., Hess, M., Yeung, H. W. C., Dicken, P., & Henderson, J. (2004). 'Globalizing' regional development: a global production networks perspective. *Transactions of the Institute of British geographers*, 29(4), 468-484.
Coe, N. M., Dicken, P., & Hess, M. (2008). Global production networks: Realizing the potential. *Journal of Economic Geography*, 8, 271–295.
Coe, N. M., & Yeung, H. W. C. (2015). *Global Production Networks: Theorizing Economic Development in an Interconnected World*. Oxford University Press.
Coe, N. (2016). The geographies of global production networks. In P. Daniels, M. Bradshaw, D. Shaw, J. Sidaway, and T, Hall (eds.), An *Introduction to Human Geography: Issues for the 21st Century*. Pearson.
Coe, N. M., & Yeung, H. W. (2019). Global production networks: Mapping recent conceptual developments. *Journal of Economic Geography*, *19*(4), 775–801.
Coe, N. M., & Yang, C. (2022). Mobile Gaming Production Networks, Platform Business Groups, and the Market Power of China's Tencent. *Annals of the American Association of Geographers*, *112*(2), 307–330.
Crawford, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
Coleman, M. (2009). Sovereignty. In: R. Kitchin & N. Thrift (Eds.), *International Encyclopedia of Human Geography*. Elsevier.
Collins Dictionary. (2022). Hyperproduction. https://www.collinsdictionary.com/dictionary/english/hyperproduction
Conger, K. & Cameron, D. (2018). Google is Helping the Pentagon Build AI for Drones. *Gizmodo,* 06.03.2018.https://gizmodo.com/google-is-helping-the-pentagon-build-ai-for-drones-1823464533.
Cox, J. W. & Hassard, J. (2010). Triangulation. In A. J. Mills, G. Eurepos, and E. Wiebe (eds.), *Encyclopedia of Case Study Research*. SAGE Publications.
Cuenca, P. (2022). The Infrastructure Behind Serving DALL·E Mini. *Weights & Biases*, 16.06.2022. https://wandb.ai/dalle-mini/dalle-mini/reports/The-Infrastructure-Behind-Serving-DALL-E-Mini--VmlldzoyMTI4ODAy
Denzin, N. K. (2012). Triangulation 2.0. *Journal of mixed methods research*, *6*(2), 80-88.

Dicken, P. (2004). Geographers and 'Globalization': (Yet) Another Missed Boat?. Transactions of the Institute of British Geographers, 29, 5–26.

Dicken, P., & Thrift, N. (1992). The organization of production and the production of organization: why business enterprises matter in the study of geographical industrialization. *Transactions of the Institute of British geographers*, 279-291.

Durrenberger, E., & Doukas, D. (2008). Gospel of wealth, gospel of work: Counterhegemony in the US working class. *American Anthropologist*, 110(2), 214-224.

Dyer-Witheford, N., & Peuter, G. (2009). *Games of Empire: Global Capitalism and Video Games*. Minneapolis: University of Minnesota Press.

Dyer-Witheford, N., Kjøsen, A. M., & Steinhoff, J. (2019). *Inhuman Power. Artificial Intelligence and the Future of Capitalism*. London: Pluto Press.

Ekbia, H. R., & Nardi, B. A. (2017). *Heteromation, and Other Stories of Computing and Capitalism*. MIT Press.

Edwards, P. N. (2003). Infrastructure and Modernity: Force, Time, and Social Organization in the History of. In T. Misa, P. Brey, and A. Feenberg (eds.), *Modernity and Technology*. MIT Press.

Fang, L. (2019). Google hired gig economy workers to improve artificial intelligence in controversial drone-targeting project. *The Intercept*, 04.02.2019. https://theintercept.com/2019/02/04/google-ai-project-maven-figure-eight/.

Fernandez, R., Adriaans, I., Klinge, T. J., & Hendrikse, R. (2020). The Financialisation of Big Tech. *SOMO (Stichting Onderzoek Multinationale Ondernemingen*. https://www.somo.nl/the-financialisation-of-big-tech/

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative inquiry*, *12*(2), 219-245.

———. (2011). Case Study. In N. K. Denzin and Y. S. Lincoln, (eds.), *The Sage Handbook of Qualitative Research.* SAGE Publications.

Ford, M. (2018). *Architects of Intelligence: The truth about AI from the people building it*. Packt.

Foster, C., & Graham, M. (2016). Reconsidering the Role of the Digital in Global Production Networks. *Global Networks*, 17(1), 68-88.

Foroohar, R. (2021). Big Tech is playing a financial shell game. *Financial Times*, 12.12.2021. https://www.ft.com/content/99ca12c5-498e-4ee4-8046-a27cd0a1038b

Freedman, E. (2018). Software. In M. Kackman & M. C. Kearney (Eds.), *The Craft of Criticism: Critical Media Studies in Practice* (pp. 318–330). Routledge.

Friedman, T. (2005). *The World Is Flat: A Brief History of the Twenty-first Century*. Farrar, Straus and Giroux.

Gaboury, J. (2021). *Image Objects: An Archaeology of Computer Graphics*. MIT Press.

Galloway, S. (2018). *The four: the hidden DNA of Amazon, Apple, Facebook, and Google*. Penguin.

Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research policy*, 43(7), 1239-1249.#

Gereffi, G. (1999). A commodity chains framework for analyzing global industries. *Institute of Development Studies*, *8*(12), 1-9.

———. (2018). *Global value chains and development: Redefining the contours of 21st century capitalism*. Cambridge University Press.

Gereffi, G., Humphrey, J., & Sturgeon, T. (2005). The governance of global value chains. *Review of international political economy*, 12(1), 78-104.

Gershgorn, D. (2017). The data that transformed AI research – and possibly the world. *Quartz*, 26.07.2017. https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/.
Gillespie, T. (2010). The politics of platforms. *New Media & Society*, 12(3), 347-364.
———. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
Glaser, B. G., & Strauss, A. L. (2006). *The discovery of grounded theory: Strategies for qualitative research*. Routledge.
Goldstein, B., & Newell, J. P. (2020). How to track corporations across space and time. *Ecological Economics*, *169*, 106492.
Gorwa, R. (2019). What is platform governance? *Information, Communication & Society*, *22*(6), 854–871.
Grabher, G. (2020). Enclosure 4.0: Seizing Data, Selling Predictions, Scaling Platforms. *Sociologica*, 14(3), 241-265.
Grabher, G., & van Tuijl, E. (2020). Uber-production: From global networks to digital platforms. *Environment and Planning A: Economy and Space*, 52(5), 1005-1016.
Graeber, D. (2018). *Bullshit Jobs: A Theory.* Simon & Schuster.
Graham, M. (2020). Regulate, replicate, and resist–the conjunctural geographies of platform urbanism. *Urban Geography*, 41(3), 453-457.
Graham, M., & Dittus, M. (2022). *Geographies of Digital Exclusion: Data and Inequality*. Pluto Press.
Gray, J., Bounegru, L., & Venturini, T. (2020). 'Fake news' as infrastructural uncanny. *New media & society*, *22*(2), 317-341.
Gray, M. L., & Suri, S. (2019*). Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Eamon Dolan Books.
Hao, K. (2020). We read the paper that forced Timnit Gebru out of Google. Here's what it says. *MIT Technology Review*, 04.12.2020. https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/
Harvey, D. (2006). *Spaces of global capitalism*. Verso.
Havice, E., & Pickles, J. (2019). Articulating value and missing links in 'Geographies of Dissociation'. *Dialogues in Human Geography*, 9(1), 73-77.
Hayles, N. K. (1999). *How We Became Posthuman: Virtual Bodies In Cybernetics, Literature, and Informatics*. University of Chicago Press.
Heaven, W. G. (2021). AI fake-face generators can be rewound to reveal the real faces they trained on. *MIT Technology Review*, 12.10.2021. https://www.technologyreview.com/2021/10/12/1036844/ai-gan-fake-faces-data-privacy-security-leak/
Hegarty, P. (2004). *Jean Baudrillard: Live Theory*. Continuum.
Helmond, A. (2015). The platformization of the web: Making web data platform ready. *Social Media+ Society*, 1(2), 2056305115603080.
Hepworth, K. (2014). Enacting logistical geographies. *Environment and Planning D: Society and Space*, *32*(6), 1120-1134.
Hess, M. (2018). Global Production Networks. In *The International Encyclopedia of Geography: People, the Earth, Environment, and Technology*, edited by Douglas Richardson. Malden and Oxford: John Wiley & Sons.
Hogan, R. (2006). *Personality and the Fate of Organizations*. Routledge.
Horner, R. (2017). Beyond facilitator? State roles in global value chains and global production networks. *Geography Compass*, *11*(2), e12307.

Horner, R., & Alford, M. (2019). The roles of the state in global value chains. *Handbook on Global Value Chains*, 555–569.

Howson, K., Ferrari, F., Ustek-Spilda, F., Salem, N., Johnston, H., Katta, S., Heeks, R., & Graham, M. (2021). Driving the digital value network: Economic geographies of global platform capitalism. *Global Networks*, OnlineFirst.

Hudson, R. (2008). Cultural political economy meets global production networks: a productive meeting?. *Journal of economic geography*, *8*(3), 421-440.

Hughes, A., Morrison, E., & Ruwanpura, K. N. (2019). Public sector procurement and ethical trade: Governance and social responsibility in some hidden global supply chains. *Transactions of the Institute of British Geographers*, *44*(2), 242-255.

Hull, M. S. (2012). Documents and bureaucracy. *Annual Review of Anthropology*, *41*, 251-267.

Infrastructural Geographies (2022). https://www.geog.cam.ac.uk/research/infrastructure/

Irani, L. C., & Silberman, M. S. (2013). Turkopticon: Interrupting worker invisibility in Amazon Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 611-620).

Johnson, D. (2017). Production. In: L. Ouellette and J. Gray (eds.), *Keywords for Media Studies*. New York University Press.

Kenney, M., & Zysman, J. (2020). The platform economy: restructuring the space of capitalist accumulation. Cambridge journal of regions, economy and society, 13(1), 55-76.

Kenney, M., Rouvinen, P., Seppälä, T., & Zysman, J. (2019). Platforms and industrial change. *Industry and Innovation*, 26(8), 871-879.

Khan, L. M. (2016). Amazon's antitrust paradox. *Yale lJ*, 126, 710.

Kitchin, R. (2009). Space II. In: In: R. Kitchin & N. Thrift (Eds.), *International Encylopedia of Human Geography*. Elsevier.

Klinge, T. J., Hendrikse, R., Fernandez, R., & Adriaans, I. (2022). Augmenting digital monopolies: A corporate financialization perspective on the rise of Big Tech. *Competition & Change*.

Klose, A. (2015). *The Container Principle: How a Box Changes the Way We Think*. MIT Press.

Knight, W. (2022). DALL-E Mini Is the Internet's Favorite AI Meme Machine*. Wired Magazine*, 27.06.2022. https://www.wired.com/story/dalle-ai-meme-machine/

Krafft, P. M., Young, M., Katell, M., Huang, K., & Bugingo, G. (2019). Policy versus Practice: Conceptions of Artificial Intelligence. Available at SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3431304.

Kwa, C. (2002). Romantic and baroque conceptions of complex wholes in the sciences. In B. Smith, J. Law, R, Weintraub and A. Mol (eds.). *Complexities: Social Studies of Knowledge Practices.* Duke University Press.

Lai, K. P., & Samers, M. (2020). Towards an economic geography of FinTech. *Progress in Human Geography*, 0309132520938461.

Langley, P., & Leyshon, A. (2017). Platform capitalism: the intermediation and capitalization of digital economic circulation. *Finance and society.*, *3*(1), 11-31.

Langlois, G., & Elmer, G. (2019). Impersonal subjectivation from platforms to infrastructures. *Media, Culture & Society*, 41(2), 236-251.

Larkin, B. (2013). The politics and poetics of infrastructure. *Annual Review of Anthropology*, 42, 327-343.

Law, J. (2004). *After method: Mess in social science research*. Routledge.

Law Commission of Ontario (2021). Comparing European and Canadian AI Regulation. https://www.lco-cdo.org/en/our-current-projects/ai-adm-and-the-justice-system/comparing-european-and-canadian-ai-regulation/
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521(7553), 436.
Lim, K. F. (2018). Strategic coupling, state capitalism, and the shifting dynamics of global production networks. *Geography Compass*, 12(11), e12406.
Lipsey, R. G., Carlaw, K. I., & Bekar, C. T. (2005). *Economic transformations: general purpose technologies and long-term economic growth*. Oxford University Press.
Luitse, D., & Denkena, W. (2021). The Great Transformer: Examining the Role of Large language Models in the Political Economy of AI. *Big Data & Society*, *8*(2).
MacQuarrie, C. (2010). Holistic Designs. In A. J. Mills, G. Eurepos, and E. Wiebe (eds.), *Encyclopedia of Case Study Research*. SAGE Publications.
Manovich, L. (2002). *The language of new media*. MIT Press.
Massey, D. (1991). A Global Sense of Place. *Marxism Today* (38): 24-29.
———. (1995). *Spatial Divisions of Labour: Social Structures and the Geography of Production*. London: Macmillan International Higher Education.
———. (2005). *For space*. London: Sage.
Maskell, P. (2001). The firm in economic geography. *Economic Geography*, *77*(4), 329-344.
Mayer, F. W., & Phillips, N. (2017). Outsourcing governance: States and the politics of a 'global value chain world'. *New Political Economy*, *22*(2), 134–152.
Mazzucato, M. (2018). *The value of everything: Making and taking in the global economy*. Hachette UK.
McCormack, D. (2012). Geography and abstraction: Towards an affirmative critique. *Progress in Human Geography*, *36*(6), 715-734.
McGrath, S. (2018). Dis/articulations and the interrogation of development in GPN research. *Progress in Human Geography*, 42(4), 509-528.
McKelvey, F. (2011). A Programmable Platform? Drupal, Modularity, and the Future of the Web. *Fibreculture*, *18*, 232–254.
———. (2018). *Internet Daemons: Digital Communications Possessed*. Minneapolis. University of Minnesota Press.
Merriam-Webster (2022a). Force. https://www.merriam-webster.com/dictionary/force
———. (2022b). Infrastructure. https://www.merriam-webster.com/dictionary/infrastructure
———. (2022c). Ecosystem. https://www.merriam-webster.com/dictionary/ecosystem
Mills, C. W. (2000). *The sociological imagination*. Oxford University Press.
Mol, A. (2002). *The Body Multiple: Ontology in Medical Practice*. Duke University Press.
Moles, P. & Terry, M. (1997a). Market Capitalisation. In: P. Moles and N. Terry, *The Handbook of International Financial Terms*. Oxford University Press.
Moles, P. & Terry, M. (1997b). Asset Specificity. In: P. Moles and N. Terry, *The Handbook of International Financial Terms*. Oxford University Press.
Mosco, V. (2018). Political Economies of Media Technologies. *Oxford Research Encyclopedia of Communication.*
Mouffe, C. (2020). *The Democratic Paradox*. Verso.
Narayan, D. (2022). Platform capitalism and cloud infrastructure: Theorizing a hyper-scalable computing regime. *Environment and Planning A: Economy and Space*.
Natale, S. (2021). *Deceitful Media: Artificial Intelligence and Social Life After the Turing Test*. Oxford University Press.

Neilson, B., & Notley, T. (2019). Data Centres as Logistical Facilities: Singapore and the Emergence of Production Topologies. *Work Organisation, Labour & Globalisation*, 13(1), 15-29.
Nicoll, B., & Keogh, B. (2019). *The Unity Game Engine and the Circuits of Cultural Software*. Springer International Publishing.
Nieborg, D. B., & Poell, T. (2018). The platformization of cultural production: Theorizing the contingent cultural commodity. *New media & society*, *20*(11), 4275-4292.
Nieborg, D. B., & Helmond, A. (2019). The political economy of Facebook's platformization in the mobile ecosystem: Facebook Messenger as a platform instance. *Media, Culture & Society*, 41(2), 196-218.
OECD (2001). Economic Production. https://stats.oecd.org/glossary/detail.asp?ID=725
———. (2022). OECD AI Principles Overview. https://oecd.ai/en/ai-principles
Offert, F. (2021). Latent Deep Space: Generative Adversarial Networks (GANs) in the Sciences. *Media+Environment*. https://doi.org/10.1525/001c.29905.
Olson, M. (2010). Document Analysis. In A. J. Mills, G. Eurepos, and E. Wiebe (eds.), *Encyclopedia of Case Study Research*. SAGE Publications.
Parks, L., & Starosielski, N., (2015). *Signal traffic: Critical studies of media infrastructures.* Champaign: University of Illinois Press.
Parks, L. (2017). Infrastructure. In L. Ouellette and J. Gray (eds.), *Keywords for Media Studies*. New York University Press.
Pavlich, G. (2010). Paradigmatic Cases. In A. J. Mills, G. Eurepos, and E. Wiebe (eds.), *Encyclopedia of Case Study Research*. SAGE Publications.
Peck, J. (2017). *Offshore: Exploring the Worlds of Global Outsourcing*. Oxford: Oxford University Press.
Peck, J., & Phillips, R. (2020). The Platform Conjuncture. *Sociologica*, 14(3), 73-99.
Peck, J., & Tickell, A. (2002). Neoliberalizing Space. *Antipode*, 34(3), 380-404.
Perrotta, C., Gulson, K. N., Williamson, B., & Witzenberger, K. (2021). Automation, APIs and the distributed labour of platform pedagogies in Google Classroom. *Critical Studies in Education*, 62(1), 97-113.
Peters, J. D. (2015). *The Marvelous Clouds*. University of Chicago Press.
Pickles, J., & Smith, A. (2016). *Articulations of capital: global production networks and regional transformations.* John Wiley & Sons.
Plantin, J. C., & Punathambekar, A. (2019). Digital media infrastructures: pipes, platforms, and politics. *Media, Culture & Society*, 41(2), 163-174.
Plantin, J. C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, 20(1), 293-310.
Ponte, S., & Sturgeon, T. (2014). Explaining governance in global value chains: A modular theory-building effort. *Review of international political economy*, *21*(1), 195-223.
Rahman, K.S., & Thelen K. (2019). The Rise of the Platform Business Model and the Transformation of Twenty-First-Century Capitalism. *Politics and Society*, 47(2), 177–204.
Rochet, J. C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the european economic association*, *1*(4), 990-1029.
Rogers, R. (2013). *Digital Methods*. MIT Press.
Rossiter, N. (2014). Logistical Worlds. *Cultural Studies Review*, *20*(1), 53–76.
Russell, A. L. (2012). Modularity: An Interdisciplinary History of an Ordering Concept. *Information & Culture: A Journal of History*, *47*(3), 257–287.

Sadowski, J. (2019). When Data is Capital: Datafication, Accumulation, and Extraction'. *Big Data & Society*, *6*(1),
———. (2020). The Internet of Landlords: Digital Platforms and New Mechanisms of Rentier Capitalism. *Antipode*, *52*(2), 562–580.
Saba, M. (1980. Productive Forces. *Struggle!*, 11.11.1980. https://www.marxists.org/history/erol/ca.secondwave/is-productive-forces.htm
Saldanha, A. (2003). Review Essay: Actor-Network Theory and Critical Sociology. *Critical Sociology*, *29*(3), 419-432.
Salminen, J., & Rajavuori, M. (2019). Transnational sustainability laws and the regulation of global value chains: comparison and a framework for analysis. Maastricht Journal of European and Comparative Law, 26(5), 602-627.
Schmidt, F. A. (2019). Crowdsourced Production of AI Training Data. *Hans Boeckler Stiftung Working Paper.* https://www.boeckler.de/pdf/p_fofoe_WP_155_2019.pdf.
Seyfert, R. (2021). Algorithms as Regulatory Objects. *Information, Communication & Society*, 1–17.
Smith, A. (2015). The state, institutional frameworks and the dynamics of capital in global production networks. *Progress in Human Geography*, *39*(3), 290–315.
Smuha, N. A. (2020). From a 'Race to AI' to a 'Race to AI Regulation'*: Regulatory Competition for Artificial Intelligence*. 27.
Soskis, B. (2014). Both More and No More: The Historical Split between Charity and Philanthropy. *Bradley Center*. https://www.hudson.org/research/10723-both-more-and-no-more-the-historical-split-between-charity-and-philanthropy.
Srnicek, N. (2016). *Platform capitalism.* Polity.
———. (2020a). Data, compute, labour. *Ada Lovelace Institute*, 30.06.2020. https://www.adalovelaceinstitute.org/blog/data-compute-labour/
———. (2020b). Paths forward for the study of the digital economy. In J. Muldoon and W. Stronge (eds.), *Platforming Equality: Policy Challenges for the Digital Economy*. Autonomy.
———. (2021). Value, Rent, and Platform Capitalism. In J. Haidar and M. Keune (eds.), *Work and Labour Relations in Global Platform Capitalism*. Elgar.
Stake, R. (2005). Qualitative case studies. In N. K. Denzin and Y. S. Lincoln (eds.), *The Sage handbook of qualitative research*. SAGE Publications.
Star, S. L. (1999). The ethnography of infrastructure. *American behavioral scientis*t, 43(3), 377-391.
Star, S. L., & Strauss, A. (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. Computer supported cooperative work (CSCW), 8(1-2), 9-30.
Starosielski, N. (2015). *The Undersea Network.* Duke University Press.
Steinhoff, J. (2022). Toward a political economy of synthetic data: A data-intensive capitalism that is not a surveillance capitalism? *New Media & Society*.
Stephenson, S. R., & Agnew, J. A. (2016). The work of networks: Embedding firms, transport, and the state in the Russian Arctic oil and gas sector. *Environment and Planning A: Economy and Space*, 48(3), 558-576.
Swain, J., & King, B. (2022). Using Informal Conversations in Qualitative Research. *International Journal of Qualitative Methods*.
The Economist (2022). How Smarter AI Will Change Creativity. *The Economist*, 09.06.2022. https://www.economist.com/leaders/2022/06/09/artificial-intelligences-new-frontier

Tsing, A. (2009). Supply Chains and the Human Condition. *Rethinking Marxism*, *21*(2), 148–176.
Tubaro, P., & Casilli, A. A. (2019). Micro-work, artificial intelligence and the automotive industry. Journal of Industrial and Business Economics, 46(3), 333-345.
Turner, F. (2006). *From Counterculture to Cyberculture*. University of Chicago Press.
Vanwynsberghe, R., & Khan, S. (2007). Redefining case study. *International journal of qualitative methods*, *6*(2), 80-94.
van der Vlist, F. N., Helmond, A., Burkhardt, M., & Seitz, T. (2022). API Governance: The Case of Facebook's Evolution. *Social Media+ Society*, *8*(2).
Van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press.
———. (2020). Seeing the forest for the trees: Visualizing platformization and its governance. *New Media & Society*, 1461444820940293.
Van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.
Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97–112.
Venturini, T., Bounegru, L., Gray, J., & Rogers, R. (2018). A reality check (list) for digital methods. *New Media & Society*, *20*(11), 4195-4217.
Vila Seoane, M. F. (2021). Chinese and US AI and cloud multinational corporations in Latin America. In T. Keskin and R. D. Kiggins (eds.), *Towards an International Political Economy of Artificial Intelligence*. Palgrave Macmillan, Cham.
Weatherby, L. (2018). Delete Your Account: On the Theory of Platform Capitalism. Los Angeles Review of Books, 24.04.2018. https://lareviewofbooks.org/article/delete-your-ac-count-on-the-theory-of-platform-capitalism/
Walker, M., & Winders, J. (2021). Where is artificial intelligence? Geographies, ethics, and practices of AI. *Space and Polity*, *25*(2), 163-166.
Weller, S. (2006). The embeddedness of global production networks: the impact of crisis in Fiji's garment export sector. *Environment and Planning A*, *38*(7), 1249-1267.
Werner, M. (2019). Geographies of production I: Global production and uneven development. *Progress in Human Geography*, 43(5), 948-958.
———. (2021). Geographies of production II: Thinking through the state. *Progress in Human Geography*, *45*(1), 178–189.
Whittaker, M. (2021). The Steep Cost of Capture. *Interactions*, *28*(6), 50–55.
Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Networked but commodified: The (dis) embeddedness of digital labour in the gig economy. *Sociology*, *53*(5), 931-950.
Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* Public Affairs.

**Appendices**



**Appendix A: Corpus of Document Analysis (Chapter 4)**


Amazon Science (2020). How NASA uses AWS to protect life and infrastructure on earth. Amazon Science Blog, 14.01.2020. https://www.amazon.science/how-nasa-uses-aws-to-protect-life-and-infrastructure-on-earth

Amazon Web Services (2017). Netflix & Amazon Kinesis Data Streams Case Study. https://aws.amazon.com/solutions/case-studies/netflix-kinesis-data-streams/

———. (2021). Perseverance lands on Mars, cloud-ready to explore. AWS Public Sector, 22.02.2021. https://aws.amazon.com/blogs/publicsector/perseverance-lands-mars-cloud-ready-explore/

———. (2022a). AWS Inferentia. https://aws.amazon.com/machine-learning/inferentia

———. (2022b). AWS Global Infrastructure. https://aws.amazon.com/about-aws/global-infrastructure/

Das, P., Ivkin, N., Bansal, T., Rouesnel, L., Gautier, P., Karnin, Z., ... & Venkateswar, K. (2020). Amazon SageMaker Autopilot: a white box AutoML solution at scale. In *Proceedings of the Fourth International Workshop on Data Management for End-to-End Machine Learning.*

Dickson, B (2020). DeepMind's big losses, and the questions around running an AI lab. *VentureBeat*, 27.12.2020. https://venturebeat.com/2020/12/27/deepminds-big-losses-and-the-questions-around-running-an-ai-lab/

Donnelly, C. (2019). AWS 'does not dominate' cloud market, claims Amazon's UK public policy chief. *ComputerWeekly.com*. 09.01.2019. https://www.computerweekly.com/news/252455512/Amazon-does-not-dominate-cloud-and-online-retail-market-UK-public-policy-chief-claims

Fang, L. (2018). Leaked emails show Google expected lucrative military drone AI work to grow exponentially. *The Intercept*, 31.05.2018. https://theintercept.com/2018/05/31/google-leaked-emails-drone-ai-pentagon-lucrative/

Gartner (2021a). Gartner Says Worldwide IaaS Public Cloud Services Market Grew 40.7% in 2020. *Gartner Newsroom*, 28.06.2021. https://www.gartner.com/en/newsroom/press-releases/2021-06-28-gartner-says-worldwide-iaas-public-cloud-services-market-grew-40-7-percent-in-2020

Gershgorn, D. (2018). Google gave the world powerful AI tools, and the world made porn with them. *Quartz,* 07.02.2018. https://qz.com/1199850/google-gave-the-world-powerful-open-source-ai-tools-and-the-world-made-porn-with-them/

Google. (2022). Advancing AI for Everyone. https://ai.google/

He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems*, *212*, 106622.

Heller, M. (2020). Google Waives $1.5B in DeepMind Startup Debt. *CFO*, 17.12.2020. https://www.cfo.com/applications/2020/12/google-waives-1-5b-in-deepmind-startup-debt/

House Democrats (2020). Investigation of Competition in Digital Markets: Majority Staff Report and Recommendations, *Subcommittee on Antitrust, Commercial and Administrative at of the Committee on the Judiciary*, 116th Cong. https://judiciary.house.gov/news/documentsingle.aspx?DocumentID=5025

Kunert, P. (2019). UK taxman spent six times more with AWS last year than cloud firm paid in corporation tax. *The Register*, 11.06.2019. https://www.theregister.com/2019/06/11/aws_uk_tax_cloud/

———. (2021). AWS tops up the Bezos rocket fund thanks to more money from Brit tax collection agency. *The Register*, 19.02.2021. https://www.theregister.com/2021/02/19/aws_hmrc_contracts/

Jurowetzki, R., Hain, D., Mateos-Garcia, J., & Stathoulopoulos, K. (2021). The Privatization of AI Research (-ers): Causes and Potential Consequences--From university-industry interaction to public research brain-drain?. *arXiv preprint arXiv:2102.01648*. https://arxiv.org/abs/2102.01648

Langston , J. (2020). Microsoft announces new supercomputer, lays out vision for future AI work. *Microsoft AI Blog*, 19.05.2020. https://blogs.microsoft.com/ai/openai-azure-supercomputer

Liang, J., Meyerson, E., Hodjat, B., Fink, D., Mutch, K., & Miikkulainen, R. (2019). Evolutionary neural automl for deep learning. In *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 401-409).

Microsoft (2019a). Age of Intelligence. *Microsoft India Whitepaper,* February 2019. https://news.microsoft.com/en-in/features/microsoft-india-ai-whitepaper-age-of-intelligence/

———. (2019b). OpenAI forms exclusive computing partnership with Microsoft to build new Azure AI supercomputing technologies. Microsoft News Center, 22.07.2019. https://news.microsoft.com/2019/07/22/openai-forms-exclusive-computing-partnership-with-microsoft-to-build-new-azure-ai-supercomputing-technologies/

———. (2022). Microsoft Cloud strength fuels third quarter results. *Microsoft News Center,* 26.04.2022. https://news.microsoft.com/2022/04/26/microsoft-cloud-strength-fuels-third-quarter-results-2/

Mucha, T., & Seppala, T. (2020). Artificial Intelligence Platforms–A New Research Agenda for Digital Platform Economy. *ETLA Working Papers* 76, 06.02.2020. https://www.etla.fi/en/publications/artificial-intelligence-platforms-a-new-research-agenda-for-digital-platform-economy/

Neate, R. (2019). New study deems Amazon worst for 'aggressive' tax avoidance. *The Guardian*, 02.12.2019. https://www.theguardian.com/business/2019/dec/02/new-study-deems-amazon-worst-for-aggressive-tax-avoidance

———. (2021). Amazon had sales income of €44bn in Europe in 2020 but paid no corporation tax. *The Guardian,* 04.05.2021. https://www.theguardian.com/technology/2021/may/04/amazon-sales-income-europe-corporation-tax-luxembourg

Nikolenko, S. I. (2021). *Synthetic data for deep learning*. Springer Nature.

Outeiral, C. (2021). AlphaFold 2 is here: what's behind the structure prediction miracle. *Oxford Protein Informatics Group*, 19.07.2021.

https://www.blopig.com/blog/2021/07/alphafold-2-is-here-whats-behind-the-structure-prediction-miracle/

Phillips, R., Pyle, J., & Palan, R. (2021). The Amazon Method. How to take advantage of the international state system to avoid paying tax. *Study for The Left in the European Parliament*, 22.05.2021. https://www.transform-network.net/en/blog/article/amazon-method-of-tax-dodging-exposed-in-new-research/

Robertson, A. (2018). I'm using AI to face-swap Elon Musk and Jeff Bezos, and I'm really bad at it. *The Verge*, 11.02.2018. https://www.theverge.com/2018/2/11/16992986/fakeapp-deepfakes-ai-face-swapping

Sandbu, M. (2018). How internet giants damage the economy and society. Financial Times, 20.02.2018. https://www.ft.com/content/17460a66-1592-11e8-9e9c-25c814761640

Scott, K. (2020). Microsoft teams up with OpenAI to exclusively license GPT-3 language model. Official Microsoft Blog, 22.09.2020. https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Hassabis, D. (2017). Mastering the game of go without human knowledge. *Nature*, *550*(7676), 354-359.

Simonite, T. (2019). Microsoft Is the Surprise Winner of a $10B Pentagon Contract. *Wired Magazine*, 25.10.2019. https://www.wired.com/story/microsoft-surprise-winner-dollar10b-pentagon-contract/

Statista (2022). Distribution of Google segment revenues from 2017 to 2021. *Research Department,* July 2022. https://www.statista.com/statistics/1093781/distribution-of-googles-revenues-by-segment/

Stolton, S. (2022). EU readies Amazon antitrust settlement. *Politico*, 14.07.2022. https://www.politico.eu/article/eu-readies-amazon-antitrust-settlement/

University of York (2022). What is reinforcement learning? https://online.york.ac.uk/what-is-reinforcement-learning/

Vincent, J. (2019). Former Go champion beaten by DeepMind retires after declaring AI invincible. *The Verge*, 27.11.2019. https://www.theverge.com/2019/11/27/20985260/ai-go-alphago-lee-se-dol-retired-deepmind-defeat

Wiggers, K. (2020). DeepMind's improved protein-folding prediction AI could accelerate drug discovery. *VentureBeat,* 30.11.2020. https://venturebeat.com/2020/11/30/deepmind-claims-its-ai-can-predict-how-proteins-will-fold-with-state-of-the-art-accuracy/

Wheeler, S. (2020). The Hardware in Microsoft's OpenAI Supercomputer Is Insane. *Eningeering.com*, 01.06.2020. https://www.engineering.com/story/the-hardware-in-microsofts-openai-supercomputer-is-insane

Yui, T. (2019). Understanding Neural Networks. *Towards Data Science*, 02.06.2019. https://towardsdatascience.com/understanding-neural-networks-19020b758230

Zoph, B., & Le, Q. V. (2016). Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*.

**Appendix B: Corpus of Document Analysis (Chapter 5)**


Barshefsky, C. (2020). EU digital protectionism risks damaging ties with the US. *Financial Times*, 02.08.2020. https://www.ft.com/content/9edea4f5-5f34-4e17-89cd-f9b9ba698103

Caffara, C. & Morton, F. (2021). The European Commission Digital Markets Act: A translation. https://voxeu.org/article/european-commission-digital-markets-act-translation

Chazan, G. (2019). Angela Merkel urges EU to seize control of data from US tech titans. *Financial Times*, 12.11.2019. https://www.ft.com/content/956ccaa6-0537-11ea-9afa-d9e2401fa7ca

Cini, M., & Czulno, P. (2022). Digital Single Market and the EU Competition Regime: An Explanation of Policy Change. *Journal of European Integration*.

Delcker, J. (2019). Germany's plan to control its own data. *Politico*, 12.09.2019. https://www.politico.eu/article/germanys-plan-to-control-its-own-data-digital-infrastructure/

EPRS (2020). Digital sovereignty for Europe. EPRS Ideas Paper. https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI(2020)651992_EN.pdf.

European Commission (1979). European society faced with the challenge of new information technologies: A community response. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:51979DC0650&from=EN

———. (2018). Coordinated Plan on Artificial Intelligence. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0795

———. (2019). Policy and investment recommendations for trustworthy Artificial Intelligence. https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence

———. (2020a). Towards a next generation cloud for Europe. https://ec.europa.eu/newsroom/dae/redirection/document/70089

———. (2020b). European industrial technology roadmap for the next generation cloud-edge offering. https://ec.europa.eu/newsroom/repository/document/2021-18/European_CloudEdge_Technology_Investment_Roadmap_for_publication_pMdz85DSw6nqPppq8hE9S9RbB8_76223.pdf

———. (2021a). Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

———. (2021b). New Rules for Artificial Intelligence – Questions and Answers. *European Commission Press Corner*, 23.04.2022. https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_21_1683/QANDA_21_1683_EN.pdf

———. (2022). Questions and Answers: Digital Markets Act: Ensuring fair and open digital markets. *European Commission Press Corner*, 23.04.2022. https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2349

European Parliament (2006). A "rapporteur" - the person who presents reports to Parliament. *European Parliament*, 27.07.2006. https://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+IM-PRESS+20060725STO09938+0+DOC+XML+V0//EN

GAIA-X. (2022). Vision & Mission. https://gaia-x.eu/what-is-gaia-x/vision-and-mission/

Goujard, C. & Cerulus, L. (2021). Inside Gaia-X: How chaos and infighting are killing Europe's grand cloud project. *Politico*, 26.10.2021. https://www.politico.eu/article/chaos-and-infighting-are-killing-europes-grand-cloud-project/

Google (2021). Google Response to the DETE Consultations on DSA & DMA. Department of Enterprise, Trade and Employment, January 2021. https://enterprise.gov.ie/en/consultations/consultations-files/google-dsa-submission.pdf

Heikkilä, M. (2022). A quick guide to the most important AI law you've never heard of. *MIT Technology Review*, 13.05.2022. https://www.technologyreview.com/2022/05/13/1052223/guide-ai-act-europe/

Heikkilä, M. & Delcker, J. (2020). EU shoots for €10B 'industrial cloud' to rival US. *Politico*, 15.10.2020. https://www.politico.eu/article/eu-pledges-e10-billion-to-power-up-industrial-cloud-sector/

Joint Research Centre (2021). AI Watch: 2020 EU AI investments. *Publications Office of the European Union*, 27.09.2021. https://ai-watch.ec.europa.eu/publications/ai-watch-2020-eu-ai-investments_en

Juncker J. C. (2014). A New Start for Europe: My Agenda for Jobs, Growth, Fairness and Democratic Change. *European Commission*, 06.05.2021. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52015DC0192

Lyons, K. (2021). Ireland's status as tax haven for tech firms like Google, Facebook, and Apple is ending. *The Verge*, 07.10.2021. https://www.theverge.com/2021/10/7/22715229/ireland-status-tax-haven-google-facebook-apple

Microsoft. (2021). Microsoft supports new rules for gatekeepers. Microsoft EU Policy Blog, 03.05.2021. https://blogs.microsoft.com/eupolicy/2021/05/03/microsoft-supports-new-rules-for-gatekeepers/

Olivi, G. (2022). The future of European AI regulation: Q&A with Brando Benifei, member of the European Parliament and co-Rapporteur on the AI regulation. *Business Going Digital*, 01.04.2022. https://www.businessgoing.digital/the-future-of-european-ai-regulation-qa-with-brando-benifei-member-of-the-european-parliament-and-co-rapporteur-on-the-ai-regulation/

Sanchez-Graells, A. (2021). Where does the proposed AI Act place public procurement? *Blog*, 21.10.2021. https://www.howtocrackanut.com/blog/2021/10/21/where-does-the-proposed-eu-ai-act-place-procurement

Schmitz, L., & Seidl, T. (2022). Protecting, Transforming, and Projecting the Single Market. Open Strategic Autonomy and Digital Sovereignty in the EU's Trade and Digital Policies. *SocArXiv*. https://osf.io/preprints/socarxiv/wjb64/

Stupp, C. (2019). European Cloud Project Draws Backlash From U.S. Tech Giants. The *Wall Street Journal*, 01.11.2019. https://www.wsj.com/articles/european-cloud-project-draws-backlash-from-u-s-tech-giants-11572600600

Von der Leyen, U. (2020). Shaping Europe's digital future: op-ed by Ursula von der Leyen, President of the European Commission. *European Commission*, *19*. https://ec.europa.eu/commission/presscorner/detail/es/ac_20_260

Zuboff, S. (2022). The EU has fired the starting gun in the fightback against Big Tech. Financial Times, 02.05.2022. https://www.ft.com/content/31f49915-0f85-48b0-bf81-131960318967

**Appendix C: Corpus of Document Analysis (Chapter 6)**

Akkaya, I., Andrychowicz, M., Chociej, M., Litwin, M., McGrew, B., & Petron, A. (2019). Solving Rubik's Cube with a Robot Hand. *arXiv preprint* arXiv:1910.07113 http://arxiv.org/abs/1910.07113.

Borkman, S., Crespi, A., Dhakad, S., Ganguly, S., Hogins, J., Jhang, Y.-C., & Kamalzadeh, M. (2021). Unity Perception: Generate Synthetic Data for Computer Vision. *arXiv preprint* http://arxiv.org/abs/2107.04259.

Bremmer, I. (2021). The Technopolar Moment. *Foreign Affairs*, 19.10.2021. https://www.foreignaffairs.com/articles/world/2021-10-19/ian-bremmer-big-tech-global-order

Brud (2022). *Brud is seeking Sr Software Engineer (UE4/C++*. https://forums.unrealengine.com/t/brud-is-seeking-sr-software-engineer-ue4-c-sr-real-time-lighter-ue4/145442

Captain, S. (2018). How Google's DeepMind will train its AI inside Unity's video game worlds. *FastCompany*, 26.09.2018. https://www.fastcompany.com/90240010/deepminds-ai-will-learn-inside-unitys-video-game-worlds

China Law Translate. (2022). Provisions on the Administration of Deep Synthesis Internet Information Services (Draft for solicitation of comments). https://www.chinalawtranslate.com/deep-synthesis-draft/

Cook, D. (2012). Unity interview: Engineering democracy'. *VG247.Com*. https://www.vg247.com/unity-interview-engineering-democracy

Crunchbase. (2022). *List of Epic Games's 18 Acquisitions*. https://www.crunchbase.com/search/acquisitions/field/organizations/num_acquisitions/epic-games-2

Davis, N. (2021). How Volkswagen Group of America visualizes vehicles for 2030 and beyond. *Unity Blog*. https://blog.unity.com/manufacturing/how-volkswagen-group-of-america-visualizes-vehicles-for-2030-and-beyond-0

D'Anastasio, C. (2022). Gaming Giant Unity Wants to Digitally Clone the World. *Wired*, 18.01.2022. https://www.wired.com/story/gaming-giant-unity-wants-to-digitally-clone-the-world/

Edelstein, S. (2018). How gaming company Unity is driving automakers toward virtual reality. *Digitaltrends.Com*. https://www.digitaltrends.com/cars/unity-automotive-virtual-reality-and-hmi/

Epic Games (2021). Announcing MetaHuman Creator: Fast, High-Fidelity Digital Humans in Unreal Engine. *Epic Games News*, 10.02.2021. https://www.epicgames.com/site/en-US/news/announcing-metahuman-creator-fast-high-fidelity-digital-humans-in-unreal-engine

———. 2022a. The Automotive Field Guide'. https://ue.unrealengine.com/ENT-1845-11-2020-Automotive-Field-Guide_LP-Automotive-Field-Guide.html

———. 2022b. 'MetaHuman Creator End User License Agreement'. https://www.unrealengine.com/en-US/eula/mhc

Farseev, A., Yang, Q., Filchenkov, A., Lepikhin, K., Chu-Farseeva, Y.-Y., & Loo, D.-B. (2021). SoMin. Ai: Personality-Driven Content Generation Platform'. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 890–893.

Gartner. (2021b). *By 2024, 60% of the data used for the development of AI and analytics projects will be synthetically generated'*. *Gartner Blog*, 24.06.2021. https://blogs.gartner.com/andrew_white/2021/07/24/by-2024-60-of-the-data-used-for-the-development-of-ai-and-analytics-projects-will-be-synthetically-generated/

Hollister, S. (2020). Unity's IPO filing shows how big a threat it poses to Epic and the Unreal Engine. *The Verge, 24.08.2020*. https://www.theverge.com/2020/8/24/21399611/unity-ipo-game-engine-unreal-competitor-epic-app-store-revenue-profit

Jennings, R. (2021). What's the deal with fictional influencers? *Vox*, 19.10.2021. https://www.vox.com/the-goods/22734620/fictional-influencers-fourfront-fake-tiktokers

Lange, D. (2022). The industrial metaverse: Where simulation and reality meet'. *VentureBeat, 26.01.2022*. https://venturebeat.com/2022/01/26/the-industrial-metaverse-where-simulation-and-reality-meet/

Lim, S. (2020). Puma introduces a virtual influencer in South East Asia called Maya. *The Drum, 17.03.2020*. https://www.thedrum.com/news/2020/03/17/puma-introduces-virtual-influencer-south-east-asia-called-maya

Looking Glass. (2021). Making ruDALL-E fine tuning quick and painless. https://colab.research.google.com/drive/15vFLeepkSTr1qd4xs31g9kMEiwkWP0sh#scrollTo=y-s9LfFBSyty

Murgia, M. (2021). Facebook to build metaverse with start-up that had US military contracts. Financial Times, 24.12.2021. https://www.ft.com/content/18d50b48-f11d-49b2-9287-e5595b25f656

Ong, T. (2020). Virtual Influencers Make Real Money While Covid Locks Down Human Stars. *Bloomberg,* 29.10.2020. https://www.bloomberg.com/news/features/2020-10-29/lil-miquela-lol-s-seraphine-virtual-influencers-make-more-real-money-than-ever

Oreskovic, A. (2018). Why EA's former boss believes the 3D tech that powers video games will make way more money outside of gaming. *Business Insider*, 14.09.2018. https://www.businessinsider.com/unity-ceo-john-riccitiello-opportunity-beyond-gaming-2018-9?r=UK&IR=T

Oleynikov, D. (2021). How Sber Built ruDALL-E: Interview with Sergei Markov. *Blog Article*, 29.12.2021. https://serokell.io/blog/how-sber-built-rudall-e

Pandey, P. (2018). Deep Generative Models. *Towards Data Science*. https://towardsdatascience.com/deep-generative-models-25ab2821afd3

Rasmussen, M. (2021). Video Game Engines Bring Virtual Influencers to Life. *Virtual Humans*, 21.10.2021. https://www.virtualhumans.org/article/the-video-game-engines-bringing-virtual-influencers-to-life

Ruthotto, L., & Haber, E. (2021). An introduction to deep generative modeling. *GAMM-Mitteilungen*, *44*(2), 202100008.

Rothman, J. (2018). In the Age of AI, is Seeing Still Believing? *The New Yorker,* 12.11.2018. https://www.newyorker.com/magazine/2018/11/12/in-the-age-of-ai-is-seeing-still-believing

Sber (2021a). ruDALL-E neural network by Sber for image generation based on Russian-language descriptions now available on ML Space platform. *Sberbank Press Release*, 02.11.2021. https://www.sberbank.com/news-and-media/press-releases/article?newsID=a26a208d-6c72-4f8a-a3b7-aefe1112cbae&blockID=7&regionID=77&lang=en&type=NEWS

———. (2021b). Sber creates ruDALL-E, first multimodal neural network generating images based on Russian-language description. *Sberbank Press Release*, 15.12.2021. https://www.sberbank.com/news-and-media/press-releases/article?newsID=dd220542-ab0b-4c3f-91d4-0e40dd33f0d4&blockID=7&regionID=77&lang=en&type=NEWS

Solis, D. (2022). *These birds do not exist*. https://twitter.com/Daniel-Solis/status/1487913576929103884

Unity Technologies (2018). *Unity Technologies Increases Commitment to Automotive with Dedicated Team.* Press Release, 17.03.2018. https://unity.com/our-company/newsroom/unity-technologies-increases-commitment-automotive-dedicated-team-industry

Tesla. (2021). *AI Day*. https://www.youtube.com/watch?v=j0z4FweCy4M

———. (2022). *Software Engineer, Autopilot Simulation*. https://www.tesla.com/careers/search/job/software-engineerautopilotsimulation-101985

The Influencer Marketing Factory (2022). Virtual Influencers Survey. https://theinfluencermarketingfactory.com/virtual-influencers-survey-infographic

Tiu, E. (2020). Understanding Latent Space in Machine Learning. *Towards Data Science*, 04.02.2020. https://towardsdatascience.com/understanding-latent-space-in-machine-learning-de5a7c687d8d

Tobin, J. (2020). *Lecture 22 Sim2Real and Domain Randomization – CS287-FA19 Advanced Robotics at UC Berkeley*. Instructor: Pieter Abbeel. YouTube. https://www.youtube.com/watch?v=ac_W9IgKX2c

Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., & Abbeel, P. (2017). Domain randomization for transferring deep neural networks from simulation to the real world. *IEEE/RSJ International Conference on Intelligent Robots and Systems.*

Unity Technologies. (2018). Unity Technologies Increases Commitment to Automotive with Dedicated Team. *Unity Technologies Press Release*, 17.03.2018. https://unity.com/our-company/newsroom/unity-technologies-increases-commitment-automotive-dedicated-team-industry

———. (n.d.). Efficient development of simulated environments for autonomous vehicle training. https://unity.com/how-to/simulated-environments-for-autonomous-vehicle-training

Webster, A. (2020). Virtual artist Miquela debuted a music video at Lollapalooza. *The Verge*, 31.07.2020. https://www.theverge.com/2020/7/31/21348119/miquela-hard-feelings-music-video-lollapalooza

Wikimedia (2022). GAN Architecture. https://commons.wikimedia.org/wiki/File:A-Standard-GAN-and-b-conditional-GAN-architecturpn.png

Whittaker, L., Kietzmann, T. C., Kietzmann, J., & Dabirian, A. (2020). All Around Me Are Synthetic Faces: The Mad World of AI-Generated Media'. *IT Professional*, *22*(5), 90–99. https://doi.org/10.1109/MITP.2020.2985492..

Wood, E., Baltrušaitis, T., Morency, L.-P., Robinson, P., & Bulling, A. (2016). Learning an appearance-based gaze estimator from one million synthesised images'. *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, 131–138.

Woolf, M. (2022). *Make Your Own AI-Generated Pokémon*. Google Colab, 13.01.2022. https://colab.research.google.com/drive/1A3t2gQofQGeXo5z1BAr1zqYaqVg3czKd?usp=sharing