ONTOGENETIC EPISTEMOLOGY OF AI:

An Investigation into the Role of Machine Learning in Knowledge Production Through its Genesis and Genealogy

Student: Kaush Kalidindi

Supervisor: Prof. Dr. Ir. M. Boon (Mieke)

Second Reader: Dr. M.H. Nagenborg (Michael)

Philosophy of Science, Technology, & Society Programme (2021-2023)

University of Twente, Netherlands.

ABSTRACT

Through a reconstruction of Gilbert Simondon's work using perspectives from contemporary philosophy of science, I pursue an ontogenetic study of machine learning —an investigation into the process by which machine learning models become what they are and do what they do —with the aim of understanding and addressing their role in knowledge production. This investigation unfolds as two parts: Genesis and Genealogy. By studying the genesis of the ML-model, I seek to identify the role (or lack thereof) of human cognition in the formation of the ML-model and illustrate its significance for our knowledge about the world. By studying the genealogy of the ML-model, I aim to address how different scientific communities and their respective knowledge practices are affected by their interaction with and integration of machine learning models.

Contents

CHAPTER 1

ONTOGENETIC STUDY OF TECHNICAL OBJECTS

INTRODUCTION

As AI technologies increasingly take center stage in societal discourses, they invite rigorous reflections from a multitude of disciplines, with psychology examining the impacts of AI on human behavior, sociologists studying the broader changes to societal structures that it instigates, political science reflecting on how it influences governance and power dynamics, philosophy of mind investigating the nature of consciousness in AI, prompting questions about its potential for genuine understanding in a cognitive sense, and legal scholars investigating the regulatory measures, intellectual property issues concerning AI applications. Amid this multidisciplinary reflection, two branches of philosophical exploration emerge as particularly salient: i) the ethics and philosophy of technology (PhilTech) and ii) Philosophy of Science (PhilSci).

PhilTech grapples with the profound moral implications and societal impacts of AI, raising essential questions about various ethical concerns arising out of its integration into societies and its potential to disrupt our existing norms and values. However, when reflecting on these issues, PhilTech in approaching AI as an object of investigation, fails to take into account the process that leads to it, the various underlying research practices that make possible the production of AI-applications and infuse them either intentionally or unintentionally with the socially disruptive potential they carry with them. Herein lies the particular space that falls under the purview of Philosophy of Science. PhilSci dives into the very heart of AI's methodological and foundational underpinnings, querying its epistemological basis, its underlying assumptions, and the nature of the scientific practices AI-applications arise out of.

Given that many of the ethical concerns arise from the technical structures underlying AI-technologies, an intersection of PhilSci and PhilTech (a philosophy of technoscience) can help provide a more comprehensive understanding and normative insights concerning AI's role in society. Such an intersection can be traced to the burgeoning field of Ethics & Epistemology of AI (Russo et al., 2023). Initiated by philosophers of science such as Federica Russo, EthEpiAI aims to address ethical issues concerning AI stemming from the ways in which Machine Learning models represent knowledge about the world (epistemology) and it does so by enabling an understanding of the process that leads to them.

It is within this interdisciplinary field of the philosophy of technoscience that this project is situated in. This thesis follows an investigation into AI with the same aims and motivations as that of EthEpiAI in that I do seek to uncover the ways in which Machine Learning models represent knowledge about the world. However, instead of starting from PhilSci literature (as in the case of Russo et al. (2023)), I carry out this investigation by drawing on the work of the 20th century philosopher of technology, Gilbert Simondon.

Gilbert Simondon is a French philosopher whose contributions to the field of philosophy of technology in his *Du mode d'existence des objets techniques* (1958), have largely gone unnoticed, with his work only being translated into English more than half a century after its initial publication. Decades before the advent of the field of technoscience, Simondon's work already addresses the particular intersecting space between PhilTech and PhilSci, in his efforts to relate technologies to culture by providing an conceptual framework to understand their underlying modes of technical operation and the process of invention.

Central to Simondon's (1958/2011) project is the idea that technologies (or technical objects as he would refer to them as[1]) carry an identity that is independent of their practical ends. This is an identity concerning the knowledge they embody, both of their own internal operational schema (how a

[1] I shall follow this terminology of the "technical object" in the rest of this thesis, so as to better distinguish its material nature as that of a concrete technical artefact, as opposed to the a Heidegger's (1977) understanding of 'technology' as a human activity or Ellul's (2021) characterization of 'technology' as worldview.

technical object operates the way it does) and the knowledge about the real-world phenomena that they manipulate in their operation. This 'epistemic' identity, for Simondon, can only be accessed by understanding the ontogenesis of the technical object, the process by which a technical object becomes what it is and does what it does. Simondon's focus on the 'becoming' of the technical object is analogous to the focus of EthEpiAI on the process that leads to the development AI-Technologies.

Unlike previous philosophers of technology in the French tradition who drew on Simondon's work such as Bernard Stiegler (1998) and Yuk Hui (2016), I approach Simondon as philosopher of science, and in doing so, I shall reconstruct his work by situating it within and relating it to contemporary PhilSci literature, albeit in the context of Machine Learning and its role in knowledge production.

Research Question:

In drawing on Simondon's work and situating it within contemporary PhilSci literature, the primary research question that this thesis aims to address is, **How can an ontogenetic investigation of machine learning—an investigation into the process by which machine learning models become what they are and do what they do—enable an understanding of the role it plays in knowledge production and help address how it is (and whether it should be) incorporated into knowledge practices.**

Sub-questions:

- [Part 1 : Genesis] When employing machine learning in knowledge production, how is our knowledge about the world affected by the process through which machine learning models come into existence?
- [Part 2 : Genealogy] How does the mode of knowledge production specific to machine learning models interact and integrate with other epistemic communities and their respective knowledge practices?

To address the above research question (and the sub-questions), I shall first set the stage in this chapter and explain what constitutes ontogenesis (1.1) and in doing so contrast it with the ontological approach of Hylomorphism, which Simondon seeks to critique in his work. Having explained ontogenesis as an approach to ontology, I shall then in 1.2 outline what constitutes an ontogenetic study of technical objects and then illustrate how it relates to approaches in PhilTech and PhilSci (1.3 & 1.4). I will then in 1.5 conclude my overview of an ontogenetic study by outlining what its significance is in the context of Machine Learning models.

After setting the stage in this chapter, the rest of this thesis is split into two parts, investigating the two aspects of ontogenesis, Genesis (Part I; Chapters 2 & 3) and Genealogy (Part II; Chapters 4 & 5).

Chapter 2 will investigate the genesis of technical objects and in doing so bring to surface the role of human cognition in the activity of invention. The central question this chapter would then seek to answer is: How is our knowledge about the world affected by the role of human cognition in the genesis of the technical object? Chapter 3 continues this discussion in the context of Machine Learning and addresses the question: How is our knowledge about the world affected by the role (or lack thereof) of human cognition in the genesis of the Machine-Learning model?

While Genesis (Part I) focus on the coming into existence of a single technical object, Genealogy (Part II) shifts attention to how the technical object carry a long lineage of geneses, with technical objects moving from the community they arise out of and being instrumental in the production of new knowledge and invention of new technical objects in other communities. Chapter 4 explores this genealogical aspect of ontogenesis and aims to answer the question: How do scientific communities extend their knowledge practices by interacting with and integrating technical objects that emerge from other scientific communities? Chapter 5 extends this discussion to Machine Learning models to answer the question: To what extent, and under which conditions, can scientific communities extend their knowledge practices by incorporating machine learning models, compared to how they incorporate other technical objects?

1.1 ONTOGENESIS AS A MOVE AWAY FROM HYLOMORPHISM

Ontogenesis concerns the development of the individual; it is the process by which the pre-individual *becomes* the individual, or as Simondon would put it, it is the event that *individuates* the individual. Despite being traced in its origins to Greek metaphysics[2], ontogenesis, today, is rarely studied in philosophical literature that engages with ontology and is instead most commonly employed in biology to refer to the development of an organism/ecosystem[3]. Central to Simondon's larger project is a revitalization of this ontogenetic approach to metaphysics against the dominant hylomorphic tradition. Hylomorphism, as a philosophical doctrine traced back to Aristotle's substantialism, advocates *being* as a composition of form (*morphe*) and matter (*hyle*) and that form actively determines matter, which remains passive and inert (Piatti, 2016). Despite not being explicitly maintained, hylomorphism has been proven powerful and persisted within numerous philosophical and scientific schools of thought: from the Kantian distinction between a priori forms of intuition and sensible matter, to this idea from developmental genetics that genetic code governs the development of the individual being and from anthropology the idea that material culture stems from the imposition of human forms on the environment. (Voss, 2020).

For Simondon, the hylomorphic schema is insufficient for explaining the "genesis" of the individual because of its pre-supposition of a union of form and matter, and because it subsumes matter to form in its characterization of fixed forms as imposed on homogenous matter (Barthélémy, 2012). For instance, someone who works with wood as a material cannot just impose the idea of a pre-determined chair onto the wooden block. They would have to recognize its porous/non-porous, elastic/resistant nature and yield to the range of chair designs permitted by the material properties of wood. Therefore, Simondon would argue against the idea of a worker who has her product in mind before she produces it or a scientist who would understand technology as a mere application of already

---

[2] As the *causa efficiens (κινοῦν).* See Aristotle, *Physics II 3, 194b29* in Apostle (1970)

[3] See Hill et al. (2010) for an overview of the usage of ontogenesis in developmental biology.

determined laws onto problems. Instead, there is an intermediate space between form and matter where ontogenesis is played out, a space characterized by a simultaneity of forms being materialized, and matter being rendered formable. Rather than forms being imposed on matter, ontogenesis acknowledges the qualities in matter that forms bring out and facilitate (Marks, 2006, p. 5).

1.2 ONTOGENESIS OF THE TECHNICAL OBJECT

In his *Du mode d'existence des objets techniques* (1958), Simondon's critique to hylomorphism unfolds as an investigation of ontogenesis in the context of technology. An *ontogenesis* of technical objects distinguishes itself from traditional substantialist and essentialist ontologies —in that the latter would identify the technical object-artefact based on what it 'is' and what it 'does' (e.g., its utility, the materials that it is made of, the effects it has on society and how it is in turn affected by society) whereas an ontogenetic study would investigate the event-qua-process by which the technical object becomes what *it is* and does what *it does* (e.g., how it is invented, how it evolves, how it is maintained). For instance, in the case of the analog clock as a technical object, a hylomorphic ontology that focuses on its existence would identify the technical object by the *form* that the human imposes onto it. Accordingly, it would identify the clock as that which represents the flow of time (by its utility), that which fueled economic growth during the industrial revolution (by its effect on society), that which was societally constructed to meet the demands of a mechanical life (by how it is affected by society).

In contrast, an ontogenetic study of the clock would shift attention to, say, how the human-engineer identifies mechanisms by which potential energy in the clock can be translated into a consistent movement that can be rendered cognizable (in the act of invention; Chapter 2), and transformation of technical schema of the clock from one driven by a pendulum to one driven by a loaded-spring (in its evolution; Chapter 5), and the necessary human interventions to ensure the reliability and precision of its functioning over time.

From the case of the mechanical clock, it is evident that what an ontogenetic study of technology draws attention to, that a philosophy rooted in the existence of a technology does not, is i) to its internal technical schema and ii) the role of the human in discovering and establishing this technical schema as a material bridge (during invention) between a human reality (e.g. time-keeping as the form-intention) and natural reality (e.g. physical properties of the loaded spring).

Simondon outlines how the historical cultural neglect of the technical activity led to a misconstrued identity of technology based on its utilitarian outcomes. This is because a unique technical schema can be operationalized towards a wide range of practical ends, and a specific usage can be fulfilled by several technical schemas. Simondon illustrates this misidentification using the following example:

> "a steam engine, a gasoline engine, a turbine, and an engine powered by springs or weights are all equally engines, but there is a more genuine analogy between a spring engine and a bow or a cross-bow than between the spring engine and a steam engine; the engine of a pendulum clock is analogous to a winch, while an electric clock is analogous to a door bell or a buzzer. Usage unites these heterogeneous structures and operations under the banner of genera and species (...)" (Simondon, 1958/2011, p.25)

This generalized mis-identification (or rather, reduction) of technical objects by their ends, results, and effects and not by their technical schemas is attributed by Simondon as a symptom of modern culture and a cause of alienation between the human and technical object. Following Simondon, it would then be important to recognize that the technical schema of say, the spring-engine, in my previous example, has an identity independent of its utility as the analog clock. Half a century after being written, Simondon's project finds great relevance in our era characterized by an increased cultural alienation that arises out of the distance between complex technical schemas internal to modern technologies and the communities that employ them. More importantly, what makes Simondon's contributions of particular significance now is that this mis-identification of technical objects by their results is not just a symptom of culture but is also embedded within academic and scientific research that concerns technology. In what follows in the next two sections of this chapter, I shall sketch an overview how an ontogenetic approach to technical objects relates to contemporary approaches in Philosophy of Technology, and Philosophy of Science.

1.3 HYLOMORPHIC UNDERPINNINGS OF PHILTECH

The hylomorphic schema, as laid out above, with its emphasis on what happens at the two ends, *form* and *matter*, and its subsequent failure to acknowledge what happens between the two ends can be traced to literature in Philosophy of Technology. PhilTech approaches technical objects by their results (their ends, their utilities, and their effects) but not based on their technical schemas and the practices they arise out of. To add more weight to this claim is Federica Russo's recent observation of the same. In her Techno-Scientific Practices (2022), she takes on the task of creating a bridge between the fields of PhilTech, STS, and Philosophy of Science (PhilSci), as a part of which she shifts attention to the origins of modern technologies in scientific practices and identifies how technical instruments have not been analyzed as a part and parcel of the scientific process (p. 65). As a part of her critique, she argues that "[PhilTech] already approaches technology and artifacts as "finished" final products and too little on the process that leads to them" (2022, p. 43). Russo's observations thereby align with my own motivations in drawing on Simondon's work to address the particular space, of the genesis of technical artefacts, that PhilTech fails to accommodate.

Approaches in PhilTech play an essential role in modern society in their descriptive and normative contributions to the relationships between technology and society at different levels and there are great risks, as will be elaborated in later chapters, of not taking into account the ontogenetic aspect of the very technologies that are objects of their research. In summary, drawing attention towards the ontogenetic aspect of technical objects brings to surface the particular ways in which the technical activity exerts its own deterministic influence (in shaping how technologies are later instrumentalized for a particular utility or an effect on society) which is otherwise either unquestioned or remains as an implicit ideology in most, if not all, approaches in PhilTech.

1.4 ONTOGENESIS AS EXTENDING THE PRACTICE-TURN IN PHILSCI

It is common knowledge that technology is a part and parcel of scientific research, and one wouldn't be wrong in directing their attention to scientific practices and philosophy of science when investigating the genesis of technical objects. While it is true that technologies do emerge out of scientific research, the characterization of the ways in which this emergence is played out remained hylomorphic within philosophy of science as well, at least until recent years.

In the same way that the hylomorphic schema remained dominant within PhilTech, in its reduction of the technical object to its results, technical instruments in traditional positivist philosophy of science have been reduced as mere applications of scientific theories. In this theory-centered paradigm of philosophy of science, technical objects-qua-instruments had value only in so far as they were instrumental in evaluating and justifying a theory (Boon, 2015, p. 4) and any empirical achievements of these instruments in subsequent technological developments have instead been credited as a testament to the success of theories they were allegedly derived from (Boon, 2015, p. 27). If within the bare hylomorphic schema, *being* is the result of form imposed on matter, and if PhilTech is hylomorphic in so far as it understands technology as the imposition of human necessities on material reality, then traditional PhilSci mirrors the hylomorphic schema in its conception of 'instruments' as the imposition of scientific theories on physical reality.

However, this supremacy of theory and the conception of the instrument as being in its service has been recently deconstructed with the practice turn in PhilSci. Traced in its origins to the works of new experimentalists such as Ian Hacking and Allan Franklin, the practice turn brought to surface the active role that instruments play not just in confirming pre-made theories but in the construction of theories as well (Russo, 2022, p. 41). The hylomorphic schema is thereby avoided because instruments are not simply derived from the imposition of scientific knowledge. In fact, knowledge of phenomena itself is co-produced in the interactions between theories and instruments (Russo, 2022, p. 67).

Transitioning away from a theory-centered approach also meant that the practice turn brought with it a pluralist conception of science in challenging the notion of there being a single comprehensive account of scientific knowledge. For instance, Roland Giere's (2004) seminal argument that scientific statements do not directly refer to an aspect of reality, but that scientists *use* certain models to *represent* aspects of reality for certain *purposes,* carried with it a practical motivation, thereby making it possible for mutually inconsistent models of the same phenomena to co-exist in the scientific project in so far as they are directed at different purposes. The focus on *purposes* also signified that scientific inquiry is not exclusively driven towards uncovering 'truth', but is also driven with the goal of developing instruments that produce phenomena of utilitarian relevance in later technologies (Boon, 2015, p. 8, 22). The practice turn with its deconstruction of the supremacy of the theory thereby shifts the focus of science from questions of truth to questions of utility and reliability.

Herein lies a danger that Philosophy of Science in Practice ought to remain wary of, in that the focus on the utilitarian relevance of scientific knowledge, and not on truth, opens doors for the return of an alternate hylomorphic schema, identical to one that dominates PhilTech, with scientific activities and their technical outputs being reduced to their utilitarian relevance. The practice turn, in moving away from the supremacy of the theory, inverts the hierarchy of values rendering science as not more true but more useful. This inversion while dethroning the theory, continues to preserve an opposition between the norm of utility and the norm of truth, leading to an anti-realist stance on scientific knowledge, as is traced in the works of Boon (Boon, 2020b) and Van Fraassen (2010).

It is precisely within this opposition that Simondon's ontogenetic project finds its greatest relevance for PhilSci in its nuanced demarcation of the *operational* from the *practical*. In Simondon's words,

> "The technical operation is not arbitrary, pliable in every way to the whims of the subject according to the randomness of immediate utility; the technical operation is a pure operation that puts into play the veritable laws of natural reality; the artificial is something natural that has been solicited, not something false or human that has been mistaken for something natural" (Simondon, 1958/2017, p. 260)

The technical operation, situated between nature and science, thereby retains an epistemic identity, grounded in the norm of truth that is independent of its utilitarian ends and is yet capable of being operational-ized towards these ends[4]. It is important to note that Simondon's demarcation of the *operational* from the *practical,* as avoiding an epistemic nominalism, is not a notion that is entirely foreign to contemporary PhilSci. The necessary groundwork has already been done. Michela Massimi, for instance, does make a similar distinction in her Perspectival Realism (2022). She identifies scientific knowledge production to consist of two distinct perspectival modes. Perspectival$_1$ [P_1] (from a vantage point) constitutes the situatedness of scientific representation, in that the way a phenomenon is identified and assessed is affected by the experimental, theoretical, and technological resources specific to a particular scientific community. Perspectival$_2$ [P_2] (towards a vantage point) on the other hand refers to the directionality of the representation itself (Massimi, 2022). This entails the practical ends *towards* which a particular phenomenon is represented as opposed to how it is represented *from* a particular vantage point.

What Simondon would add to Massimi's demarcation is that technical resources aren't simply resources as a part of P_1 and the human isn't simply passively connecting these two modes, P_1 and P_2. Instead, technics serves as a bridge that connects these two modes. The technical object which produces a technologically relevant phenomenon of utilitarian value is not one that already exists as a resource or is arrived at simply by an application of existing theoretical and experimental resources. It is one that is invented during the technical activity by an active human cognitive operation that connects P_1 and P_2, as one that is grounded in the real[5]. As opposed to simply directing P_1 (resources) towards P_2 (ends), the idea that the technical object embodies a knowledge owing to its bridging of the two realms while being grounded in the real renders every technical object as one that is serving both epistemic and utilitarian tasks. Or as Simondon would put it, "there are not two categories of technical objects, those serving utilitarian tasks and those serving knowledge; any technical object can be scientific and vice-

[4] This line of argument has the potential to be extended into a realist stance on scientific knowledge from within a philosophy of science in practice – that is a task for a future research project.
[5] The role of human cognition that Simondon introduces isn't foreign to philsci as well. (Boon, 2022) identifies the same and this will be contextualized and elaborated in the next chapter.

versa." (Simondon, 1958/2017, p. 252). The type of epistemology Simondon advocates here with respect to technical objects is similar to one that PhilSci employs with models. Russo (2022) shows how models mediate between reality and theory and allows one to learn about the two sides they connect whereas technologies are only employed (p.100). For Simondon, however, technical objects themselves, would embody both these realms, utilitarian and epistemic. An ontogenetic study of technical objects thereby presupposes an epistemological dimension as well. This does not mean that model-instrument distinction is conflated, but rather that one now begins to approach the model not as model, but as a modelling –a practice that takes place during the act of invention in translating an abstract technical object (one grounded in P_1) to a concrete technical object (one grounded in the laws of the real world), all while being directed towards P_2, and yet remaining independent of it.

It is a pity that Simondon's work remained underrepresented for more than half a century within Anglo-American philosophy literature. However, if it had gotten the attention it deserved at the time of writing, it would have confronted the then mid-20th century philosopher of science who was just beginning to acknowledge the practice turn; it would have perhaps denied her the optimism with which she leapt into the norm of utility and perhaps even denied the rest of the scientific community from recognizing the benefits of an engineering-paradigm in science. In the current era, however, the time is right for Simondon's project to be integrated into both PhilSci and PhilTech and also as that which integrates the two. My elaboration of Simondon's distinction between the *practical* and *operational* through the works of contemporary philosophers of science Russo, Massimi, and Boon is in fact an anticipation of how in later chapters, I shall reconstruct Simondon's project not as one that externally confronts the philosopher of science, but one that is built from within PhilSci itself; a reconstruction that does not confront the scientist-engineer with a warning of the dangerous path ahead (as it would have if it got the same reception at the time when Simondon wrote it) but instead as a guide in helping navigate the epistemological issues that currently lie at the frontiers of technoscience, namely those epistemological issues concerning the complex technical object that is the Machine Learning model.

1.5 TOWARDS AN ONTOGENETIC EPISTEMOLOGY OF AI

My reconstruction of Simondon's project as an Ontogenetic Epistemology of Technical Objects (OETO) is definitely guided by this idea that the 'time is right', with i) the practice turn in PhilSci developing to a point where Simondon's project would find its best relevance as was elaborated in the previous section, ii) PhilTech becoming aware of its own self-defeating disconnection with the process by which technologies come into existence, and lastly, iii) with authors like Russo (2022) and Radder (2009) pushing forward the need for an integration between PhilSci and PhilTech.

At a broader level, it is important to note that OETO is then an intuitive response to the above outlined developments. Although pursuing an ontogenetic study would no doubt help identify and refine certain methodologies within both PhilTech and PhilSci, it is definitely not the case that these fields cannot continue operating without an OETO, or that there is a grave need for them to fundamentally re-think their ontological commitments to be aligned with those of an ontogenesis.

However, what makes OETO, not just an intuitive response but a *necessary* response, is the particular role it can play in guiding these fields in their confrontations with the 'alien' subject that is AI. In what follows, I shall sketch an overview of why existing resources in PhilTech and PhilSci are insufficient in addressing certain epistemological concerns that arise out of the scientific development and societal implementation of Machine Learning models and identify the role that an OETO can play in this context.

Firstly, as was discussed in 1.3, the hylomorphic undertones of PhilTech approaches imply that the process by which technologies come into existence (ontogenesis) is not explicitly acknowledged and that technical objects are instead assessed based on their results (e.g., utilities, effects on society, etc.), or sometimes based on what is needed for their ontogenesis, but not ontogenesis itself (e.g., capital that goes into scientific research; societal, political and cultural influences in driving the course of technological innovation, etc.). Such a strategy has been proven successful in the past in the anticipation and evaluation of emerging and potentially disruptive technologies, but it reaches a limit when PhilTech enters into the domain of AI. This limit can be illustrated as follows: let's say there is a strong societal need for a technology that fulfills a certain necessity. There are two technical schemas that fulfill this exact necessity in identical ways producing identical results in all contexts. It is the case that one of these schemas arises out of a fully human written algorithm and the other schema corresponds to a complex machine learning model with minimal human intervention. In this context, PhilTech approaches would be incapable of justifiably distinguishing between these two cases given that from the outside (when one looks at what goes in, and what comes out, as in a hylomorphic schema), the two technical activities appear to be identical. An OETO in the context of ML models can help de-mystify the technical activity because of which ML models become what they are and do what they do, and can help PhilTech make more informed decisions regarding their implementation in different social, political, and cultural spheres.

Secondly, it is well-recognized by now, in contemporary PhilSci, that knowledge of phenomena is not produced by human agents alone by a direct application of theoretical, and methodological resources. Instead, Russo (2022) outlines how knowledge is co-produced in the interactions between human scientists and technical instruments and how instruments do not just mediate, but embody a certain epistemic agency. With her revitalization of the concept of poesis, Russo brings to surface the interaction and partnership between human and artificial epistemic agents in knowledge production. However, it is important, especially in the context of ML models, to investigate how this partnership /interaction is played out and how it is different from traditional instruments. Furthermore, it is important to recognize that when PhilSci literature approaches technical objects as artificial epistemic agents, they fail to acknowledge the human reality and the role of human cognition (or the lack thereof) in the ontogenesis of technical object which is otherwise labelled “artificial”. In doing so, one runs the risk of confronting the ML model as something external and alien to the human-scientist who participates with it in scientific practices, whereas in reality, the ML model, like other technical objects, follows a long line of evolution, carrying traces of both human and material features that cumulatively shape it into what it is now.

More importantly, the advent of the AI paradigm also marks the practice-turn in PhilSci confronting a limit. This is because the ML model, with its emphasis on empirical results, its lack of theoretical underpinnings, and the subsequent black-boxing of methodological, epistemological operations, serves as the absolute embodiment and hallmark of what it means to fully embrace the norm of *utility* over the norm of *truth*. How different scientific disciplines will incorporate ML models into their practices in the coming years will speak a lot about the scientific enterprise as a whole and where its commitments lie. An unhindered adoption of ML models in scientific practices would mark the completion of practice-turn, with an absolute shift from a traditional positivistic philosophy of science to one that defines what is *true* by what is *useful*. An OETO in the context of AI, with its demarcation of the *operational* from the *practical,* would serve as a guide in keeping the dominance of *utility* at bay. It would do so by equipping scientific practices with resources to reinvigorate their commitments to the norm of *truth* by strengthening their methodological and epistemological commitments, while continuing to direct technological innovation towards the norm of *utility*.

Simondon's own motivations in investigating the genesis of technical objects stemmed from technology being reduced to its *finality* (results) and the technical object not being understood through the technical activity that it arises out of. Simondon's work was thereby directed at suppressing the cultural alienation between the human and the technical object that arises out of this misidentification. However, in the age of AI, investigating the genesis is not just relevant for suppressing this alienation and better representing the ML model in human culture. Ontogenesis, in the context of ML models, tackles deeper epistemological concerns as well.

In Part I: Genesis, as I layout the role played by human cognition (or the lack thereof) in the invention of the ML model, it is not to demystify the construction of the ML model to reach a better understanding of AI. Instead, it is valuable in so far as recognizing the particular role that human cognitive resources, and theoretical resources play in the construction of the ML model can help better address questions of accountability and reliability of ML-driven knowledge production.

In Part II: Geneology, as I trace the genealogical lineage of ML models from its statistical foundations to their operationalization in distinct knowledge practices, it is not the question of milestoning their increasing empirical successes in diverse fields. Instead, it is to investigate how machine learning as an epistemic tool transforms itself by interacting and integrating with existing epistemic communities and knowledge practices. Only then can we identify whether some forms of integration of ML models are more justified than others.

PART I

GENESIS

CHAPTER 2

GENESIS OF THE TECHNICAL OBJECT

In the last chapter, I have shown how PhilTech and Positivistic PhilSci preserve a hylomorphic schema in their understanding of technical objects as derived from applying existing scientific knowledge onto human needs. This chapter uncovers how this is not such a straightforward process and that we do not directly arrive at technologies from applying existing knowledge. Instead, human cognitive capacities play a crucial role in the genesis of the technical object. This claim will be substantiated by drawing on Simondon's conception of invention as an event that is made possible by the human creative capacity to imagine a future technical object by identifying and organizing existing scientific models in precise ways. I will further strengthen this claim, and in doing so go beyond Simondon's own arguments, by drawing on PhilSci literature to show how human cognitive capacities play a role not just during invention, but also during the construction of the scientific model itself. Having identified the role of human cognition in both *construction* of the scientific model, and the subsequent *invention* of the technical object, this chapter will then outline the epistemological significance of this role, namely, how is our knowledge about the world affected by the fact that the genesis of the technical object relies on human cognitive capacities? Given the questionable role of human cognition in the genesis of the Machine-Learning object, answering this question will help better understand the adequacy of Machine Learning as an epistemic practice in Chapter 3.

2.1 INTRODUCTION

For Simondon, an investigation into the genesis of technical objects directs our attention to the particular moment of invention, an ontogenetic event before which the technical object only exists virtually, as a non-actualized potential in nature. The ontogenetic event, as laid out by Simondon over the course of his book, is one that always has both an ontological and epistemological dimension. This is because the genesis of the technical object does not just lead to its material manifestation as its existence; it also brings with it the knowledge of the internal operational schema that it embodies; and knowledge about the phenomenon that the technical object produces (or manipulates).

A technical object that is invented is then tied to the phenomenon that it seeks to produce (or manipulate) as a result of its functioning. For instance, a gyroscope is tied to our understanding of the phenomenon of conservation of angular momentum, which allows it to maintain its orientation; and a photographic filter is tied to the production of polarization as a phenomenon that reduces glare and reflection in an image. In these examples, it is important to recognize that the invention of these technical objects, as directed towards the production of particular phenomena, presupposes some form of epistemic access to those phenomena as a condition of their genesis. Scientific models provide this initial epistemic access to the world (or a phenomenon in the world) as a pre-requisite to the genesis of the technical object.

Although Simondon does not explicitly refer to scientific models, he distinguishes between the *abstract* technical object and the *concrete* technical object. *Abstract* technical objects are a translation of an intellectual system (like how Giere (2004), conceptualizes scientific models as abstract objects derived from theory). *Concrete* technical objects, in contrast, are derived from the abstract technical object, and are embedded into real-word causes and effects following the activity of invention (Simondon, 1958/2017, p. 25-29). I thereby infer from Simondon's work that the *abstract technical object* he points to is indeed similar to the way the *model* is employed in the semantic view in PhilSci. The semantic view, traced in the accounts of Suppe (1989) and Giere (2010). Cartwright (1983) and Morgan & Morrison (1999) (as cited in Boon 2020a) corresponds to the notion that there is a representational relationship between scientific models and real-world systems analogous to how semantic signs and symbols represent real world entities. The invention of the concrete technical object for Simondon, then starts from these 'representational'-models or what he calls abstract technical objects, but remains a process that is not fully directed by them. Instead, the human plays an active role, as will be discussed in section 2.2, by weaving together elements from these models and other epistemic resources towards particular practical ends by employing cognitive resources that are unique to her. Drawing on contemporary developments in PhilSci, particularly the conception of scientific models as epistemic tools and the emphasis on the process of model construction (Boon and Knuuttila, 2009), I shall in 2.3, deviate from this Simondonean conception of the genesis of the technical object in arguing how the ontogenetic event is not just confined to the activity of invention alone but that it extends to the construction of the scientific model as well. To make my position explicit, I will be extending Simondon's attribution of the human role to not just the activity of invention of the technical object but also the construction of the scientific model, in that in the same way the genesis of the technical object does not follow, for Simondon, from a direct application of scientific models (abstract technical objects), the activity of model construction too does not arise from a direct translation of existing theoretical resources. It is to be noted that from here onwards, to maintain clarity for the reader and avoid confusion, I will be referring to *scientific models* as *models* indeed, and not as *abstract* technical objects in the way Simondon refers to them, and that whenever I use the term '*technical object*', I'm in fact referring to a '*concrete*' technical object in particular and not the *abstract* technical object that is the *model*.

The genesis of the technical object i.e., the ontogenetic event, then, follows two activities: i) the construction of the model, ii) the invention of the technical object. The aim of this chapter is to investigate the ontogenetic event in both these activities and identify the role of human cognition, with the next chapter extending this investigation in the context of machine learning in particular. Accordingly, I will first, in 2,2 investigate what constitutes the activity of invention as laid out in Simondon's work. After laying out the need for human cognitive abilities during the activity of invention, I will show how the human fulfills this need through her unique creative capacities, not only during the invention of the technical object, but also in the construction of the scientific model as well (2.3). Lastly, 2.4 will explore the epistemological implications of there being an active human involvement in the construction of the model and in the invention of the machine. In other words, I will be concluding this chapter, by answering the question, how is our knowledge about the world (both the world that the technical object operates on, and the world within the technical object through which it operates) affected by the fact that human cognition plays a necessary role during the genesis of the technical object?

2.2 INVENTION OF THE TECHNICAL OBJECT

Before I dive into Simondon's characterization of the activity of invention, it is important to distinguish invention's unique contribution to the ontogenetic event from that of the construction of the scientific model. Models play a crucial role in scientific practices in that they provide an interface between the epistemic agent-scientist and the real-world system under investigation (Russo, 2022, p. 93). While there is a vast body of literature in PhilSci on what constitutes a model, Contessa (2010, as cited in Boon (2020b)) distinguishes their ontological status into the following three types: material (e.g., anatomical models of body organs used by medical professionals), mathematical (e.g, epidemiological models such as the SIR model to understand the spread of diseases), and fictional models (e.g., the plate-tectonics model which imagines the movement of earth's lithosphere as an interaction between plates, while the actual process is driven by complex geophysical phenomena). Each of these types of models play an instrumental role in producing knowledge about a phenomenon, which is further employed in the invention of a technical object that produces (or manipulates) this phenomenon (Boon, 2015). Why is it that there is then a distinct need for the activity of invention when scientific models already provide a comprehensive understanding of the underlying phenomena? Isn't Fermi and Szlizard's mathematical model of nuclear chain reaction sufficient to build a nuclear reactor? Isn't Carnot's ideal heat engine model sufficient for building internal combustion engines?

It is important to note that the model itself does not carry with it all the sufficient epistemic resources for the invention of the technical object. This is because of the following reasons: Firstly, models themselves, even if they are material, cannot serve as *concrete* technical objects because of their dependency on the contexts of their construction, in that they cannot function beyond the scope of the controlled conditions created within say, the laboratory; Secondly, models do not give us a direct, unmediated epistemic access to an objective reality (Russo, 2022, p. 101). The ways in which scientific models correspond to a real-world system is further mediated by the theoretical and experimental resources specific to the epistemic community that they arise out of (Massimi, 2022, p. 22); and lastly, models are instrumental in knowledge production in so far as they help isolate the relevant factors and simplify an otherwise complex phenomenon (Russo, 2022, p. 101-102) or as Giere (2006, as cited in Russo, 2022, p. 103) puts it, they serve as maps, in guiding epistemic agents in reaching a particular purpose, but they cannot alone fulfill this purpose.

If scientific models can only mediate, isolate, and guide access to phenomena as discussed above, how is it possible that the inventor-engineer is able to invent *concrete* technical objects that produce phenomena in the real world? Answering this epistemological question is crucial to understand the true genesis of the technical object for Simondon and to do this, he directs his attention to the event of invention itself, which I will elaborate in what follows.

Among the several engines and assemblages that Simondon employs as examples to lay out his arguments concerning the genesis of the technical object, of particular significance in the case of invention is the Guimbal Turbine, in a hydro-electric powerplant (1958/2017, p. 57). In what follows, I shall draw on this example to illustrate how scientific models are insufficient to account for the genesis of technical objects and identify the additional resources that the inventor-engineer needs to bring into the technical activity during invention.

From my discussion earlier on the role of models in knowledge production, it would be the case that the knowledge of the phenomenon, of potential energy in water being able to be converted into electrical energy (from say, Faraday's models of electromagnetic induction), precedes the invention of the hydro-electric generator. However, it is not simply a direct application of this (past) knowledge that produces the Guimbal Turbine. The model despite being constructed with the epistemic purpose of building a future hydro-electric generator, can have a determinate influence only in so far as it can help engineers make use of the understanding of the process of generating electricity from water (or perhaps creating a highly simplified prototype in the laboratory). In practice however, the internal electric parts would need to be constantly kept dry; the generator would have to be frequently turned off to dissipate the heat; and the friction within the turbine needs to be minimized to not just make it more efficient but also to ensure that the turbine does not deteriorate over time.

In the Guimbal turbine as a concrete technical object, however, there are two main elements, water and oil, each having their own multi-functional potentials. The water brings with it 1) a potential energy to activate the turbine, and 2) it helps dissipate heat away from the generator. The oil, on the other hand, can 1) lubricate the generator, 2) insulate the turbine, and 3) prevent water from leaking because of differences in pressure of the oil inside the box, and pressure of the water outside. Although the knowledge of these individual potentials precedes the invention of the turbine, it is at the moment of invention, that they are coupled into a concrete synergetic system that is self-maintaining. A coupling where the different functional capacities of both water and oil are interlinked to enable the reliable functioning of the generator, with water dissipating the heat from the generator while it simultaneously drives the turbine, and the oil insulating and lubricating the generator while simultaneously keeping the water from seeping into the crankcase (Simondon, 1958/2017, p. 57).

This intricate, synergetic coupling is one that did not exist (in the past) before invention, and does not follow directly from existing knowledge (as it would in a hylomorphic schema), and for this reason, Simondon refers to the emergence of this coupling as a "conditioning of the present by the future" (1958/2017, p. 60). Herein lies the active role that human cognition plays in the genesis of the technical object because, for Simondon, only a thought capable of foresight and creative imagination can accomplish this conditioning, in so far as it requires a capacity to identify, isolate, and organize existing elements as symbols in such a way that they represent a future technical ensemble that does not exist yet (1958/2017, p. 60).

How then does human cognition fulfill this role in the invention of the technical object? To be able to invent the machine that fulfills a specific purpose, the human inventor, for Simondon, needs to represent to herself, in thought, the way of functioning (such as electricity generation) that coincides with a technical operation (such as that between the water and oil in the Guimbal turbine) that accomplishes that function (1958/2017, p. 249). In making this representation, the theoretical resources, and the situatedness of the inventor-engineer in a particular cultural epistemic community, would definitely play a role but only in so far as they help serve as a repository of models that she can potentially make use of. For instance, one could argue that the creative ability of, say, the poet remains independent of the vocabulary and grammatical structures of the language she chooses. In the same way, the ability of the inventor-engineer to represent to herself a future technical ensemble retains an identity independent from the various epistemic resources she has access to.

Simondon further illustrates this role unique to human cognition through an analogical relation between the human and the machine, a relationship that doesn't lead to anthropomorphization of the machine (as one that is commonly found in AI discourses[6]) but one that points to a kind of techno-morphization of human thought — that during invention, there is a relation between the mental functioning of the human and the physical functioning of the machine (1958/2017, p. 151). In Simondon's words, "To invent is to make ones thought function as a machine might function, neither according to *causality*, which is too fragmentary, nor according to *finality*, which is too unitary, but according to the dynamism of lived functioning, grasped because it is produced, accompanied in its genesis" (Simondon, 1958/2017, p. 151). Therefore, it is only because the inventor-engineer can conceive of the not-yet-existing-machine's operation in her own thoughts, i.e., to organize the epistemic resources she has access to (say, the model of the hydro-electric generator, fluid dynamics of water and oil, etc) in ways that coincide with the practical end that the machine would be directed towards, was it possible for the machine to come into existence.

Having laid out what constitutes the activity of invention and having identified the role of human cognition during this activity, I will now turn to the preceding activity of the construction of the scientific model in the following section.

2.3 CONSTRUCTION OF THE SCIENTIFIC MODEL:

[6] See Salles et al. (2020); Watson (2019)

Drawing attention to the activity of invention in the genesis of the technical object, in the previous section, serves to overcome the hylomorphic schema in showing how technical objects are not arrived at by a direct application of scientific resources and in doing so, brings to surface the active role played by the human in this activity. Although Simondon's own project does not address it, drawing attention to the activity of the construction of the scientific model, like the activity of invention, also makes explicit the different modelling practices employed by the scientist-engineer, enabled by her own unique cognitive capacities. Such a perspective on scientific models, one that emphasizes the constructing of the model as opposed to the representational qualities that it is otherwise known by (as is the case in the semantic conception of the model outlined in 2.1), can be traced in Boon and Knuutuila's (2009) conception of models as epistemic tools. This alternative account of models recognizes their epistemic role by shifting attention from the way a scientific model 'is'. Instead, the conception of models as epistemic tools emphasizes how it is the activities of model construction and manipulation (as a continuation of construction) that equip scientists with useful knowledge (i.e., in their enabling of different modes of scientific reasoning, in helping develop theories, and in being instrumental in the later invention of technical objects). For instance, Boon and Knuutila (2009) draw on the case of the construction of Carnot's ideal heat engine to illustrate how at different points in the process of construction, the human-modeler actively makes choices such as what aspects of the real engine the model abstracts, what assumptions it makes, and what purposes[7] it is directed towards. These choices, which from hereon, I shall refer to as methodological choices, constitute the construction of the model.

Methodological choices would include i) the choice of the type of model to build depending on the purpose it would be directed towards; 2) which parts of the target system that the modeler takes into account and how she would abstract and idealize certain parts to be relevant for the modelling practice. 3) which existing scientific knowledge (other models and theoretical resources) to make use of in the construction of the model and 4) how the model is attached to the real world (through say, data

[7] Purposes in the case of Carnot's ideal heat engine would include the need for an explanation of a limit to the performance of heat engines. This purpose would direct different choices such as what parameters are taken into account, what existing theoretical resources to employ. (see Boon and Knuutuila, 2009, p. 18)

and measurements) (Boon, 2020a, pp. 15–17).[8] These methodological choices enable a justification of the model and allow us to approach models as intelligible entities given that we can reconstruct these choices, in so far as they are made by the scientist engineer.

Of significance in this activity of model construction is that the ways in which different methodological choices are handled is actively enabled by the human and her own unique cognitive resources. In the same way an activity of invention presupposes, for Simondon, an anticipation of a future technical object by the human-inventor, the human-modeler would have to anticipate future materials and new phenomena that can be productive towards the epistemic purposes she intends to satisfy. Similar to Simondon's arguments in support of the active role played by human cognitive resources in the invention of the technical object, Boon (2022) advocates the need for the same in the activity of model construction as well. She argues against the traditional belief that science is objective, and rational in the sense that 'the world speaks for itself', and brings to attention the otherwise trivialized contribution of human cognition in scientific practices (p. 124). Drawing on a Kantian epistemology, Boon makes the case for how the human cognitive apparatus possesses unique capacities such as the ability to identify relationships and analogies, and to imagine, conceptualize and transform observations into new concepts, all of which are essential for scientific research (p.114). Boon's insights further strengthen Simondon's argument that existing epistemic resources alone are insufficient to account for the genesis of technical objects. What Boon's work would add to Simondon's conception of the genesis is that it is not just within the intermediate space between the model and the functioning of the *future* technical object that human cognition plays its role, but also in the construction of the model itself. In the same way the creative capacities are directed towards the identification of a synergetic coupling between multiple pluri-functional elements (as is the case in the Guimbal turbine) during invention of technical object, the same capacities also find their significance in identifying relations and imagining new concepts that are productive in the construction of scientific models.

[8] Boon, (2020a) provides a systematic overview of these methodological choices articulated in a set of questions that guide modelers in the activity of construction and can enable other scientists to re-construct models to derive knowledge from them. These questions make up what is commonly referred to as the B&K method in PhilSci literature.

If these two ontogenetic events, the *construction* of the model and *invention* of the technical object, are both carried out by making active use of the unique cognitive capacities that the human brings into the technical activity, does that necessarily mean technical objects cannot exist at all without this role? What if there is an extremely useful technology that is offered to a human community by an alien species? If it is identical to other technical objects with the only difference being that it is one in which the human did not participate in the ontogenetic event, can we simply employ this alien-object in the way we employ other technical objects? From a hylomorphic perspective, there would indeed be no difference for the human between these two objects. There are, however, epistemological concerns that arise out of the use of technical objects that do not emerge from human activities of *construction* and *invention*. The next section brings these concerns to the surface in so far as making these concerns explicit will be later instrumental in the next chapter in distinguishing the genesis of traditional technical objects from that of machine learning models.

2.4 EPISTEMOLOGICAL IMPLICATIONS OF THE ROLE OF HUMAN COGNITION IN THE GENESIS OF THE TECHNICAL OBJECT.

Having established the role played by human cognition in the genesis of technical objects, I shall now draw insights from this to give an overview of the epistemological implications of this role. Particularly, I aim to address the question, how is our knowledge about the world (both the world that the technical object operates on, and the world within the technical object through which it operates), affected by the fact that human cognition plays a role in the construction of models and the invention of technical objects? In answering this question, I will first outline this idea of technical objects as bearers of knowledge made possible by the human reality embedded into them during invention. I will then illustrate how the process of invention itself (and that of construction as well) tells a story, one that can be made intelligible to those that later employ the technical object.

2.4.1 Technical Objects as Bearers of Knowledge:

Because the technical object is one that is thought and invented by the human, and also one that operates in the real world, it carries with it a mixture of both a human reality[9] and a natural reality (Simondon, 1958/2017, p. 251). Note that this mixture does not mean a compromise between the two realities, as it would be in the case of scientific models whose access to natural reality is mediated (or rather compromised) by existing epistemic resources (human reality). Unlike the model, the technical object operates in the real world, embedded into its chains of causes and effects. Simondon thereby refers to this mixture as one that is *stable*, in that it doesn't obscurely represent the natural world through human lenses, but that it rather "gives its human content a structure comparable to that of natural objects, and allows for the integration of this human reality into the world of natural causes and effects; the relation of man to nature, rather than being only lived and practiced obscurely [through say, the theory or the model], takes on a status of stability, of consistency, making it a reality that has laws and an ordered permanence" (Simondon, 1958/2017, p. 251). The technical object being grounded in the natural world, operating through its laws, is then a bearer of knowledge in that it can be studied inductively in the same way a scientist studies a natural phenomenon.

In PhilSci literature, Baird (2004, as cited in Russo, 2022, p.140) acknowledges this capacity of instruments to be bearers of knowledge in that, like theories, they can also provide explanations and predictions. To make his case, Baird draws on how Thomas Davenport was able to invent the electric motor without the knowledge of electromagnetic theory, and that this invention later served to be instrumental in the development of the theory itself (Russo, 2022, p.140). While contemporary PhilSci does acknowledge this capacity of technical instruments to bear knowledge and the active role that they play in the development of theories, it fails to make explicit the fact that it is the role of human cognition during invention which infuses the instrument with this knowledge-bearing capacity. If we do not take into account this human cognitive role, the technical object would be instrumental in knowledge production only in so far as how any other natural object would be. In the case of the

[9] Human reality here (and whenever it is later mentioned) refers to all the theoretical resources, the human constructed models, and the human thought during invention.

natural object, one arrives at knowledge, only after approaching the object as a phenomenon through the lenses of various epistemic resources. What distinguishes the technical object from the natural object is that its mode of operation coincides with the human thought that it emerges out of. This is what renders the technical object not just as an instrument in the 'process' of knowledge production, but it in itself as a bearer of knowledge, because it doesn't have to pass through the lenses of additional theoretical resources to be able to be understood[10]. This is what makes any invented technical object one that cannot be reduced to its practical ends, in that it retains an identity independent of it, one as a bearer of knowledge, signified by its operational schema materialized during the act of invention.

2.4.2 The Two Stories: *Construction* of the Model and *Invention* of the Technical Object.

In my discussion above, the idea there is a singularity between the mental functioning of the human and the physical functioning of the machine at the moment of invention, and that the technical object emerges as materialization of a human thought, implies that no matter how complex the internal schema of an invented technical object may seem from the outside, it can be rendered meaningful and intelligible through language. The genesis of the technical object therefore always tells a story, one of its invention, in so far as the human thought that materializes as the operation of the machine, during invention, is thought through linguistic elements.

The importance of being able to tell a story becomes all the more relevant when one takes into account its role in scientific practices. Boon (2020b) shows how it is necessary for scientific models to be able to tell a coherent story in order to be considered valid, in that the model being able to be captured entirely in linguistic elements would allow other scientists to reconstruct how the model is constructed. Boon further argues how non-linguistic elements such as diagrams, pictures, and graphs may indeed aid in telling this story, but the story remains one that can be told through language (p.14). Russo (2022, p.114-116) also holds the ability to tell a coherent story, of how the model is built and

[10] One nevertheless still needs access to some epistemic resources but only those that one would need to mentally reconstruct the act of invention.

tested, as a condition for its validity, in that it enables other scientists not only access to the model's construction, but also equips them with the ability to inquire and challenge the methods and results.

There are then two stories, one of the *construction* of the scientific model, and the other of *invention* of the technical object. In the same way the story about the model finds its value in enabling the scientific community to reliably reconstruct and adopt certain models into their own individual practices, the story about the invention of the technical object would find its value not only in 'users' that later employ it, but also other scientists that make use of it in their practices as 'instruments' in knowledge production. While the importance of being able to tell the story about the model is apparent and well-recognized, the same in the context of the invention of the technical object does not get the attention it deserves. Simondon traces this failure to be the source of the alienation between the technical object and the human, in that not being able to reconstruct the act of invention obscures the user from "knowing" the machine by its technical schemas and instead reduces it to its practical ends. To simply use and employ the machine, without being able to reconstruct the information of the act of invention that it carries with it, Simondon says, is like employing a "book that would be used as a wedge or pedestal" (Simondon, 1958/2017, p. 253). To anticipate my discussion in the next section, these two stories in the context of Machine Learning models, would be i) the story of the construction of the ML-architecture (e.g. why a particular neural network is built with the particular size and structure that it has); and ii) the training of the ML-model (e.g. why does the ML-model output the result it does, which is the object of explainable AI research).

In summary, this chapter, in investigating the genesis of the technical object brought to surface the epistemological importance of the role of human cognition in the acts of model *construction* and technical-object *invention*. The following chapter starts with a similar aim, that of investigating the ontogenetic event, albeit in the context of machine learning models. Throughout this investigation, I will identify different spaces where human participation in the genesis of the ML-object is distinguished from that in traditional technical objects, and bring to surface the epistemological implications of this difference for knowledge practices that make use of machine learning.

CHAPTER 3:

GENESIS OF THE MACHINE LEARNING OBJECT

3.1 INTRODUCTION

Having laid out what constitutes the genesis of technical objects and the epistemological importance of the role of human cognition during this genesis, I shall now turn to an investigation of the genesis of Machine Learning models. I shall pursue this investigation as informed by insights from PhilSci and Simondon's work that I have sketched in the previous section, and in doing so, I shall reconstruct the ways in which this genesis plays out and identify the possible epistemological implications of the same. More importantly, I'm interested in addressing the particular questions that my discussion in the previous section brought to surface, namely, what is the role of human cognition in the genesis of the ML model? Does the ML model carry with it a human reality in the way the traditional technical object does? Can the genesis of the ML model tell a story in the way the constructions of scientific models and inventions of technical objects do?

Before I layout my investigation, one important clarification is to be made: Is the ML model a *model* in the way a scientific model, as laid out in the previous section, is a *model*? Or is it instead a *concrete* technical object? Although the term, '*machine learning model'* is used across academic and professional settings, it is important to distinguish between those aspects of this entity (?) that fall under a model-ing practice and those that usually find their role in instrument-al practices. To avoid confusion, I shall use the two terms, 'ML-*Architecture'* and 'ML-*Object*' instead of the term '*model'*.

My position is that it is the ML-architecture that is the product of a modelling practice. It is a *model* in so far as it is constructed as a scientific model would be (I shall elaborate on this in the following sub-sections). The architecture does not refer to an already *trained*-model. Instead, it refers to a schema such as a *decision-tree* or a *neural network* which is later employed to build (or train) what I call the ML-object (which in computer science discourses is nevertheless referred to as a *model*). The

ML-object that is trained and optimized to a problem, substituting invention (or at least aspects of it), I will argue, is not a *model*, but one that is a concrete technical object, albeit with a questionable process of invention. The ML-object's identification as a *model* in academic discourses, I argue, is misguided: the ML-object may take the place *of* the model, in that it substitutes a task in a scientific practice which was previously the task of the model; the ML-object may also be a *part of* a modelling practice, like how any other scientific instrument is employed in the construction of models. The only exceptions are self-supervised architectures such as the variational auto-encoder (VAE) and Word2Vec, where the values of the hidden layers themselves, and not the predictions, serve as encoded representations of phenomena, identical to how scientific models isolate and simplify a complex phenomenon through their own epistemic lenses. My identification of the ML-architecture as a model and the trained ML-object as a technical object may seem misguided and is definitely not without limitations but I urge the reader to restrain from judgements before I make this clearer in the following sections.

3.2 THE *CONSTRUCTION* OF THE ML-ARCHITECTURE

In this section, I shall first compare ML-architectures to traditional scientific models and illustrate how information plays out as the phenomenon that is the object of their representation (3.2.1). I shall then draw a distinction between a pre-industrial ML practice and a post-industrial ML practice, to show how the former involved an active participation of the human in the construction of the architecture (3.2.2), whereas the latter downplayed this human role in order to privilege the empirical adequacy of the ML-objects subsequently built through the architecture (3.2.3). I will then conclude this section by drawing on my insights from the previous chapter to show the epistemological implications of the industrial turn in ML-research for (philosophy of) Science.

3.2.1 ML-Architectures as Information Transducers

ML architectures serve as the ground upon which future ML-objects are trained. Examples of ML-architectures include Decision trees, Support Vector Machines, and the different forms of convolutional, recurrent, and adversarial neural networks that reign dominance in the AI-world today. Architectures themselves do not have a predetermined end they are built towards, but certain presuppositions of the types of informational content they can extract from the data (e.g. sequential data, spatial data). This is similar to how scientific models provide certain understandings of a phenomena in question based on their own theoretical presuppositions. In order for my identification of the ML-architecture, in its construction and operation, as a scientific model (and as an outcome of a model-ing practice) to hold, I would first need to justify the target-system (qua phenomenon) that the architecture is directed towards.

Firstly, it is important to recognize that the conception of scientific models as epistemic tools overcomes the semantic notion that there is a direct representational relationship between the model and some real-world system. Instead, the modeler often employs various levels of abstractions and idealizations in building the model at the end of which the model may not necessarily share any representational relationship with the real-world system but can however be used in the production, manipulation, and understanding of that system. For instance, Euclidean geometry can be productive in building, manipulating, and understanding different geometrical spaces and objects but it is in no way tied to a particular set of real-world systems. The Euclidean distance formula would retain an identity independent of the contexts that it would later be applied in and is abstracted and idealized to a point where it withholds no content (i.e., reference to a concrete real-world system). Euclid himself may have however employed and experimented with different real-world systems in constructing the models of Euclidean geometry but the phenomenon that the construct-ed model would refer to remains one without a concrete content.

Like the Euclidean model, the phenomenon that is the object of the ML architecture is one without content, it is information in itself. Information here is understood in the way Simondon puts it, as a variability of forms (1958/2017, p. 150). It does not have content in that it is not form itself, but the variations that forms take, given a particular quantity of energy. Energy here is the carrier of information, like how audio communications in telephone networks are carried via electric pulses. The maximum energy and zero energy would mean there are no pulses whatsoever, which would render either as no sound, or a constant sound, but not a meaningful varying sound. For audio communication to take place via the electric wires, energy has to be regulated in precise ways, with calculated increases and decreases, each corresponding to a particular sound. At the same time, the overall maximum-threshold for energy has to be kept at its lowest to avoid noise in the medium, i.e., the quantity of energy needs to be minimized to a point that best serves the spectrum of human voices.

Information here is in these precise variations in energy, but it is not an absolute variability in that information distinguishes itself from pure randomness. In Simondon's words "Information is thus halfway between pure chance and absolute regularity" (1958/2017, p. 150). To illustrate further, take the example of the 20th Century CRT TV. The TV cable gets its identity as a carrier of information not just because it can carry one particular image or fixed set of images, but because it can carry all variations of images that fit its energy-quantity (resolution and color-spectrum). And yet, the TV cable stops being a carrier of information if it projects purely random images, as is the case of the static white noise when no signal passes through the TV cable.

In the same way, ML-architectures are modelled in a way to carry certain types of information (a particular energy-quantity spectrum), like how the TV works within its limits of resolution and color-spectrum, but it is not one particular form or a set of forms that it operates on. At the same time, every ML architecture has some mode of regulation that makes its operations distinguish itself from a purely random operation (through say, the loss-function, and back-propagation[11]). What then distinguishes the TV-signal from the ML-architecture is that the former is optimized to carry these variations of forms (energetic-pulses) with the highest fidelity, whereas the ML-architecture is designed to transduce a set of variations of forms in a particular energy-quantity into other variations of another energy-quantity in a precise manner.

---

[11] For an overview of backpropogation and loss function as regulatory methods in neural networks, See: https://towardsdatascience.com/how-does-back-propagation-work-in-neural-networks-with-worked-example-bc59dfb97f48

To better understand this operation, take the case of a convolutional neural network (CNN) that is designed to classify a binary image of a hand-written digit into its digital equivalent. The model would take the 64 pixels x 64 pixels image as input and output a digit between 0 and 9. We can identify how there is a compression irrespective of whether the model outputs the correct label, in that the 4096 bits [64x64] are transduced into a mere 10 bits.[12] This compression, or rather transduction, is the condition by which the ML architecture operates and it does so from its genesis, albeit in purely random ways, leading to purely random results. The regulation, however, operationalized by say, the loss-function, makes changes to the channel's internal nodes (the neurons) in such a way that the compression from 4096 to 10 bits takes place in a way that results in the correct label. Choices of the modes in which this compression and regulation plays out are not predetermined but one that the human-scientist actively makes. In what follows, I shall sketch the role of the scientist-engineer in making these choices in the construction of the ML architecture.

3.2.2 The Role of the scientist-engineer in the construction of the ML-architecture

From the previous section, it is evident that the construction of different ML architectures is driven towards carrying certain types of information (energy-quantity) and regulating this information in particular ways. This is analogous to how scientific models are driven towards regulating certain target systems as phenomena, and how this follows some form of regulation (abstractions and idealizations) by the epistemic resources the modeler employs, resulting in a meaningful representation of the target-system.

The construction of the ML-architecture, as in the case of the construction of the scientific model, relies on the active role that human creative capacities bring into play. Constructing the convolutional neural network, for instance, is made possible only by the human ability to conceptualize informational patterns in visual media as those that can be derived from abstracting

[12] Note that bits here are employed only to illustrate the process of transduction. In practice, neither a pixel nor a digit corresponds to a single bit.

different portions of an image at different levels (LeCun et al., 1989). Similarly, the construction of the recurrent neural network presupposes the idea that semantic information in textual data can be better derived sequentially than spatially (Jordan, 1997). As is the case in the scientific model, this construction does draw on existing theoretical resources as well. For instance, Information Bottleneck Theory (IBT) (Saxe et al., 2019), an epistemic resource commonly employed by computer scientists, advocates the idea that high-level patterns and generalizations are made possible by changes in the number of neurons in hidden layers. If the input layer in a neural network has over 4096 (64x64) neurons corresponding to a 64 pixels x 64 pixels image, then there being only 128 neurons in a certain hidden layer would mean that all the information in those 4096 neurons has to somehow pass through a tight 128 neuron layer. IBT advocates that this informational compression within hidden layers is what allows high-level patterns and generalizations to be made possible within the neural network.

Computer scientists would thereby make use of IBT in making choices (such as the overall depth of the network and sizes of individual layers) while building their ML architectures. Because the various presuppositions of information and choices of the shape and size of the network are actively made by the scientist, the construction of the ML-architecture does tell a story. And it is a story that can be meaningfully interpreted by other scientists, making it possible to operationalize this architecture in their own practices, recreate it, make changes to it, or perhaps even question its validity by building an alternate architecture that contradicts the presuppositions and choices made.

This practice that I have sketched above, with the scientist-engineer constructing the architecture with her own conscious decisions and choices, guided by theoretical resources, and being able to render this construction as a story, is a practice unique to the early decades of ML research and in recent years, as I will argue in what follows, has failed to sustain itself. Drawing on Sevilla et al. (2022) who identify a simultaneous exponential increase in both computational power and number of publications in ML research after 2010, I hereby make a distinction between a pre-industrial turn and a post-industrial turn in ML research demarcated by the year 2010.

3.2.3 The Industrial turn in ML Research

The years leading to 2010 saw great leaps in computational power which made it possible for the first time to implement architectures such as CNNs (LeCun et al., 1989), RNNs (Jordan, 1997) and LSTMs (Hochreiter & Schmidhuber, 1997) in industrial settings. In addition, what makes this particular year of significance for me is a competition called the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[13], a competition[14] that invited researchers to develop an ML-model with the highest accuracy for recognizing a thousand different types of objects from natural images. The aftermath of this challenge saw a wide-spread societal recognition of machine learning's abilities and invited investments by corporations towards its industrial applications, which until then was mostly confined to academic fields. This challenge, which continued every year ever since, marks the first time that the importance of empirical achievements (the results) were prioritized over the process that leads to the construction of the architecture at the research community level.

Like the ImageNet challenge, following years have seen a flurry of popularized benchmarks and challenges such as the COCO[15], TIMIT[16], SQuAD[17], BioASQ[18], and several others. Most, if not all, research papers in machine learning which present new architectures or modifications to existing ones, have an entire section dedicated to their architecture's performance on these benchmarks. This is further fueled by the idea that one needs to be able to beat the state-of-the-art model in these benchmarks to be able to publish their work in popular journals such as the IEEE (Yuan, 2020).

The post-industrial turn in ML research thereby brought with it a series of architectures that privileged empirical results at the cost of a coherent and justified story of construction. Lipton & Steinhardt (2018), for instance, in their literature review of trends in contemporary ML research, note

[13] https://www.image-net.org/challenges/LSVRC/

[14] What distinguishes this competition from preceding ones in the field of AI research such as the RoboCup and Loebner Prize is that ILSVRC is particularly directed at solving an epistemological task. It presupposes some form of knowledge embedded into the ML-object that can help it identify a wide range of images in the way a human would.

[15] https://cocodataset.org/

[16] https://catalog.ldc.upenn.edu/LDC93s1

[17] https://nlp.stanford.edu/blog/cs224n-competition-on-the-stanford-question-answering-dataset-with-codalab/

[18] http://bioasq.org/

how researchers, despite having achieved significant breakthroughs in various computational tasks, were often unable to identify and explain which factors are responsible for empirical gains. Whatever empirical successes they did present, it remained unclear whether it was the architecture that was responsible, or if it was the choice of hyperparameters, or instead the distribution of the dataset itself.

To complicate this further, not being able to provide reasons for why they built the architecture in the way they did, did not stop researchers from ‘making up’ a story instead. Lipton & Steinhardt (2018, p.2) note how research papers on novel architectures offer speculations in the guise of explanations, which are later interpreted as authoritative given the presumed expertise of the authors and affiliations to popular tech companies. This failure to identify sources for empirical gains and increasing presence of ill-understood architectures (often referred to as the model zoo) have raised concerns regarding a lack of rigor within the field and the perception of Machine Learning as a form of alchemy (Hutson, 2018)

3.2.4 Epistemological implications of the Industrial Turn in Machine Learning Research for Philosophy of Science.

The practice-turn in (philosophy of) science, as sketched in the introduction, brought with it an emphasis on practical ends, but it nevertheless retained a norm of truth in its commitment to the use of methodological and epistemological criteria, as evidenced by norms such as the validity of the model being derived from a coherent story of construction. However, the post-industrial turn in ML research, as laid out above, in its utter disregard to process of architecture construction, marks the practice-turn in science confronting a limit, with the norm of utility being so excessively privileged that they render previous methodological and epistemological criteria unnecessary to validate the model.

Furthermore, architectures such as the *Transformer* (Vaswani et al., 2017), native to the popularized GPT model, render existing theoretical resources such as the IBT that I have laid out above irrelevant for the model’s performance, because changes in the architecture become insignificant after one pours in an exceedingly large amount of computational power and massive datasets (Malik, 2021). Not only is this problematic because it is detrimental to the democratization of ML research, with only those who have access to these resources at these scales being able to deploy ML architectures, but also because it de-incentivizes research into other architectures and de-values existing novel architectures that were products of human ingenuity. For instance, before the advent of the power- and data-hungry *Transformer*, the LSTM architecture was commonly used in language tasks such as machine translation.

Developed in 1997, and only implemented industrially in mid-2010s, LSTMs owe their existence to human creative potential. Not only did they solve the then technical problem of the vanishing gradient[19] which made long-term semantic dependencies practically impossible, they did so using an intricate human design. By introducing specialized memory cells, the researchers, Hochreiter and Schmidhuber, crafted a mechanism that could retain and forget information over extended sequences, akin to the workings of human memory (Hochreiter & Schmidhuber, 1997). This conceptualization of LSTMs showcases the brilliance of human cognitive abilities in drawing inspiration from natural systems, as well as the relentless pursuit of innovative solutions to intricate challenges.

In contrast, the *Transformer*, which by now has effectively replaced the LSTM in all language tasks, despite its slightly higher empirical gains, remains a rudimentary architecture, with its value primarily traced in its ability to employ large datasets and make use of a greater computational power. With the innovativeness of the architecture aside, what makes the LSTM significant is that the question, "how is the architecture able to solve language tasks?" would have an answer, one traced to its construction by its human developers. The *Transformer* on the other hand, despite several misguided speculations in research communities, does not provide a justified answer (M. Sullivan, 2023).

Lastly, to conclude my discussion in this section, it is important to reiterate that the technical object is both one that serves a practical end, and one that bears knowledge in its materialization of a human thought. This knowledge, in the context of the abstract technical object that is the scientific model, is

[19] See Hochreiter (1998) for an overview of the vanishing gradient problem.

what makes the model valid. The post-industrial turn with its exclusive privileging of empirical results, as is evidenced by architectures such as the Transformer, effectively reduces the technical object to its instrumental identity. In the context of information and the process of transduction, Simondon highlights how it is in fact very easy to build machines that accumulate and transmit large amounts of energy, and that human ingenuity lies instead in building machines that efficiently transduce this energy (1958/2017, p. 156). A research community that holds an application such as ChatGPT and its underlying power- and data-hungry (energy-hungry) *Transformer* architecture as the exemplar of its achievements, would in effect be identical to one that privileges the use of a heavier stone with a greater force instead of a human-invented hammer.

3.3 THE FORMATION OF THE ML-OBJECT

The ML-object distinguishes itself from the ML-architecture, in the same way an atomic power plant would distinguish itself from the nuclear fission model. Like the scientific model which enables certain operations and constrains others, the ML-architecture too enables the capturing certain types of informational patterns and constrains others. The ML-object is one that is grounded on the architecture and yet is not fully determined by it. The architecture itself cannot be used towards a practical end; it has no finality (result). The scientist-engineer chooses an architecture and optimizes it to a particular task, by say, adjusting its parameters, and training it on a dataset. The genesis of the ML-object is then the optimization event itself, that by which an instance of the architecture becomes a fully-trained 'model' ready to be operationalized in instrumental contexts.

This section investigates the optimization event that makes the ML-object what it is. In particular, I seek to answer the question, what can the genesis of the ML-object through the optimization event, and the role played by the human in this event, tell us about the reliability of the knowledge we can produce using the ML-object? Answering this question unfolds as the sub-questions: How is the human reality embedded in the 'formation' of the ML-objects different from the human reality embedded into 'invented' technical object? How is the story of the formation of the

ML-object different from the story told by the genesis of technical objects and the construction of (some) ML-architectures? Note that I use the term *formation* here to distinguish this ontogenetic event from that of *invention*, and the particular differences between the two will be made clearer in what follows.

To answer the above outlined questions, I shall first, in 3.3.1, locate the context/community in which the genesis of the ML-object takes place. This will help illustrate the type of knowledge practice the ML-object is substituting or being instrumentalized in. Drawing on the role of human cognition in the invention of the technical object, as laid out in 2.3, I will illustrate, how the scientist-engineer does not participate in the genesis of the ML-object but nevertheless creates the conditions necessary for the genesis. In doing so, I will show how the scientist-engineer then succumbs to a hylomorphic schema, and consequently becomes a worker instead. The outcome of my investigation into the genesis of the ML-object will be laid out in 3.3.2 where I show how the lack of human participation entails the ML-object as not having a story (and therefore not bearing knowledge) in the way technical objects do. I then, in 3.3.3, trace how approaches in explainable AI nevertheless try to project a story onto the ML-object, even when it doesn't have one, and I will then show why this is problematic. Lastly, I will conclude this section by pointing out how certain interpretable architectures can help reconstruct a story that is faithful to the formation of the ML-object, even if the human does not participate in this formation.

3.3.1 The Context of Formation of the ML-Object

Before I investigate the role played by the human in the genesis of the ML-object, I shall first address the contexts where this genesis takes place. Although construction of ML-architectures is situated in computer science communities, the formation of the ML-object rarely happens within it[20]. For instance, Meteorology would make use of ML-architectures from computer science communities to build weather forecasting networks, but these networks are trained and operationalized within the

[20] The only cases where the ML-object itself is operationalized within the CS community would be those where its results on say, certain benchmarks, serve as justifications for the architecture's validity.

practice of Meteorology, in so far as the newly-formed ML-object substitutes existing weather models such as those based on atmospheric physics.

It could be the case that meteorologists outsource this task to an ML-engineer, but the ML-object being situated in the knowledge practice of weather prediction would in fact suggest the ML-engineer temporarily becoming the meteorologist, as opposed to, say, meteorology becoming a sub-practice of computer science. In this example, it is also important to note that although the ML-object substitutes the physics-based model, it cannot claim the label 'model' because i) the ML-object is not one that is constructed in the way scientific models are (this will be elaborated further), ii) it in itself does not directly draw on existing theoretical resources from meteorology, except indirectly through surface level choices made by the meteorologist-scientist like identifying relevant variables and curating the dataset, and iii) scientific models are able to provide predictions because they are able to represent and explain weather phenomena through their own lenses, as opposed to ML-objects that offer predictions of weather, without being able to say anything about the phenomena as such.

Like the weather-predicting ML-object in Meteorology, we find similar instances in other epistemic communities, with earthquake-detecting neural networks showing the potential to substitute traditional geophysical models in Seismology, and CNNs being increasingly used in the place of traditional medical diagnostic models in medical communities. In all these epistemic communities, the process of *construction,* by which their traditional models came into to being, were those where the human would play an active role, as was described in section 2.3. However, this process, in the context of the ML object, only begins at the moment when the training starts, and ends when sufficient accuracy has been achieved to stop the training. The *construction* here, if it were to be called so, happens within closed doors with no participation from the human.

3.3.2 The Scientist as a Worker in the Formation of the ML-Object

The *training* of the ML-object happening within closed-doors does not mean there is no human reality in this construction. Although the human does not participate in the process of construction itself, she does play an active role in enabling it: It is the human that defines the problem as that which can be handled by the machine, and identifies which ML-architecture is most suitable to handle that problem[21]; it is the human that curates the data that the network is trained on; it is the human that identifies which features (variables) are most relevant for the problem; and lastly, it is the human that evaluates the network and tunes its hyperparameters if its performance is not satisfactory[22]. All these activities in the construction of the ML-object are also reliant on the human's own creative abilities. Is all this not enough to say that the ML-object is constructed by the human?

No! Simondon would argue that the human-scientist here is simply creating the conditions necessary for the genesis of the ML-object and that the process of matter taking form, which in this case is the randomized-weights in individual neurons gradually changing to best fit the problem during training, remains not only out of the human's control, but is also obscured from her. To illustrate this further, here's a passage of Simondon's commentary on the worker who molds clay to create a certain artefact:

> "He [the worker] prepares the clay, makes it malleable, without lumps, without air bubbles, and correlatively prepares the mold; he materializes the form by making it into a wooden mold, and makes matter pliable, capable of receiving information; then, he puts the clay into the mold and presses it; but it is the system constituted by the mold and the pressed clay that is the condition of the process of taking form; it is the clay that takes form according to the mold, not the worker who gives it its form. The working man prepares the mediation, but he doesn't fulfill it; it is the mediation that fulfills itself on its own once the conditions have been created; even though man is very close to this operation, he does not know it" (Simondon, 1958/2017, p. 249)

The human-scientist in preparing the conditions necessary for the ML-object's genesis then, Simondon would say, stops being a scientist and instead takes on the role of a worker, in that she remains alienated

---

[21] Also, note that although the human-scientist has agency in identifying and choosing the right architecture, this choice may only be justified for architectures such as LSTMs and CNNs, as discussed in 3.2.2 , whose construction itself tells a story. A scientist making use of a post-industrial turn architecture such as the *transformer*, which does not carry with it a coherent story of construction, would be similar to a 17th century alchemist.

[22] See Kumar (2015) for an overview of the steps involved in training a neural network.

from the ontogenetic event, namely the process of taking form itself. I do however acknowledge that an actual scientist may be provoked by this accusation. Surely an activity like this is common in scientific practices. Scientists often use complex instruments like electron microscopes and mass spectrometers in their practices and they do not really 'participate' in the internal process through which these instruments do what they do. They calibrate the instrument if it does not work, or if they have reason to believe that it isn't working as well as it should, but they only make use of its outputs and readings, right? How is this any different from using the ML-object as an instrument in the scientific process? Doesn't this render all scientists as workers?

What distinguishes the ML-object from other scientific instruments like an electron microscope is that the latter, in being a technical object which is *invented*, carries with it the knowledge of the human thought that materializes it in the ontogenetic event (as was discussed in 2.3). Scientists that make use of these *invented* instruments can and most often times are able to, reconstruct the same operation using their own cognitive resources. They know how and why the instrument works the way it works and being able to do this is of utmost importance in scientific practices because only then can the scientist be aware of the instrument's own limits and capacities. A scientist continuing to employ the instrument without this knowledge, in the practice of producing new knowledge, is then in effect not a scientist but a worker. The ML-object however, in not being one that is a product of an invention, but instead a self-fulfilling formation, made possible by the human-created conditions denies, the scientist-worker knowledge of its ontogenetic event. This also marks the looming of a hylomorphic schema into the scientific practice itself. In the same way the worker only knows what goes into the clay-mold and what comes out, the scientist-worker only knows what goes into the machine (the conditions) and what comes out (the ML-object) but not what happens within.

Now, if the ML-object is one that is not *invented*, and thereby one that cannot tell a story as an *invention* would, does it make it impossible for us to nevertheless derive this story? Can we not, after the ontogenetic event, reconstruct the story assuming it *was* a product of *invention*? As frivolous as this question may seem, this is the approach that contemporary post-hoc explainable AI (XAI) approaches take. In what follows, I shall investigate the epistemological assumptions made by these approaches and in doing so address their shortcomings and identify spaces in which they may continue to be valid.

3.3.3 Deriving a Story in the Absence of One

With the aim of maintaining reliability and trustworthiness while simultaneously employing complex black-boxed ML architectures, recent years have seen the advent of post-hoc explanatory methods, where a second model is created to explain the first model. This secondary model, unlike the black-boxed ML-object is one that is constructed by a human and thereby does offer an intelligible story as explanation. However, post-hoc explainable methods operate on the assumption that this secondary model is identical to the ML-object. Sullivan (2022) for instance advocates how one can arrive at an understanding of the inner-workings of the ML model by resolving what she calls, link uncertainty. The resolution of link uncertainty would entail establishing a link between the scientific community's understanding of the target system and the ML-object's predictions. For Sullivan, resolving link uncertainty can help understand the ML model because explanations derived from the human scientific activity can serve as explanations for how the model ends up making predictions. However, establishing this link rests on the assumption that there being an equivalence of the same input features and the same predictions between the community's human-constructed model and the machine learning model is sufficient to conclude that they both employ the same mode of reasoning in their operation.

Such an absolute fidelity between the human conceptual model and the ML-object is not only unattainable, as is evidenced by Rudin (2019), who in her work shows how it is impossible to reconstruct calculations made within the ML-object in human-constructed models, but is also self-contradictory, in that if there was indeed a conceptual model that is identical in its operation to the blackboxed ML-object, then one wouldn't have to use the ML-object in the first place and could just make use of the conceptual model instead. Furthermore, a long-held belief that justified the use of XAI methods is that even if it is black-boxed, the ML-object does extract certain high level patterns from the data (generalizations) and that we can estimate what these patterns by methods such as Saliency

Maps[23] and Variable Importance[24]. However, Zhang et al., (2021) question the generalizing ability of neural networks by identifying how complex neural networks are able to achieve high levels of accuracy despite intentional errors in data. For instance, in a dataset of images of cauliflower and broccoli, there being a thousand different variations of the cauliflower image, but all these images having the same label say "*Cauliflower",* meant that if the ML-object is performing with a high accuracy, then it must have been able to extract some real-world pattern about what a cauliflower is, despite differences between individual cauliflower-images. However, what Zhang et al.'s (2021) analysis shows is that the ML-object continues to perform with a high level of accuracy even if the images were mis-labelled. What this means is that if in the same dataset, the labels "cauliflower" and "broccoli" were assigned randomly to images of cauliflowers and broccolis, a model with sufficient training data and computational power would still be able to perform the classificatory task with the highest accuracy. An insight from Zhang et al.'s findings is that the ML-object does not necessarily find high-level patterns (generalizations) from the data, and that it instead optimizes itself to fit each data-point. If the ML-object does indeed perform well on test-dataset, then it is testament to how its own training data is so astonishingly exhaustive that any singular instance from the test-dataset would be almost identical to at least one of the instances in the training data. Building a secondary model to explain the black-boxed ML-object is then problematic in so far as the generalizations made by the human-constructed model may in no way align with the mode of reasoning employed by the ML-object

Continuing to use these secondary models to understand the ML-object, would still be acceptable as long as one doesn't consider their insights as 'explanations' but instead as 'speculations' or as how Rudin (2019) puts it, 'approximations' of the ML-object's modes of operation. Still, Sullivan's conception of link-uncertainty does find its value, if not in enabling an understanding of the ML-object itself, at least in enabling the scientists to make use of its functioning, as a guide to construct their own models of the target-system.

[23] Saliency maps can help determine which portions of an image are ignored and which portions are paid attention to in the model but they fail to address how the relevant information in an image is being used.
[24] Variable Importance is used to identify which features (variables) are being used to make a prediction.

3.3.4 Reconstructing a Faithful Story through Interpretable ML-architectures

So, if there is no intelligible story of the genesis of the ML-object and one cannot faithfully project a story onto it, does that render the ML-object fundamentally incapable of bearing knowledge in the way technical objects do?

While this is more often than not the case with black-boxed architectures such as neural networks, there do exist fundamentally interpretable architectures which carry the marks of their genesis in their existence, and one can reliably reconstruct a story from studying their internal schemas. These include Decision Trees, Gradient Boosting Machines, and regression based models (Molnar, 2020). ML-objects derived from these architectures stand as testaments to the knowledge bearing capacities of technical objects, in that one can legitimately study their internal schemas to gain knowledge about the phenomena they are directed towards.

If Architectures such as these truly enable the scientist-engineer access to their inner-workings, why aren't they employed as much? The answer ML-engineers usually give is that there is a tradeoff between *accuracy* and *interpretability* (Bratko, 1997), and that in certain high-stakes contexts, a greater accuracy is to be prioritized to ensure the best possible outcomes, even if that means the model is way less interpretable and cannot be studied.

This trade-off between *accuracy* and *interpretability* has nevertheless been challenged in recent years with researchers illustrating how this phenomenon does not exclusively arise from the architecture itself and that if the data is structured sufficiently, with a meaningful representation of features, there would be no significant performance difference between more complex neural networks and simpler interpretative architectures (Dziugaite et al., 2020; Rudin, 2019).

Despite these challenges, a continued widespread emphasis on the use of black-boxed neural networks in contemporary ML research also points to a more sinister influence. Given the shift from academic to industrial settings that the post-industrial turn ML research has taken (as I've shown in 3.2.3), Rudin (2019) also notes how there has been an incentive for researchers with industrial affiliations to privilege black-boxed models because companies can make better profits from intellectual property afforded to a black box (p. 6). Such a sinister practice has reached even greater frontiers with companies like OpenAI, not only using fundamentally black-boxed models, but also refusing to share the data they are trained on (Greshgorn, 2020). What this implies is that very little story that these ML-objects do tell, at least about their conditions of genesis if it isn't for the process itself, is also obscured. This is not only problematic because it does not allow us to validate the ML-object's results, but it raises concerns of a possible epistemic violence, with these models relying heavily on the knowledge contributions made by different epistemic communities and yet not acknowledging them. The next chapter, on genealogy, will further elaborate these modes of epistemic violence, that come into play when ML-objects interact with other knowledge practices.

PART II

GENEALOGY

CHAPTER 4:
GENEALOGY OF TECHNICAL OBJECTS

4.1 INTRODUCTION

An ontogenetic study of technical objects cannot be complete without accounting for their genealogical dimension. The type of ontogenesis sketched out and investigated in the context of Machine Learning in Chapters 2 & 3 is that of the ontogenetic event, the coming into existence of a technical object. However, ontogenesis as an approach to ontology, transcends a singular event. In the same way the genesis of the organism follows not just the "birth" as an absolute event, but a long lineage of genetic events that eventually lead to the organism as a part of a species, the genesis of the technical object too, carries with it a genealogy of those that came before it, and those whose future geneses the object will be operationalized in.

Such an "evolution" of the technical object is made possible because although the technical object is invented by the human, as a materialization of her thought, it can be detached from the inventor and the context of *invention*. Simondon says "the machine has a sort of impersonality which allows it to become an instrument for another man; the human reality that it crystallizes within itself is alienable, precisely because it is detachable" (1958/2017, p. 250). This detachment from the original context of invention and a re-attachment in completely foreign contexts means that another human, despite not having invented the technical object herself, is nevertheless able to re-construct its *invention*, provided she is able to re-think the same thought as that of the inventor at the moment of *invention*.

It is only because the technical object bears the knowledge (the story of *invention*) of its operational schema, and only because this knowledge can be re-constructed by another human, was it possible for, say, the spring engine, which was initially used in the 15th century mechanical clock, to detach itself from the clock and be made use of in inventing the 16th century firearm. The impersonality of the technical object (its detachable and re-attachable character), for Simondon, establishes what he calls "technicity" as an inter-individual relationship. Technicity here is a mental and practical universe in which human beings communicate through what they invent (1958/2017, p. 252).

The genealogy of the technical object would then follow a series of communications. These communications are not just sequential with the technical object being invented, and re-invented in a linear path as it passes through different communities; they can also be complex and irregular, with multiple technical objects being grouped in a single activity of invention, as in the case of the first automobiles which trace their genealogy to both the combustion engine and the bicycle. What I would further make explicit in this conception of genealogy that Simondon does not is the active role played by the scientific model alongside invention; in that between every *invention* and re-*invention* is also a construction of the scientific model.

To explain the genesis of the technical object, I have previously approached *construction* and *invention* as individual activities with the former leading to the latter. Taking a genealogical lens however brings to surface how they are interconnected: with each scientific practice, relying on technical instruments (that have previously been invented) to construct scientific models, which are later made use of during the *invention* of new technical objects, which further find their epistemological role in other scientific communities in being instrumentalized in the *construction* of new scientific models.

Taking into account the role of model construction as a part of the genealogical study further emphasizes the importance of the epistemic identity that the technical object embodies in addition to its practical identity; that with each new community it enters into, it makes possible the production of new knowledge within that community by enabling the construction of new scientific models that were previously impossible.

Investigating this epistemological role of the technical object through a genealogical lens is the aim of this chapter. In particular, I seek to answer the question, How is it that scientific communities are able to extend their knowledge practices by making use of technical objects that emerged from communities foreign to them? In other words, if technicity is indeed this inter-human communicatory relationship, does that mean all technical objects can establish this communication in the same way between all epistemic communities[25]? Can some technical objects integrate with certain epistemic communities better than others? Answering these questions in this chapter will later be instrumental in investigating (in chapter 5) whether and how the use of Machine learning can help scientific communities extend their knowledge practices.

To answer the above questions, I shall first, in section 4.2, outline how technical objects are able to establish communicatory relationships (or what I will call technical bridges) between epistemic communities. I will then, in section 4.3, identify the other side, where such technical bridges cannot be established, and illustrate how this inability renders certain technical objects as detrimental to (or rather violent towards) other epistemic communities.

4.2 TECHNICAL BRIDGES AND CONTINUOUS MODES OF PROGRESS

To speak of technical objects as moving between sciences, one needs to first acknowledge that the two scientific communities (one that it emerges out of, and the other that it is later made use of in) between which the technical object finds itself in are not identical and are in fact heterogenous[26]. Subsequently, to speak of heterogenous communities within science, one needs to first acknowledge that the scientific project itself is not a product of a single genealogy of theoretical and empirical developments as the

[25] Note that from here onwards, I use the term epistemic communities in place of scientific communities. This is to expand my investigation to not just particular communities that are labelled "scientific" in the traditional sense but also other communities that engage with the production of knowledge. This includes indigenous communities like, say, the Inuit community in the arctic regions, who over a millennia developed their own understanding of the environment, wildlife, and climate patterns.

[26] In that it is not necessarily a technical bridge (a communicatory relationship) between communities if say, one is looking at the emergence of the electric motor from the electromagnet, all happening within the field of electrical engineering.

logical positivists would characterize it as, but that it is instead a heterogenous ensemble of epistemic communities with no single fixed foundation. Such is the conception of science that developments in PhilSci have paved the way to in the aftermath of the logical positivist movement: starting from this idea that scientific knowledge is unstable and always contestable (Kuhn, 1962; Lakatos, 1976) to how there are plurality of scientific perspectives each building models of the same phenomena in ways that are inconsistent with others (Massimi, 2022).

This new post-logical positivist conception of science characterized by the existence of a plurality of scientific perspectives, each of which enables and guides the construction of scientific models from its own repository of theoretical and methodological resources (Boon, 2020b), entails that scientific knowledge claims can be held valid only in relation to a given perspective. Such a polarization of epistemic communities and the knowledge claims they produce may give the illusion that the scientific enterprise is fragmented and disconnected, with no inherent unity or interdependence between areas of research.

Here-in lies the particular space that the notion of technical bridges finds its significance, in that technical objects, by being able to move through disciplinary boundaries, establish relationships between sciences —relationships that are not theoretical or methodological, but instead technical, with, in Simondon's words, "each science being capable of making use of a certain number of other sciences for its own benefit, which it uses as technical sources in order to carry out the effect it studies" (1958/2017, p. 125). Therefore, while scientific knowledge claims are only valid within well-defined scientific perspectives enacted by specific epistemic communities, technical objects do not necessarily emerge out of the same epistemic community where they are later used in to advance scientific knowledge. It is in fact common to find technical objects emerging out of models from one scientific discipline being used in another foreign discipline.

For instance, advancements in quantum physics building on top of Einstein's concept of stimulated emission led to the construction of photonic lasers(Yam, 2004). Decades later, the technology was later instrumental in various applications in biology. In particular, CRISPR-Cas9 employs lasers to create holes in cell membranes, enabling the insertion of genetic material directly into the cell (Yang et al., 2022). This technology is now being incorporated into the field of medicine to develop gene therapies that have the potential to treat a wide range of genetic diseases. In this example, photonic lasers, as technologies emerging from quantum physics, are now serving as technical resources in the field of biology towards the development of the CRISPR-Cas9 technique, which is in turn employed in the field of medicine.

Similarly, radio telescopes which were first developed in the field of astronomy to observe celestial phenomena were later adopted in the field of geology to create Deep Earth Imaging systems, enabling scientists to construct detailed images of the earth's interior. These systems are now adopted in Environmental Science for predicting earthquakes and volcanic activity, being instrumental in disaster planning and mitigating potential damage from seismic events.

The existence of these technical bridges between sciences suggests that the plurality of scientific disciplines is not a disjoint sent but one that is intersecting. Michela Massimi, in her Perspectival Realism (2022), refers to such an intersection to be indicative of a cross-perspectival assessment of knowledge claims, in that different scientific communities can justify the reliability of knowledge claims they generate through their dependence on material resources produced using knowledge from another community, which is in turn justified by another. Such a cross-perspectival assessment, which in itself is a mode of knowledge production, is possible because the technical object remains one that can be detached from the scientific model that it emerges out of. While scientific models are confined in their validity to the epistemic community they are situated in, technical objects that are built by making use of scientific models, in so far as they are invented, operate based on the laws of the real-world (as was discussed in 2.3). This is what justifies the use of a particular technical object by an epistemic community even if the community has a different understanding of certain phenomena from that of the community whose models the technical object was drawn from.

Now, for a technical bridge to be established, i.e., for an epistemic community to incorporate a technical object of a foreign origin into its own knowledge practice, two conditions have to be met:

Condition of Reconstructability: Firstly, it should be possible for the knowledge carried by the technical object (the story of invention) to be reconstructed by the recipient-scientist. This is because only when she know its capacities and limits, and how it does what it does, can the recipient-scientist be able to reliably and justifiably use the technical object qua instrument in her knowledge practice (for say, constructing new scientific models). For instance, take the case of a mass-spectrometer as a technical bridge between pharmacology and physics. Initially developed by the British physicist, J J Thomson, mass spectrometers (Dronsfield, 2010) are now being used by pharmacologists to study drug reactions, and aid the discovery of new drugs. To be able to reliably use the mass spectrometer in her practice, the pharmacologist-scientist should be able to reconstruct its invention to "know" the instrument first. She should be able to recognize its limits (of say, the resolution, and accuracy) and how it could be sensitive to different environmental conditions such as air quality, temperature and humidity.

Condition of Integrability : Secondly, it has to be the case that the technical object can be integrated alongside existing epistemic resources specific to the community; in that the use of the technical object should not invalidate the community's own knowledge practice. For instance, native American tribes relied on a slash and burn agricultural technique. This practice has sustained the natives for thousands of years, providing a varied, healthy diet with low environmental impact (Fraser, 2014). Adopting a technical object such as the plow in place of this technique, can only come at the cost of invalidating their practice, and all the knowledge that they have gathered over the years. In contrast, take the case of a biologist, employing a microscope instead of the naked eye. The microscope can be integrated alongside existing epistemic resources in that the biologist can continue her practice in the way she did before, with the microscope aiding and improving her own knowledge practice, without substituting it.

Now, the establishing of a technical bridge, when these two conditions are met, that is if the technical object is "known" by the recipient-scientist and can be integrated into their knowledge practice, constitutes a continuous mode of epistemic progress. I trace this mode of progress from reconstructing Simondon's commentary on 18th century technical improvements during which he notes how “when man, while preserving the fruit of his training, exchanges an old tool for a new tool whose manipulation is the same, he has the feeling of having more precise, skillful, and rapid gestures” (1958/2017, p. 130). Such an incorporation of the tool, which does not make one’s own training invalid, entails, for Simondon, a continuous progress because the tool-user perceives this progress. A scientist that exchanges an optical microscope for an electron microscope recognizes how her own practice can be enhanced, made easier, or open possibilities for applying her epistemic resources that previously did not exist given the limitations that the older optical microscope had.

Continuous modes of epistemic progress would then refer to the adoption of those technical objects that can establish a technical bridge between two communities, whose incorporation brings improvements to the recipient’s scientific practice, while allowing her to retain the theoretical and methodological resources specific to her epistemic community. Instances of continuous epistemic progress would include the invention of quartz clocks, which replaced mechanical clocks in many scientific applications due to their superior accuracy; and the invention of the digital pH meter which offered a more precise and quick method for measuring acidity and alkalinity compared to the traditional litmus paper.

It is important to note that not all technical objects can establish technical bridges between communities and lead to an epistemic progress within the community they interact with. Sometimes a technical object can be detrimental to one or several epistemic communities and I shall outline these cases in the following section.

4.3 AGGRESSIVE MODES OF PROGRESS AND TECHNICAL POWERBROKERS

In addition to continuous modes of epistemic progress, which entailed an integration of a foreign technical object into an epistemic community’s own knowledge practice, there are what I call, aggressive modes of epistemic progress, traced to those inventions whose introduction is characterized by a substitution of the knowledge practice altogether. I trace this mode of epistemic progress from Simondon’s commentary on 19th century inventions such as the automatic weaving loom, and the forging press which brought a sense of progress at the cost of invaliding the then practices of manual weavers and blacksmiths respectively (1958/2017, p.131). In building this notion of an aggressive mode of epistemic progress, I also draw inspiration from Gayathri Spivak's (1988, as cited in Dotson, 2011) concept of epistemic violence, which refers to the silencing of the marginalized communities through a removal of their ability to speak for themselves by invalidating their systems of knowledge, beliefs, traditions, and language.

It is important to note that aggressive epistemic progress in the context of scientific practices differs from that in industrial practices. For instance, development of new procedures and instruments in medicine and surgery brings changes that may seem aggressive in their rendering of previous techniques obsolete. However, I would argue that these changes still signify a continuous mode of epistemic progress in so far as they remain aligned with the theoretical and methodological resources specific to the discipline. Cases of aggressive epistemic progress in science would then include those inventions which served as catalysts for scientific revolutions or as justifications for privileging one scientific practice over another. For instance, the invention of the microscope lead to the birth of cell theory and microbiology which revolutionized our understanding of life at the microscopic level and made theories such as preformationism obsolete (Kaplan, 2019).

Unlike a continuous mode of epistemic progress, an aggressive mode cannot bring with it an establishment of a technical bridge. This is because what ensures an incorporation of a technical object within an epistemic practice despite its foreign origin is that it can meet the two conditions of reconstrutability and integrability. Aggressive modes of epistemic progress cannot form technical bridges between epistemic communities because their coming into being can only arise from a usurping of one knowledge practice by another.

If the material equivalent for continuous modes of epistemic progress is the establishment of technical bridges, then the equivalent for aggressive modes of epistemic progress is the enabling of what I would call, technical powerbrokers, in that they serve as justifications for substituting one knowledge practice for another or invalidating a knowledge practice of one community in the privileging of another. To illustrate what I mean by technical powerbrokers, I draw on Massimi's (2022) identification of two historical forms of epistemic injustice in scientific practices: i) *Epistemic Severing:* a surgical excision or removal of the contribution of particular communities from narratives about scientific knowledge production (p. 349) and ii) *Epistemic Trademarking*: a merchandising of scientific knowledge as a 'trademark' of one epistemic community at the expense of others who have historically contributed to such production (p. 362).

While Massimi outlines these two modes of epistemic injustice in the context of interactions between epistemic communities, I'm particularly interested in cases where they arise out of the development of technical objects or techniques. To illustrate what I mean by technical objects serving as powerbrokers, I shall now employ Massimi's two concepts of epistemic injustice through examples from colonial history in what follows in the two subsections below.

Epistemic Trademarking and Rubber as a technical powerbroker:

Indigenous tribes of South America have developed a knowledge practice over generations for the production of what they referred to as *caoutchouc*, which we now know by the name "rubber". They used it to make various objects such as bottles, containers, and collapsible-expandable hollow balls long before the advent of European colonizers (Domingues, 2020, p. 591). This practice constituted a situated knowledge about their environment and its natural resources with indigenous communities being able to identify which species had which properties and uses. They knew what the right conditions were for cultivating these plants, and when to harvest them (Domingues, 2020, p. 591). A French scientific expedition during the colonial era recognized of how useful rubber can be in industrial applications and in subsequent years, indigenous knowledge about its production was exploited in various late-19th century applications such as telegraph cables, bicycle tubes, and belts and bumpers in steam machinery (Domingues, 2020, p. 592). The increased demand in industrial applications and the rubber-producing plant species being native to the Amazon, lead to further exploitation of Brazilian Indians, who were forced to work in miserable conditions. The intensification of rubber production lead to subsequent decline of the indigenous population altogether and a disappearance of its epistemic role in the production of rubber (Domingues, 2020, p. 591). Indigenous knowledge, despite not having officially recognized at any time, continued to find its presence in later western scientific research on rubber (Domingues, 2020, p. 590). The technical knowledge around rubber (or rather, caoutchouc as it was originally referred to) production in the way we know it is then rendered as powerbroker, that has effectively been trademarked by western scientific communities, at the cost of invalidating the contributions made by indigenous populations in the development of these techniques.

Epistemic severing and the Compass as a technical powerbroker:

Long before the advent of the British colonization, aboriginal communities in Australia that did not have a written language practiced a novel technique of marine navigation based on song-lines that served as oral maps of the landscape (Norris & Harney, 2014). However, the introduction of the compass as a navigational instrument and its established reliability by the western nautical science community served to dismiss the legitimacy of this practice altogether with some historians such as Sharp (1964) even arguing that these techniques were too crude and primitive that it is only by accident and not by a navigational method that aboriginal communities were able to reach ocean islands (pp.7). After two centuries post-colonization, the indigenous practice has largely been forgotten owing to its dismissal and exclusion from the broader scientific community. It is only very recently that anthropologists were able rediscover the scientific basis for these techniques suggesting that aboriginal Australian communities were in fact studying the natural world in the same way as modern scientists, albeit within their own cultural context (Norris, 2016). What is important in this case is not whether song-lines as an oral tradition are as effective and precise as the compass (which they very well may not be) but rather that they did in fact serve as techniques aboriginal communities could rely on in navigating the ocean for generations. The introduction of the compass in this case served as a technical powerbroker in justifying both the dismissal of the aboriginal practice altogether from narratives of scientific knowledge production and also the subsequent privileging of western nautical science.

At this point, one may however argue that aggressive modes of epistemic progress could perhaps be attributed, in a Kuhnian sense, to signify progress of the scientific enterprise as a whole given their role in scientific revolutions. Is not the move from a *fragile* oral-tradition based method to one as robust and universal as the compass a paradigm shift in a sense?

It is important to recognize that what may seem from the outside as an extension of scientific knowledge altogether, may in fact arise out of what Spivak (1988) would call an epistemic violence, in the resulting 'silencing' of a particular community's ability to retain knowledge systems of their own, as can be seen in the case of the Australian aboriginals. Furthermore, even contemporary PhilSci literature that validate pluralist conceptions of science (such as the one that Massimi (2022) advocates in her Perspectival Realism) legitimize modes of knowledge production by several distinct historically and culturally situated epistemic communities, thereby rendering aggressive modes of epistemic progress as possibly detrimental to the scientific enterprise as a whole. It is then not necessarily how universal or 'objective' a technical object is in its operation that legitimizes it, but rather how specific communities have been able to develop certain techniques and technical objects from their own situated understanding of the world and that these techniques have served them reliably for purposes they directed them towards.

As I now conclude this chapter and turn towards ML-Objects and how they may possibly serve as either technical bridges or powerbrokers, I would like to re-emphasize that there are crucial ethical (in terms of injustices to the historical contributions of certain epistemic communities) and epistemological (in terms of a hegemonic suppression of certain modes of knowledge production in an otherwise pluralist scientific enterprise) downsides to aggressive modes of epistemic progress.

CHAPTER 5

GENEALOGY OF MACHINE LEARNING

Taking a genealogical lens in the previous chapter helped us recognize that a technical object is not confined to the contexts (and communities) of its genesis. Instead, the technical object carries with it virtually the traces of future models whose construction it can enable, and knowledge that it can help produce in its interaction with different epistemic communities. These interactions can sometimes establish what I referred to as technical bridges and can be productive for both the community they arise out of (as a validation of the community's own knowledge practice) and the community that later incorporates it (in terms of advancements it brings to the recipient-community's knowledge practice). I have also shown how in some cases, these interactions may be violent, in that the technical object, rendered as a powerbroker, can be employed towards invalidating the epistemic contributions of a particular community. In what follows in this chapter, I shall extend my discussion of the ML-object as laid out in chapter 3 in the context of its interaction with other epistemic communities.

In particular, I will first, in section 5.1, argue why ML-objects have a tendency to privilege aggressive modes of epistemic progress in their operation as technical powerbrokers and then outline how their continued use in knowledge production despite this tendency can be detrimental to the scientific enterprise. I will then in section 5.2, investigate possible ways in which this tendency can be suppressed and outline how they can continue to be useful in knowledge practices.

5.1 THE ML-OBJECT AS A TECHNICAL POWERBROKER

The empirical adequacy of the ML-object which draws in all the applause that ML-based industrial applications receive reaches a limit when it enters epistemic communities, in that empirical results alone are not sufficient to justify their use in knowledge production. As was outlined in 4.2, for ML-objects to be integrated into epistemic practices, as technical bridges, they would have to abide by the two conditions of reconstructability and integrability.

In terms of reconstructability, my in the previous chapter 3 already pointed at how ML-objects cannot tell a story of invention and (in most cases) of construction of their underlying architectures. The ML-object not being constructed in the way scientific models are ,with an active role played by human cognitive resources (as was shown in 3.3.2) entails that the recipient who intends to employ the ML-object cannot reconstruct how and why they are built the way they are. Even efforts such as XAI to nevertheless reconstruct the ML-object qua model in intelligible ways would not be justified as evidenced by my discussion in 3.3.3. While it is indeed true that both scientific models and ML-objects provide predictions regarding a phenomenon, it is important to note that the understanding (representation) of the phenomenon is what makes these predictions possible in human-built scientific models which in the case of the ML-object remains inaccessible[27]. The non-reconstructability of the ML-object would mean one cannot make use of ML-objects in the invention of new technical objects in the way one would make use of scientific models. A continued use of ML-objects despite this non-reconstructability would then point towards the negligence of the scientist-engineer in that she would in effect be employing a technical object whose capacities and limits she is unaware of. The non-reconstructability of the ML-object and the subsequent inability to enable future constructions renders the ML-object as an isolated body of knowledge in that they cannot detach from their initial contexts (of the particular conditions that enabled their training) and re-attach to new epistemic contexts (guide the construction of new scientific models) in the way technical objects usually do, (as was discussed in 4.2).

[27] It is to be noted that I'm particularly referring to black-boxed ML-objects in this chapter whenever I refer to ML-objects. The simplest way to address many of the epistemic concerns I will layout in this chapter arising from the use of ML-objects is if one can just rely on purely interpretable ML-architectures such as decision trees, and generalized additive models. These architectures can be reconstructable, in allowing their construction to be rendered intelligible and can also be integrated into knowledge practices because the scientist-engineer can carefully curate the input features as guided by her own theoretical resources and can later verify from the interpretable operational schema of the model if the mode of reasoning employed by the model aligns with her own theoretical understanding of the target phenomenon.

In terms of the integrability condition, the ML-object's epistemically isolated nature as shown above also means that they cannot be integrated into epistemic communities alongside their existing resources. This is because, say, a meteorologist cannot employ an ML-driven weather prediction model in the way she would use other instruments like thermometers in her practice to aid her model-ing of weather phenomena. It would in fact be a choice made from the outside, in that one would have to choose whether to rely on predictions from ML-object or from a meteorologist's model. A societal adoption of purely ML-driven weather prediction tools would then suggest how the privileging of empirical results enabled the ML-object to be a powerbroker for computer science[28] as an epistemic community, and leads to the sidelining of the historical contributions made by meteorology as a practice in the context of weather prediction.

The ML-object's inability to abide by the conditions of reconstructability and integrability renders them incapable of establishing technical bridges and thereby makes them unable to be incorporated into epistemic communities, at least in the way other technical objects are. A society dogmatized by the empirical achievements of the ML-object and continues adopt ML-objects in tasks of knowledge production that were previously undertaken by appropriate epistemic communities renders them as powerbrokers for computer science, in that this adoption can only come with epistemic injustices[29].

One can already trace cases of epistemic injustice in terms of both epistemic severing and trademarking within the few knowledge practices that ML-objects are being adopted in and I will outline these cases in the following section.

---

[28] As I refer to computer science as an epistemic community in different parts of this chapter, note that I'm particularly referring to the community as it became after post-industrial turn given its tight affiliations to tech companies.

[29] I would like to re-emphasize again that in this chapter, and in the thesis altogether, I'm only reflecting on ML-objects that are used exclusively in knowledge production. Consequently, my critical remarks and allegations of epistemic violence would not apply to say, ML-driven applications such as those of data entry, customer service, traffic management, email filtering etc. There can indeed be ethical concerns regarding these applications as well, probably because of their discriminatory features or their socially disruptive nature, but reflecting on these concerns is not the object of this research. Having said that, this project can nevertheless be productive for later research that does reflect on these concerns in particular.

5.2 EPISTEMIC INJUSTICES UNDERLYING THE ADOPTION OF ML-OBJECTS

In terms of an *epistemic severing*, we already see how computer-vision based ML-objects, given their efficient classification of medical images are sidelining the practices of radiologists who have long produced a knowledge base that enabled their interpretation of medical images (Reardon, 2019). This severing of the contributions made by radiologists is analogous to the case of the Australian aboriginals I've laid out in the previous chapter: with the compass being the ML-object; and the British nautical science being computer science as an epistemic community. A similar case can be made in the case of weather-prediction, with 'state-of-art' neural networks such as Huawei's Pangu-Weather, Nvidia's FourcastNet, and Google DeepMind's GraphCast (Heikkilä, 2023). The fact that these three projects are developed by industry leaders in ML-research further strengthens my identification of ML-objects as powerbrokers, with meteorologists' historical epistemic contributions being sidelined by an almost hostile entry of the CS community into a knowledge practice that was previously outside its purview.

We can also identify cases of *epistemic trademarking* with the "creative" potential of generative neural networks such as DALL-E and Stable Diffusion being deemed as major achievements by ML-research, when such a feat was in fact only made possible by the contributions of several artistic communities whose creative works were non-consensually (Xiang, 2022) melded into the datasets that these networks were trained on. Furthermore, a study by Somepalli et al. (2022) questions the actual creative capacities of generative neural networks by showing cases where their "generations" merely reproduce the training data in that the network simply pieces together foreground and background objects that it has previously memorized. The study also shows how in some cases, the generated 'artwork' is semantically equivalent to a source image within the dataset, while not being pixel-to-pixel identical. Observations such as these strengthen my argument these networks enable an epistemic trademarking, and makes one question to what extent the generated images are made possible exclusively because of the contributions made by neural network itself and those that train and own it.

From these cases of epistemic severing and trademarking I have outlined above, it is apparent that a widespread societal acceptance of ML-objects (whose initial waves we have begun to witness in recent years) can facilitate an epistemic violence to a wide range of communities, whose historical contribution to knowledge production finds itself invalidated or undervalued. In addition, the various phenomena that these communities have grown to represent and understand in their own situated ways falls under the risk of being transferred to the purview of a single epistemic community, that of the post-industrial turn computer science.

Such a homogenization and concentration of knowledge by a single community with a single set of values threatens the complex heterogenous network that contemporary pluralist conceptions of science hold dear and I shall investigate this threat in the following section.

5.3 MACHINE LEARNING AND THE THREAT OF AN EPISTEMIC CRISIS

It is important to note that the concentration of knowledge production within a single epistemic community is not just concerning because of the resulting epistemic injustice and how it threatens our ability to acknowledge different situated knowledge practices. It also has implications for the epistemic value of the knowledge produced. Self-supervised ML-objects such as the GPT built on transformer-based architectures are not directed at a curated set of parameters and measurements (as is the case in supervised ML-objects) but are instead trained on an assumption that the neural network optimizes itself to create the most complete picture given the data at hand. A totalizing optimization such as this serves to create a single 'model' that works for a wide range of purposes, as is the case in the later ChatGPT application. 'Models' such as these that paint a 'complete' picture, take into account every single extractable feature[30] found within the data and are capable of being put into use in a wide range of contexts. Sandra Mitchel (2020) illustrates how complete models can in fact be detrimental to knowledge production given that it is the different choices made by the human-modeler such as identification of features, abstractions and idealizations that make a knowledge inference possible (as

[30] Note that the identification of these features in this case remains one that is not curated by the human

was also indicated in the B&K method laid out in 2.3). A model that accommodates every single property of the target phenomenon with the highest degree of precision and accuracy, would in fact fail to provide any useful knowledge of the phenomenon and instead replicates the properties of the phenomenon as a part of its functioning (Mitchell, 2020, p. 5). A more justified alternative would therefore entail building individual models, guided by different methodological choices, modelled by specific epistemic communities, directed at particular contexts (or purposes).

For instance, given the same geospatial setting, one could build a wide range of maps, each designed for a particular epistemic purpose. A transit map of a city would be optimized to illustrate connections between different transport modes, enabling travelers to better navigate the public transportation system, plan their routes, identify transfer points, and estimate travel times. Such a map would deviate from the real geographical spatial distribution and would instead privilege the particular purpose, that of navigation, in its representation of the city. In contrast, a topographical map that privileges accurate graphic representations of natural formations like mountains, valleys and bodies of water would be optimized to help urban planners determine best locations for new infrastructure and can be critical for disaster management in helping identify say, areas that are at a greater risk of flooding.

These two types of maps of the same city would privilege different features, simplify some of them and exclude others in ways that best fit the contexts in which they would later be operationalized in. The adequacy of the model in question would then be assessed not in terms of how comprehensively it captures the mechanisms of the target phenomenon into its own functioning with the most empirical adequacy but instead based on how well they serve and fit within a particular set of purposes and contexts (Mitchell, 2020, p. 7).

Furthermore, a takeaway from my initial discussion of technical objects serving as bridges between communities was that the two communities that they bridge together were distinct heterogenous epistemic communities. It is not that there is a single dominant tradition that governs how different communities produce the knowledge they do. We live in an era where a technique as unique as Acupuncture, built on the Chinese medical community's understanding of the body as characterized by the flows of Qi, a non-physical life force (Ye et al., 2019), is being incorporated by western medical practitioners based on evidence of its therapeutic properties (White, 2009) despite western medical science's physicalist understanding of biological phenomena. It is then not necessarily the case that it is one epistemic community that establishes technical bridges with several others but rather how these bridges make a complex rhizomatic network, each influencing and validating the other, to enable what Massimi (2022) calls a cross-perspectival assessment of scientific knowledge.

Accordingly, in the case of the CS research community pulling in different phenomena under its own epistemic purview, noteworthy is the privileging of empirical adequacy (precision and accuracy) over an intelligible understanding of reality altogether. The danger I have previously mentioned in 1.4 that machine learning posed to the practice turn in (philosophy of) science, in its exclusive privileging of the norm of utility and a complete disregard to the norm of truth, would then in fact not be limited to one domain of science but all those domains that ML-objects find their way into. Such a domination of methodological values specific to the post-industrial turn CS community also echoes what Russo (2022) calls methodological imperialism, characterized by an imposition of methodological criteria specific to one field onto other fields.

I now conclude this section with the hope that arguments I have laid out above demonstrate how grave of a threat an ML-driven knowledge production can pose to the scientific project, both in its dissemination of pluralist conceptions of science and also in its disregard towards an intelligible understanding of the world.

5.4 ML-OBJECTS AS STARTING POINTS FOR FUTURE SCIENTIFIC INQUIRY

Having outlined the ML-object's inability to establish technical bridges between epistemic communities and how they can be detrimental to the scientific project altogether, I will now, in this section, identify possible ways in which epistemic communities can still benefit from black-boxed ML-objects in their knowledge practice even if they are not directly incorporated into the practice itself.

The type of interaction between epistemic communities and ML-objects that I will illustrate in this section as being productive for knowledge practices is one that neither establishes the ML-object as technical bridge nor as a technical powerbroker. The ML-object, I will argue, can help further the knowledge practice of a scientific community, if the community approaches the ML-object, not in the way it approaches other scientific instruments, neither in the way it approaches scientific models, but as it would approach an anomalous natural phenomenon that it seeks to reproduce.

What does it mean for a scientist to approach an ML-object as they would a natural phenomenon? I will illustrate what I mean with an example:

In 1991, climate scientists observed a global average temperature drop of around 0.5 degrees. Initially, there was no explanation or understanding as to why this drop has occurred. Scientists have later traced how this drop was caused by an eruption of Mount Pinatubo volcano in Philippines. This identification has led to further research in the relationship between volcanic eruptions and global temperature levels, which further lead to the identification of the role of sulfur dioxide and other aerosols (which are injected into the atmosphere during volcanic eruption) in scattering solar radiation and thereby cooling the planet. This knowledge of the phenomenon has eventually led to the development of solar radiation management as a technique that would mimic the effects of volcanic eruptions with the aim of artificially reducing temperature levels.

In the above example, the discovery of the phenomenon of the temperature drop served as a starting point for future inquiry. The scientists knew that there had to be some cause to this anomalous temperature shift and that has led them after a long line of research investigations to construct a model that mimics this phenomenon. In the same way, black-boxed ML objects can help serve as starting points for future inquiries despite them being un-understandable like the anomalous temperature shift phenomenon. For instance, Duede (2022) draws on the case of human-constructed geophysical models of earthquakes and how they have poorly functioned in relating mainshock and aftershock locations (2022, pp. 10). In this case, researchers have indeed built a fully black-boxed ML-object to predict aftershock locations based on the mainshock event. However, it is not the case that they employed this network in the way they would employ other scientific instruments. They instead interpreted the ML-object's superior empirical performance as evidence that the existing human-built scientific model can be significantly improved. They knew their initial assumption of a relationship between aftershocks and mainshocks is indeed true from the black-boxed ML-objects performance. This served as a justification for them to continue on their current path of inquiry to further refine and improve their own model. The main insight from Duede's case is that throughout the process, the ML-object remained black-boxed, and scientists remained foreign to its inner workings. Despite this opacity and despite there being no bridging between the theoretical resources themselves and the ML-object, neural network continued to influence the progress made within the discipline.

The ML-object in this case was not integrated into the knowledge practice itself. It instead served as a starting point for future research in the same way the 1991 temperature drop did. If the scientists do end up constructing a better scientific model, the justification for this model would stem from the validity of their own model construction, and the theoretical resources from geophysics they made use of, but not from the neural network. The ML-object remained outside the knowledge practice while still guiding and enabling the researchers to build models in their own ways. The role played by the ML-object in this case can find its equivalence in similar situations across knowledge practices where scientists have some preconceptions about a phenomenon before constructing models. They could use an ML-object built with these preconceptions as its conditions to verify if the line of inquiry they are about to pursue can indeed be promising.

It is important to note that the above outlined interaction between the ML-object and the epistemic community cannot be justified with all ML-architectures. In Chapter 3, I have distinguished ML-objects from ML-architectures and have shown how some architectures, particularly those before the industrial turn, do tell a story of construction in that they are built in the way they are because of intentional choices made by the architects, and they carry certain pre-established presuppositions about the kind of informational patterns they can recognize. The geophysical scientist would have to be aware of these presuppositions and be able to justify why the architecture is of the size and structure that it is, to be able to ascertain that the relationship between the aftershock and the mainshock does exist. Such a justification cannot be arrived at by the use of certain post-industrial turn architectures such as the *Transformer* which cannot tell this story (as discussed in 3.2.4). In fact, if given access to sufficient data and a high enough computational power, the use of the transformer could show a exceedingly strong relationship between the mainshock and the aftershock event than what is truly the case. This can set unrealistic expectations for the geophysicists as they build their model and can de-incentivizing them from looking at secondary factors.

After taking a very justifiably critical stance on the use of ML-objects in knowledge production in the previous sections of this chapter, I wish I could have avoided ending this chapter with this slightly positive role that ML-objects can play in knowledge production, so as to not let these few benefits undermine the importance of my earlier critique. I have however laid out this aspect of the ML-object regardless of my own stance, to make room for some actionable normative insights as opposed to advocating an absolute outlawing of any influence of ML-objects on scientific practices. I would nevertheless still re-emphasize for the reader that the scope of the normative insights in this section is indeed limited and in no way does it undermine the epistemic threat that black-boxed ML-objects pose to the scientific project.

CONCLUSION

Having pursued an ontogenetic study of ML-objects in both their genesis and genealogy, this thesis has brought to surface i) how ML-objects can be distinguished from traditional technical objects both in their construction and in their instrumentalization in knowledge practices ii) the epistemic crisis that arises out of the lack of human participation in the construction of the ML-architecture and the invention of the ML-object and ii) the subsequent epistemic injustices that stem from a continued use of the ML-object in knowledge production, despite this lack.

At a more disciplinary level, my reconstruction of Simondon's work through contemporary developments in PhilSci would serve in establishing a bridge between PhilSci and ethics and philosophy of technology (PhilTech), by equipping the latter with conceptual tools such as i) the two criteria of reconstructability and integrability and ii) the two notions of epistemic violence namely, severing and trademarking. These conceptual tools would help strengthen PhilTech (which has otherwise remained hylomorphic in its approaches towards technologies) in its own critical reflections on AI applications and take into consideration their underlying processes of construction and invention.

One could very well argue that my identification of the hylomorphic undertones of PhilTech and my emphasis on the need to take into account the ontogenetic aspect of technology may be misguided given that the developments in the so called "empirical turn" in PhilTech, characterized by approaches such as Post-Phenomenology and Value-Sensitive Design, do not just descriptively reflect on existing technologies, but also play an active normative role in their design and implementation. However, as I have shown in my account on the genesis of technical objects, it is important to recognize that the technical object retains an identity in its underlying technical schemas which remains independent of its later implementations in concrete applications. These schemas are pre-established by the activities of construction and invention thereby rendering the later design of artefacts to already take these schemas as starting points. Whether it is post-phenomenology in curating technical objects that fulfill specific mediations or Value-Sensitive Design in its attempts to design artefacts that embody a specific set of values, the range of normative choices that designers can accommodate, while being guided by these approaches, is already constrained by the type of technical schemas that arise out of model construction and technical object invention. These fields would then be able to make largely surface level changes while leaving underlying modes of operation intact, or at best, choosing from an already existing set of technical schemas made available to them.

Furthermore, acknowledging the role of human cognition in the activities construction and invention and the subsequent rendering of technical objects as bearers of knowledge (2.4.1) would also serve as a critique to Heideggerean approaches in PhilTech that reduce technology to an immaterial attitude based on what can be done with it (as enframing[31]) and the subsequent mis-identification of science as being inherently technical by Heideggerean scholars[32] in PhilTech.

The main insight for PhilTech from my critical reflections on technical objects through Simondon's work and PhilSci is then that drawing attention towards the ontogenetic aspect of technical objects brings to surface the particular ways in which the technical activity exerts its own deterministic influence (in shaping how technologies are later instrumentalized for a particular utility or an effect on society) which is otherwise either unquestioned or remains as an implicit ideology in most, if not all, approaches in PhilTech.

I also hope the insights I have provided in this thesis can help both PhilSci and PhilTech assess individual AI-applications based on to what extent their stories of construction and invention (qua formation) can be reconstructed and whether they pose any threat of epistemic violence. It is important to recognize that there is nothing fundamental to the ML-architecture itself that makes it agent of an aggressive mode of epistemic progress. It is then quite unfortunate that among the several tedious tasks that human communities engage in, it is those of serious epistemic significance and those that require human participation the most that are being substituted (or face the threat of substitution) by state-of-art AI applications such as DALL-E and ChatGPT. While it could be true that such

[31] See Heidegger (1977)

[32] See Zwart (2022)

applications may be justifiably adopted if rendered transparent in how they use and manipulate epistemic resources (as data) that other communities have historically contributed to, such a transparency cannot come from the use of XAI methods given their 'approximatory' nature (as I have shown in 3.3.3). Their adoption can nevertheless be justified if they do arise out of purely interpretable architectures. Given that ML-research is primarily being carried out either by industries or institutes with industrial affiliations, it is no surprise that the dominating ML-architectures are those that require resources which no one but the industry has access to, and these architectures being black-boxed because of an alleged technical limitation conveniently lets the industry monetize on research that on paper remains open-source. From my discussion in 3.3.4, it is evident that there is no fundamental technical limitation that forbids us from employing interpretable architectures and that if sufficient resources are pooled in, there is no reason to believe we cannot have interpretable architectures capable of performing complex computational tasks.

Although I have centered my discussions on scientific communities in particular and their ability to reconstruct the story of invention, Simondon's project in the way he intended it to be wasn't directed at just the scientific community. The kind of society that Simondon dreamt of is one where everyone who uses a technical object is able to "know" it by its true nature, by its underlying technical schemas and modes of operation. Only then can the hylomorphic schema be avoided and the human alienation from the technical object be resolved. This alienation, and users of technical objects not being able to reconstruct their inventions, has existed in societies long before the advent of the AI-paradigm. It is because we live in such a society that the average person finds it unproblematic to transition from a technical object to an ML-object: how would it matter for someone if the weather predictions they receive arise out of human-constructed models or an ML-object, if the former was equally opaque to them begin with. Such an attitude was what enabled the post-industrial turn CS community to conveniently seep into other knowledge practices in the first place and continue their privileging of empirical results as the only criteria of validity with no regard for providing an intelligible understanding of the world. For a society that adopts and relies on ML-driven knowledge production, it is not just the ML-object that is black-boxed —it is the historical contributions of different epistemic communities that are black-boxed; it is reality itself that is rendered black-boxed.

BIBLIOGRAPHY

Apostle, H. (1970). *Aristotle's Physics Books 1 and 2*.

Baird, D. (2004). *Thing knowledge: A philosophy of scientific instruments*. Univ of California Press.

Barthélémy, J.-H. (2012). Glossary: Fifty Key Terms in the Works of Gilbert Simondon. In A. Woodward, A. Murray, & J. Roffe, *Gilbert Simondon* (pp. 203–231). Edinburgh University Press. https://doi.org/10.1515/9780748645268-015

Boon, M. (2015). The Scientific Use of Technological Instruments. In S. O. Hansson (Ed.), *The Role of Technology in Science: Philosophical Perspectives* (Vol. 18, pp. 55–79). Springer Netherlands. https://doi.org/10.1007/978-94-017-9762-7_4

Boon, M. (2020a). Scientific Methodology in the Engineering Sciences. In D. P. Michelfelder & N. Doorn (Eds.), *The Routledge Handbook of the Philosophy of Engineering* (1st ed., pp. 80–94). Routledge. https://doi.org/10.4324/9781315276502-8

Boon, M. (2020b). The role of disciplinary perspectives in an epistemology of scientific models. *European Journal for Philosophy of Science*, *10*(3), 31. https://doi.org/10.1007/s13194-020-00295-9

Boon, M. (2022). How Philosophical Beliefs about Science Affect Science Education in Academic Engineering Programs: The Context of Construction. *Engineering Studies*, *14*(2), 109–133. https://doi.org/10.1080/19378629.2022.2125398

Bratko, I. (1997). Machine learning: Between accuracy and interpretability. In *Learning, networks and statistics* (pp. 163–177). Springer.

Contessa, G. (2010). Scientific models and fictional objects. *Synthese*, *172*, 215–229.

Domingues, H. M. B. (2020). *Rubber: The invisible movement of traditional knowledge*. 590–594.

Dotson, K. (2011). Tracking Epistemic Violence, Tracking Practices of Silencing. *Hypatia*, *26*(2), 236–257.

Dronsfield, A. (2010, January 5). Mass spectrometry—The early days. *EIC*. https://edu.rsc.org/feature/mass-spectrometry-the-early-days/2020189.article

Duede, E. (2022). Deep learning opacity in scientific discovery. *ArXiv Preprint ArXiv:2206.00520*.

Dziugaite, G. K., Ben-David, S., & Roy, D. M. (2020). *Enforcing Interpretability and its Statistical Impacts: Trade-offs between Accuracy and Interpretability* (arXiv:2010.13764). arXiv. http://arxiv.org/abs/2010.13764

Ellul, J. (2021). *The technological society*. Vintage.

Fraser, B. (2014, December 12). *Global change, climate change upend livelihoods, food security in the Amazon*. Forest News. https://forestsnews.cifor.org/25881/global-change-climate-change-upend-livelihoods-food-security-in-the-amazon#.VJkt0v8r7A

Giere, R. N. (2004). How Models Are Used to Represent Reality. *Philosophy of Science*, *71*(5), 742–752. https://doi.org/10.1086/425063

Giere, R. N. (2006). Perspectival pluralism. *Scientific Pluralism*, 26–41.

Greshgorn, Dave. (2020, August 20). GPT-3 Is an Amazing Research Tool. But OpenAI Isn’t Sharing the Code. | by Dave Gershgorn | OneZero. *Onezero*. https://onezero.medium.com/gpt-3-is-an-amazing-research-tool-openai-isnt-sharing-the-code-d048ba39bbfd

Heidegger, M. (1977). The question concerning technology. *New York*, *214*.

Heikkilä, M. (2023, November 7). *Weather forecasting is having an AI moment*. MIT Technology Review. https://www.technologyreview.com/2023/07/11/1076067/weather-forecasting-is-having-an-ai-moment/

Hill, D. P., Berardini, T. Z., Howe, D. G., & Van Auken, K. M. (2010). Representing ontogeny through ontology: A developmental biologist's guide to the gene ontology. *Molecular Reproduction and Development: Incorporating Gamete Research*, *77*(4), 314–329.

Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *6*(02), 107–116.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hui, Y. (2016). *On the existence of digital objects*. University of Minnesota Press.

Hutson, M. (2018). AI researchers allege that machine learning is alchemy. *Science*, *360*(6388), 861.

Jordan, M. I. (1997). Serial order: A parallel distributed processing approach. In *Advances in psychology* (Vol. 121, pp. 471–495). Elsevier.

Kaplan, S. (2019). The Renaissance and Preformation. *Occam's Typewriter*. https://occamstypewriter.org/stevecaplan/2019/06/07/the-renaissance-and-preformation/

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions. Chicago (University of Chicago Press) 1962*.

Kumar, A. (2015, February 25). Machine Learning—7 Steps to Train a Neural Network. *Data Analytics*. https://vitalflux.com/machine-learning-7-steps-train-neural-network/

Lakatos, I. (1976). *Falsification and the methodology of scientific research programmes*. Springer.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541–551.

Lipton, Z. C., & Steinhardt, J. (2018). Troubling trends in machine learning scholarship. *ArXiv Preprint ArXiv:1807.03341*.

Malik, S. (2021, February 16). *Deep Learning and Compute: Can we just keep Scaling?* UCL FTR. https://www.uclftr.com/post/ai-and-computing-power

Marks, J. (2006). Introduction. In *Deleuze and Science* (Vol. 29). Edinburgh University Press.

Massimi, M. (2022). *Perspectival Realism* (1st ed.). Oxford University Press. https://doi.org/10.1093/oso/9780197555620.001.0001

Mitchell, S. D. (2020). Instrumental Perspectivism: Is AI Machine Learning Technology like NMR Spectroscopy? *A Critical Reflection on Automated Science: Will Science Remain Human?*, 27–42.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Norris, R. P. (2016). Dawes review 5: Australian aboriginal astronomy and navigation. *Publications of the Astronomical Society of Australia*, *33*, e039.

Norris, R. P., & Harney, B. Y. (2014). Songlines and navigation in Wardaman and other Australian Aboriginal cultures. *ArXiv Preprint ArXiv:1404.2361*.

Piatti, G. (2016). *The life and the crystal. Paths into the virtual in Bergson, Simondon and Deleuze*. *3*.

Radder, H. (2009). Science, technology and the science-technology relationship. In *Philosophy of technology and engineering sciences* (pp. 65–91). Elsevier.

Reardon, S. (2019). Rise of robot radiologists. *Nature*, *576*(7787), S54–S58.

Rudin, C. (2019). *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead* (arXiv:1811.10154). arXiv. http://arxiv.org/abs/1811.10154

Russo, F. (2022). *Techno-scientific practices: An informational approach*. Rowman & Littlefield.

Russo, F., Schliesser, E., & Wagemans, J. (2023). Connecting ethics and epistemology of AI. *AI & SOCIETY*, 1–19.

Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. *AJOB Neuroscience*, *11*(2), 88–95. https://doi.org/10.1080/21507740.2020.1740350

Saxe, A. M., Bansal, Y., Dapello, J., Advani, M., Kolchinsky, A., Tracey, B. D., & Cox, D. D. (2019). On the information bottleneck theory of deep learning. *Journal of Statistical Mechanics: Theory and Experiment*, *2019*(12), 124020.

Sevilla, J. (2022, February 16). *Compute Trends Across Three Eras of Machine Learning*. Epoch. https://epochai.org/blog/compute-trends

Sharp, A. (1964). *Ancient voyagers in Polynesia*. Univ of California Press.

Simondon, G. (1958). *Du mode d'existence des objets techniques. Paris: Aubier-Montaigne*.

Simondon, G. (2017). *On the Mode of Existence of Technical Objects* (1st ed.). Univocal Publishing.

Somepalli, G., Singla, V., Goldblum, M., Geiping, J., & Goldstein, T. (2022). *Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models* (arXiv:2212.03860). arXiv. http://arxiv.org/abs/2212.03860

Spivak, G. C. (1988). Can the Subaltern Speak? *Die Philosophin*, *14*(27), 42–58.

Stiegler, B. (1998). *Technics and time, 1: The fault of Epimetheus* (Vol. 1). Stanford University Press.

Sullivan, E. (2022). Understanding from Machine Learning Models. *The British Journal for the Philosophy of Science*, *73*(1), 109–133. https://doi.org/10.1093/bjps/axz035

Sullivan, M. (2023, May 17). *The frightening truth about AI chatbots: Nobody knows exactly how they work*. Fast Company. https://www.fastcompany.com/90896928/the-frightening-truth-about-ai-chatbots-nobody-knows-exactly-how-they-work

Van Fraassen, B. C. (2010). *Scientific representation: Paradoxes of perspective*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Voss, D. (2020). 3 Modulating Matters: Simondon, Deleuze and Guattari. In R. Przedpełski & S. E. Wilmer, *Deleuze, Guattari and the Art of Multiplicity* (pp. 50–63). Edinburgh University Press. https://doi.org/10.1515/9781474457675-007

Watson, D. (2019). The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence. *Minds and Machines*, *29*(3), 417–440. https://doi.org/10.1007/s11023-019-09506-6

White, A. (2009). Western medical acupuncture: A definition. *Acupuncture in Medicine*, *27*(1), 33–35.

Xiang, C. (2022, September 26). *AI Is Probably Using Your Images and It's Not Easy to Opt Out*. VICE. https://www.vice.com/en/article/3ad58k/ai-is-probably-using-your-images-and-its-not-easy-to-opt-out

Yam, P. (2004). Everyday Einstein. *Scientific American*, *291*(3), 50–55.

Yang, W., Yan, J., Zhuang, P., Ding, T., Chen, Y., Zhang, Y., Zhang, H., & Cui, W. (2022). Progress of delivery methods for CRISPR-Cas9. *Expert Opinion on Drug Delivery*, *19*(8), 913–926.

Ye, J., Cai, S., Cheung, W. M., & Tsang, H. W. (2019). An east meets west approach to the understanding of emotion dysregulation in depression: From perspective to scientific evidence. *Frontiers in Psychology*, *10*, 574.

Yuan, Y. (2020, September 7). Viral Post Highlights 'Toxicity Problems' in the Machine Learning Community. *Synced: AI TECHNOLOGY & INDUSTRY REVIEW*.

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, *64*(3), 107–115.

Zwart, H. (2022). *Continental philosophy of technoscience*. Springer Nature.