UTKU B. DEMIR

GENERATIVE AI AND THE INSTITUTIONS OF CONTROL: AN ANALYSIS OF ALGORITHMIC MEANING-MAKING AND SUBJECTIVATION

Contents

Glossary

epoch Epochs represent the number of times the entire training dataset passed through the algorithm. Nebius-Team 2024. 40, 48

kernel In Machine Learning, the Kernel method consists of using a linear classifier to solve a non-linear problem. This is achieved by transforming a linearly inseparable set of data into a linearly separable set (Melanie 2024).. 34

token Tokens are the smallest units of text that a model processes; typically words, subwords, or characters in natural language processing tasks (Cser 2024). In current AI systems, a token often corresponds to a single word, and the process of breaking text into tokens is known as tokenization (Jurafsky et al. 2009, p. 59). For genAI models, particulary LLMs, this enables efficient computation across varying text inputs.. 30, 34, 35, 37, 42

Acronyms

AGI Artificial General Intelligence. 15, 45

AI Artificial Intelligence. 2, 8, 15, 16, 21, 24, 25, 26, 27, 28, 29, 31, 45, 50, 57, 59, 60, 61, 64, 68, 73

BwO Body without Organs. 66, 73

CNN Convolutional Neural Network. 34

D&G Gilles Deleuze & Felix Guattari. 10, 11, 12, 19, 23, 28, 37, 56, 60, 61, 62, 63, 64, 66, 68, 69, 74

DL Deep Learning. 26, 29, 30, 34, 39

DNN Deep Artificial Neural Network. 8, 29

genAI Generative Artificial Intelligence. 2, 3, 4, 5, 8, 9, 10, 15, 16, 22, 23, 24, 25, 26, 28, 30, 31, 33, 34, 35, 42, 45, 46, 48, 49, 51, 52, 53, 54, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 73, 76

GM Generative Model. 49

GOFAI Good old-fashioned AI. 26, 40

LLM Large Language Model. 3, 5, 8, 9, 15, 16, 25, 26, 31, 34, 35, 36, 37, 42, 43, 45, 52, 54, 73, 75

LM Language Model. 30

MGM Multimodal Generative Model. 30

ML Machine Learning. 27, 29

NLP Natural Language Processing. 2, 8, 26, 28, 29, 34, 46

NN Artificial Neural Network. 8, 28, 29, 31, 33, 45

RLHF Reinforcement Learning from Human Feedback. 74

RNN Recurrent Neural Network. 34

SL Supervised Learning. 29, 45

SSL Self-Supervised Learning. 30

symAI Symbolic Artificial Intelligence. 26, 27, 31

T2IM Text to Image Model. 30

UL Unsupervised Learning. 29, 30, 46

1

Introduction

In recent years, substantial advancements in the field of AI, particularly through developments in Artificial Neural Network (NN) and Deep Artificial Neural Network (DNN) architectures, have enabled the deployment of predictive models across a wide array of domains, from social media platforms and search engines to natural language processing tasks such as text classification and topic modelling. While these applications primarily focused on analysis, relevance association, personalisation, and prediction, a new paradigm has emerged in the form of genAI. Once a relatively silent front in NLP research, genAI's history dates back to the 1950s (Cao et al. 2023, p. 4). Unlike traditional models, genAI systems are capable of producing novel outputs; such as text, images, or code by extracting and operationalising intent from human-provided instructions. This shift marks a transformation not only in the goals and capabilities of AI, but also in its epistemic and operational logics. Particularly in its implementations based on transformer architecture, genAI now occupies a central role in the production, interpretation, and circulation of information and media, moving beyond automation and decision support to enabling generative processes that raise fundamental questions about agency, subjectivity, and truth.

The analysis and critique of the AI models is nothing new, the surveillance capabilities that has been established by the contemporary data analysis (e.g. Krasmann 2017), the effect of a completely data based rationality introduced by the datalogical turn (see Clough and Gregory 2015), a dividualised information flows through the profiling and association by the models running on the web (see e.g. Cheney-Lippold 2011), and the decision-making systems adopting an algorithmic governmentality (see e.g. Rouvroy 2007), and various ethical, as well as, bias related research (e.g. Kordzadeh and Ghasemaghaei 2022) have been a vibrant field in the last years. The capability genAI models especially LLMs to meaning-making (Dishon 2024; Gretzky 2024; Mishra and M. K. Heath 2024), however, , have provoked renewed inquiry. While these generative processes are rooted in a long history of statistical and computational development, contemporary architectures with their interpretation of the

vast datasets of productive human legacy introduce an immediate representational logic, one that embodies a distinct political model or *governing rationality* (Amoore et al. 2024, p. 2)[1]. While this interpretative substance enables genAI models to communicate human-like, also establishes a power structure over governing information as a governing institution (see e.g. MacKenzie and Porter 2021 or Dishon 2024). Beyond their technical capacities, these systems enact a form of governance deeply entangled with power, normativity, and new forms of subjectivisation (Eloff 2021). GenAI operates through a distributional logic that structures knowledge by modelling statistical regularities in a datafied world (Amoore 2023). These systems "traverse data foundations" to generate outputs that appear plausible within a learned distribution, but are often critiqued with the lack of deterministic causality and transparency in justification. Whether genAI models blur the line between representation and enactment; especially the outputs form LLMs , for instance, are increasingly treated as epistemically meaningful, even authorial, despite the large discussion about the lack of intentional agent which brings us to an even richer debate and literature about the agency in human autorship.

[1] The governing rationality of genAI models (see e.g. ibid.) refers to the generative structure of algorithmic meaning-making and should not be confused with Rouvroy n.d.'s concept of Algorithmic Governmentality. Whereas Rouvroy addresses algorithmic turns in neoliberal governance, governing rationality designates the internal logic established by the model itself.

TODO: Title

- ☐ This discussion line can be followed up on
- ☐ Mention somewhere soemthing like: My work focuses especially on the architectural components that made the meaning-making process of the genAI Models as comprehensive as today, especially those of transformer architecture like "attention", "latency", "gradient descent".

As Michel Serres (2019, p. 41) put, *the forces shaping our bodies* increasingly shifted from natural conditions to environments of our own making; the constructed world exerts influence more than the given one. The rapid acceleration of technological change has drawn the human condition more deeply into a culturally defined context, shaped by the flow of information and new forms of mediation than into any purely natural domain. The models capable of meaning-making are farther away from being mere classication and filtering tools, they actively participate in structuring the mediation of culture and knowledge, further amplifying their ability to (re-)structure the reality, . What genAI introduces is beyond its technical nature inquires institutional analysis (MacKenzie and Porter 2021) of the power structure deployed in this new constellation. Framing the question as *in what form of institutional nature the architecture and rationality of the contemporary genAI algorithms deploy and what conclusions these implicate on agency, subjectivisation, ciritique, and resistance*; this study situates genAI within the broader transformation of power described by Gilles Deleuze (1992) as the shift from Michel Foucault's *disciplinary societies* (1977) to *societies of control*. Whereas Foucault's account of disciplinary power emphasized enclosure, surveillance, and the moulding of subjects within bounded institutions like prisons, schools, and hospitals (Foucault 2008), Deleuze's postscript outlines a more fluid and continuous, flexible form of control. Deleuze's description of governance in the "Postscript on the Societies of Control" operates through modulation: subjects are governed not by confinement but through their data traces, captured and recomposed in real-time[2]. In such control societies, individuals become *dividuals*, decomposed into discrete, analysable data points recombinable by

[2] One frequently necessary clarification in this context is that the concept of disciplinary societies refers to a specific mode of operation of (bio-)power; speaking of control or post-disciplinary societies does not imply the disappearance or replacement of discipline. Rather, one might underscore that control is itself a continuation or transformation of discipline (see Kelly 2015)

algorithmic systems (MacKenzie and Porter 2021) vastly increasing the field of visibility on the bodies (Foucault 2008). Control does not dissolve institutions into flow; rather, it reorganizes them into mechanisms that totalise by sequencing dividuals across domains. Institutions of control no longer discipline by containment, but by aggregating, modelling, and redistributing datafied subjectivities through infrastructures such as genAI platforms.

Resistance does not have to be necessarily mean halting development or rejecting these systems outright; rather, it involves maintaining an openness to transformation while actively engaging with and shaping these changes where intervention remains possible (see Tucker 2021a, p. 227).

TODO: Title

☐ Rewrite

This thesis therefore advances the hypothesis that generative AI models function as *institutions of control*, not metaphorically, but operationally. Their architectures instantiate regimes of truth through statistical inference, acting as epistemic infrastructures that determine what can be said, imagined, or inferred. This shift raises the stakes of critique. In control societies, resistance cannot depend on unmasking ideology or demanding transparency alone. Instead, critique must become processual and counter-sequential: it must trace the operations of sequencing and propose alternative arrangements that disrupt the logic of totalisation (MacKenzie and Porter 2021) . Accordingly, I adopt a micropolitical perspective, asking not merely what genAI systems do, but how they do it. What are the machinic elements, attention mechanisms, tokenisation, transformer layers that enable the modulation of information and subjectivity? And where, if anywhere, might one locate lines of flight within these architectures? Can their operation be reappropriated as tools for critique, invention, or resistance?

Rather than positioning genAI as either emancipatory or repressive, this study approaches it as a complex institutional actor embedded in contemporary capitalism, furthermore develops a ciritical reflection of Deleuze's concept of control by deviating both in terms of the analysis of the genAI architecture, and in the critique of the institutional formation these models establish other works of Gilles Deleuze & Felix Guattari (D&G) (see e.g. Deleuze and Guattari 1983, Deleuze and Guattari 1987). As D&G argue, capitalism decodes and deterritorialises flows only to reterritorialise them elsewhere (Deleuze and Guattari 1983). In this context, genAI functions not merely as a tool of production or surveillance, but as a mechanism of epistemic reterritorialisation: producing coherence, narrativity, and alignment from fragmented inputs. It governs the production of meaning while also embedding the potential for alternative uses, ones that may exceed or redirect capitalist logics. This study thus offers a re-critique of control societies through the lens of generative architectures. Combining political theory, and technical analysis, it examines how genAI models operate on both infrastructural and epistemic levels. In doing so, it seeks to develop a renewed account of power, one grounded

in probabilistic modulation, infrastructural inscription, and the micropolitics of machine reasoning. Ultimately, I argue that while the generative capabilities challenge the pillars of the control society concept, while finding particularly insightful correspondences in other literature of D&G .

Consider the following:
• What does it mean when a machine modulates language, rather than represents it? • How does a transformer model participate in the social, not as a metaphor, but as a machinic node? • What kinds of institutions are being formed not through buildings or laws, but through attention-weighted predictions and recursive training?

2

Theoretical Framework

This chapter is mostly incomplete

One could argue that artificial intelligence is *no longer an engineering discipline* (Dignum 2023, p. 206), if it has ever truly been one.

The present thesis builds on critical perspectives from political theory and philosophy of technology to interrogate the institutional, epistemological, and political implications of contemporary generative AI systems by analysing their architectural structures and through the analysis is situated within D&G's conceptual apparatus of control, desire, and modulation. Together, these approaches provide the foundation to conceptualize generative AI not merely as computational artefacts but as infrastructural actors in the contemporary configuration of control societies.

2.1 Research Question

RQ: How does the development of the current AI algorithms, particularly those of generative AI models (genAI), embody, extend, or contradict Gilles Deleuze's concept of societies of control by modulating subjectivity through its probabilistic, non-linear operation structures?

RQ (Alternative): To what extent can the processes of meaning-production in generative AI algorithms, particularly the transformer architecture be understood through Gilles Deleuze's concepts of control and modulation, especially in relation to subjectivity construction via their probabilistic, non-linear operational structures?

RQ(Alternative): In what ways do generative AI models instantiate the logics of control society, and how might their probabilistic architectures transform the institutional production of subjectivity and meaning?

2.1.1 Possible lines of argument

- The distributional structure of generative AI, which produces content based on joint probability distributions, creates a dynamic

of non-linear causality that reflects modulation without control, where subjectivity isiiiii formed through continuous, data-driven modulation rather than direct causative action in the absence of directional control.

- Despite their current state, future genAI models offer unique ways for individuals to create lines of flight and support the becoming of nomadic subjects.
- Transformer-based genAI systems operate as infrastructural institutions that govern subjectivity through modulation rather than disciplinary enclosure; their architectures do not command but statistically orchestrate behavior and meaning.
- Rather than merely reflecting ideology, genAI systems produce truth regimes through processes of probabilistic inscription, positioning themselves as epistemic institutions within the broader infrastructure of control.
- Despite their instrumental role in capitalist modulation, genAI architectures retain machinic singularities that can be redirected toward critique, invention, and experimental modes of subjectivation, forming potential "lines of flight."

[1]

[1] **TODO**: Revise

2.2 *State of the Art*

TODO:

- ☐ This part is to be adcvanced later on
- ☐ Rouvroy n.d.
- ☐ Pasquinelli 2023

While there are various attempts to adapt the notion of control to modern digital developments, such as those by Brusseau 2020, or partially or completely rejecting the concept of control as an inadequate explanation of current digital power structures (see, for example, Hui 2015), these works either overlook the novelty and potential of genAI mechanisms or fail to explore the *dividual* dimension. Furthermore, earlier works focusing specifically on the aspect of dividualisation (see e.g. Cheney-Lippold 2011; Van Otterlo 2013) are, unfortunately, temporally limited as they were unable to analyse generative models that had not yet reached the level of advancement we see today.

Other theorists focus on various aspects of genAI models, with a common tendency to analyse ethical considerations, their application under neoliberal governmentality, and their role within surveillance capitalism (see e.g. Gillespie 2024; Haggerty and Ericson 2000; Zuboff 2019). While these aspects are not outside the scope of this research, they were central to my previous bachelor's thesis. This current study specifically focuses on the structure of probabilistic models, particularly examining their construction at present (and speculatively in the future), without delving into how their misuse (or intentional use) may play out.

Furthermore, while this study addresses the political implications of current generative AI usage, its focus is primarily on the analysis of power operations and, within this context, relevant to questions of democracy only insofa as they relate to the operation of power. This study does not, however, aim to provide a comprehensive analysis of democratic risks associated with AI, as explored by others (see e.g. Zarkadakēs and Tapscott 2020; Coeckelbergh 2024).

This research, therefore, is focused on three primary areas of debate: (1) an analysis of genAI models as entities deploying power, examining their mechanisms (see e.g. Amoore et al. 2024; Konik 2015; MacKenzie and Porter 2021); (2) modern reflections on Deleuze's theory (see e.g. Mischke 2021; Poster, Savat, and Deleuze 2010); and (3) sources analysing algorithmic structures with technical expertise (see e.g. Bender et al. 2021; Vaswani et al. 2017).

3

Subjectivity and the Shift from Discipline to Control

> The society of control is characterized not by the power of the institutions of modernity, or pre-modernity, the army, the prison, the university, the church, but instead by what he Deleuze called "the ultra-rapid forms of free-floating control that are inherent in distributed networks".
>
> — A. R. Galloway 2004, pp. 318–319, MacKenzie and Porter 2021

TODO:

- ☐ Introduce the critique of the Postscript, also reflect on ibid.
- ☐ "Every regime introduces some form of governance of desire"
- ☐ tell how control replaces the institutional formation of Foucault
- ☐ add Burroughs: Burroughs never thought about the control as bidirectional thing (mayan shaman citation)
- ☐ **To be completed with following**: Disciplinary inst. -> Control -> Postscript -> Modulation
- ☐ genAI as an institution
- ☐ But how does the AI relate to the subjectivisation etc.?

Deleuze's short and speculative essay, "Postscript on the Societies of Control" (1992), introduced a fragmentary but generative diagnosis of late-modern power structures. It describes a social assemblage with a specific machinery, an institutional framework constructed via the digital technologies, governed via the functionalities made available by computational advancements. A computational dispositif that do not merely maintain or articulate, but also shape the political, and economic architectonics; realising the shift into a more fluid and flexible operation of power. It is an uncanny tendency to associate the current in and around the novel computer technologies, it is reaching beyond the plain understanding of *cyberspace* or the *virtual*, we are dealing with artificial systems capable of meaning-making (see Dishon 2024; Kazakov 2025). AI systems are mediating how we structure the reality while expanding their existence both in terms of capability, their spectrum of actions and their presence in vastly differing areas of operation (see Kazakov 2025). The current AI scene is exposed to an accelarating domination by genAI models, and particularly LLMs operating on a strict paradigm of expansion and rapid scaling model features, increasing data ingestion, and allocation of more and more resources. It is a matter of debate if the rapid accelaration in every aspect of the AI development is also translated into the development towards a more sophisticated future with Artificial General Intelligence (AGI), singularity while we are still debating what human-like intelligence means, but nonetheless, these models are getting more capable, comprehensive and becoming more present by the day.

Kazakov (ibid.) situates the contemporary AI development within what he calls scalar Darwinism: a phase of development defined by relentless quantitative expansion rather than qualitative transforma-

tion. LLMs and other genAI systems advance primarily by scaling; more parameters, larger datasets, greater computational resources, without fundamental architectural innovation. This mode of growth reinforces existing capitalist logics, where data is treated as a resource to be extracted and leveraged, and market advantage derives from size rather than capability. As in Michel Foucault's (see 2008, p. 131) definition of neoliberal governmentality whereas *the market* claimed to be the methaphysical plane producing the best possible solutions for diverse issues without any *directional* intervention, in the age of rapid AI development the scaling of the data and models claimed to be paving the ways to the ultimate solutions humanity direly in need of [1].

Whether the truth abouth our world is contained in the data waiting for the right model to extract the correct distribution, the genAI models have already shown surprising capabilities. While, especially LLMs often defined as simple next word predictors (see e.g. Dalvi 2025), or plagiarism machines (see e.g. Chomsky, Roberts, and Watumull 2023) genAI models continue to show surprisingly well performance in different areas of expertise (see e.g. Sultanow et al. 2024). These surprising performances are not simply technical feats; they signal a qualitative shift in how artificial systems engage with reality. Modern LLMs and other genAI models no longer operate solely as analytic tools that classify or retrieve as their previous AI model counterparts; they actively participate in the production and reproduction of reality by governing the flow of information, assisting various purposes of knowledge creation, generating outputs that are treated as meaningful, actionable, and often authoritative (see e.g. Dishon 2024; Montanari 2025). While their presence is causing a profound transformation in the ecology of information and the mediation of human knowledge, they structure what is visible and legible, filter which forms of knowledge circulate, and frame how subjects encounter and relate to information. In this sense, the processes of subjectivisation today is tightly bound to this new mode of algorithmic meaning production. Other modes of information mediation is increasingly replaced by the genAI models, To understand the societal and political implications of genAI, we must first interrogate its substance of subjectivisation of these meaning making entities as distributed institutions over information.

[1] Not always this obvious, but the technolsolutionist propagation is especially strong in tech. development fronts:

— Musk 2025

TODO:

- ☐ Last part is quite weak, update
- ☐ Write a better transition

3.1 *Subjectivity*

> Knowledge is not innocent. It is not so much about discovering the truth, but rather about producing certain truths; it produces objects, or subjects, like the delinquent, and thus spaces and strategies of intervention. The individual, as Foucault has it, became knowable and thus accessible to power.
>
> — Krasmann 2017, p. 11

CCC

The question of subjectivity; its emergence, its production has long haunted Western thought, morphing with shifts in different branches of philosophy like epistemology, and metaphysics. Its philosophical genealogy stretches back to ancient concerns with soul and selfhood (see e.g. Aristotle's *De anima (On the Soul)* 1986), but its modern formulation takes decisive shape with Descartes's cogito, which installs the thinking subject as the indubitable ground of knowledge (Descartes 2008). From there, Kant's *Copernican revolution* redirects philosophical inquiry toward the a priori structures of the subject that condition possible experience (Kant 2009). These moves solidified the notion of the autonomous subject as the bedrock of Enlightenment thought, closely linked to emerging political imaginaries of agency, reason, and rights (Taylor 1989). Yet even in these rationalist formulations, tensions persist around the relationship between the interiority of the subject and its formation through language, culture, and material institutions.

It is these tensions that would later be unraveled by structuralism and post-structuralism. Structuralists such as Lévi-Strauss, Saussure, and Althusser shifted attention from interior experience to the impersonal systems; language, myth, ideology that precede and produce subjectivity (Althusser 1977; Lévi-Strauss and Lévi-Strauss 1963; Saussure, Baskin, et al. 2011). The subject, in this sense, becomes an effect of signifying structures; it is interpolated by ideological apparatuses and made legible within symbolic orders. Post-structuralist thought does not simply reverse this move but radicalizes it: in the theories of Barthes, Derrida, and Kristeva foreground the instability, iterability, and difference at the heart of these structures themselves (Barthes and S. Heath 1977; Derrida and Spivak 2016; Kristeva et al. 1980). In post-structuralism, the autonomous subject of modernity dissolves into the relational field of discourse and social practice. What appears as "subjectivity" is only an instance, a provisional effect emerging from the entangled formations of language, power, and social structures. In this sense, there is no pre-given subject behind the text; the subject exists only as a position produced within and by the text of the social. Subjectivity, in this view, is neither natural nor given; it is produced, fractured, and dynamic. Institutions, in turn, do not merely constrain subjects but participate in their ongoing fabrication. This shift opens the way for analyzing how subjects are continuously assembled across milieus like for example linguistic, institutional, technological surfaces [2]

Post-structuralist accounts of power challenge the assumption that subjectivity is an original or pre-social essence. Instead, they situate the subject as a product of institutionalised practices that operate across everyday life. From the school to the prison, the factory to the family, institutions do not simply govern behaviour; they organise routines, structure spaces, assign functions, and produce bodies

[2] **NOTE**: Not quite sure about this prologue, consider removing.

capable of carrying out specified functions. To become a subject is to be made visible and intelligible; seen, heard, corrected, and positioned within patterned social routines. These processes are not merely ideological; they are materially embedded in environments that stabilise what kinds of subject-positions are possible. Spatial arrangements such as the classroom desk, the assembly line, or the domestic threshold do not simply house activity; they condition the very terms under which individuals come to know themselves and others. What emerges is a subjectivity that is not interior or fixed, but modulated through institutional dispositifs that shape the conduct, perception, and agency of the self over time (Foucault 1995; Hardt 1998).

In Foucault's description of *careful fabrication of the subjectivities* (Foucault 1995, p. 215) the institutions were formed as efficient enclosures operating on surveillance machinery to engrave specific forms of subjectivities on bodies (see e.g. Foucault 1982, p. 783). This forging process on bodies in disciplinary societies was allowing to delegate the correction of behaviour to the individuals themselves through being exposed to a *constant state of conscious and permanent visibility* (Foucault 1995, pp. 202–203). The disciplinary machinery appears as an effective and economic machinery in comparison with the operation of power in the previous *sovereign societies* whereas the power was imposed on the bodies through public spectacle, punishment, and sovereign's right over subjects' lives; with Deleuze's words *to tax rather than to organize production, to rule on death rather than to administer life*(1992, p. 3). Discipline, the successor of the sovereign power, instead, inverts the gaze in the subjects, deploys the knowledge gathered over bodies and works as a *political technology of the body* (Foucault 1995, p. 26); it operates in a distributed manner, without forming monolithic centres emitting the power. Yet, constantly active in its diffuse structure, nests in societal functions and practices, manifested extended by its own subjects, a *microphysical manifestation of power* (see ibid., pp. 26–27).

In disciplinary societies, as Foucault famously outlined, power is exerted through the molding of individual bodies and behaviors via normative institutions (Foucault 2008). The logic was architectural and corporeal: subjects were shaped within bounded spaces, subjected to surveillance, and trained into docility through routines and assessment. Control societies, by contrast, operate through the flexible recomposition of dividual data traces,discrete fragments of identity, behavior, or preference,that can be extracted, modelled, and recombined by algorithmic infrastructures. This understanding frames the conceptual shift that underpins the present analysis: the transition from disciplinary societies to control societies. Both operate as mechanisms for the production of subjectivity, but they differ in how they structure space, time, and the flow of power.

3.2 *Postscript: Updating the Societies of Control*

TODO:

- ☐ Brusseau 2020; A. Galloway 2001; Hardt 1998; Hui 2015; Kazakov 2025; Mackenzie 2015

> You see control can never be a means to any practical end... It can never be a means to anything but more control... Like junk...
>
> — Burroughs 1979, p. 81

Following Michel Foucault's genealogy of power, from sovereign societies to disciplinary regimes, Gilles Deleuze introduces a third historical configuration: *the society of control* (Deleuze 1992) . Deleuze points to a crisis in disciplinary institutions and starts charting the replacement of enclosed institutional spaces; schools, factories, prisons, but also family[3] by diffuse, pervasive mechanisms of flexible forms of control. The analysis starts with a historical account to draw the difference between the discipinary and control societies, Deleuze notes that the 20. Century marks the transition form one society to another and the disciplinary institutions were already fading out after the WWII (Deleuze 1992, p. 3). While disciplinary regimes operated through enclosures, segregating individuals into clearly defined spaces associated with specific functions, contemporary forms of control rely on more fluid mechanisms; instead of physical boundaries, social organisation is achieved by tracking, directing, and modulating movement and behaviour across interconnected and permeable environments (see Brusseau 2020, p. 3). *The walls of the institutions are breaking down in such a way that their disciplinary logics do not become ineffective but are rather generalized in fluid forms across the social field* (Hardt 1998, p. 139).

[3] Which D&G were eager to emphasise its institutional nature of (see *Anti-Oedipus* 1983).

The shift from disciplinary societies to societies of control, as articulated by Deleuze (Deleuze 1995), marks a profound transformation in the mechanisms of governance and the infrastructures of power. In disciplinary societies, power functioned through discrete institutions; family, school, factory, and prison etc. each molding individuals within enclosed spaces, imposing fixed norms, and segmenting life into sequences of spatial and temporal regimes (Deleuze 1992, p. 4). Movement between institutions required the subject to start anew in each, as the machinery of discipline operated in parallel but remained compartmentalized. Control, a term Deleuze borrows from Burroughs (1979), signifies both this institutional shift and a fundamental change in the machinery of subjectivisation. A new post-disciplinary dispositif powered by the emerging forms of computational advancement whereas power becomes immanent to processes of circulation; it no longer functions through strict enclosure but through the real-time coding and recoding of flows of information, behavior, and subjectivity. *Modulation*, a concept Deleuze introduces instead of the molds that characterise disciplinary institutions, is central to this transition. Disciplinary *molding* imposes form from fixed sites, *modulation* is a flexible, dynamic regime that enacts control through adaptation via continuous feedback; a shift from a *form-imposing mode to a self-regulating mode* (Hui 2015, p. 74) in the production of subjectivity. In this sense, control is not simply the

replacement of discipline but its generalization and internalization within the continuous processes that traverse the social field.

The unitary subject operating in different designated roles in disciplinary institutions is going through a different kind of splitting this time. The self in control societies is unified across the smooth surface of control society but the individual's qualities and behaviour is getting analysed an and acted upon in smaller separated parts (see MacKenzie and Porter 2021, p. 5) . The individual is broken down, fragmented into dividuals; data particles, micro-traces, partial qualities circulating across digital and organizational infrastructures (Deleuze 1992), this *bundle of elements held together* replaces the unitary subject (MacKenzie and Porter 2021, p. 6). [4] Through databases, ubiquitous computation, and advanced statistical inference, these dividual traces are parsed, recomposed, and acted upon, generating personalised evaluations, outputs, and interventions. The effect is akin to a self-deforming cast, continuously adjusting to the subject in motion (ibid., p. 4).

Docility under this new regime is no longer enforced by explicit institutional intentionality, operationalized as rigid codes. Instead, control operates by creating spaces that feel open and permissive, as if the individual were free to explore, create, and tangle with possibilities. Yet both their production and its ends are subtly governed by intangible, underlying forces (Hui 2015, p. 75). In this sense, control converges the previously separate spaces of subjectivation into a single, fluid field: one no longer leaves an institution behind, and one is never fully done with the spaces that act upon the self (Deleuze 1992, p. 6). In a dividualised society, the masses are not analysed as collection of individuals anymore, partial dividual characteristics open trajectories to analyse collevtive behaviour on; contrary to the disciplinary institutions segmenting individuals and populations, control society seperates componenets of individuality (MacKenzie and Porter 2021, p. 9) . The variables are biometric information, social media history, purchasing tendencies, political leaning, voting patterns, and other elements of subjectivity, a new way of grouping working in the logic of marketing (see Deleuze 1992, p. 7).

Control does not abolish discipline; its difference is bound in intesification and hypereffectivity. As Hardt and Negri (2003) observe, *the passage to the society of control does not in any way mean the end of discipline. In fact, the immanent exercise of discipline [...] is extended even more generally in the society of control* (A. Galloway 2001, p. 83). This immanence is central: control is no longer imposed from above but embedded within the continuous flows of communication, code, and affect. It operates through protocols, feedback loops, and algorithmic infrastructures, an open-ended regime of governance that infiltrates the very capacities of subjects to act, perceive, and desire [5]. Yet while Foucault never postulated a stage beyond disciplinary societies, Deleuze's Postscript on the Societies of Control offers only a sparse sketch of what comes after enclosure-based institutionalisa-

[4] Control and its modulating nature are not necessarily digital, the terms are referring to a specific operation that became the new paradigm of capitalism. Deleuze finds a characterisation of this phenomenon in the transition from the factory to corporation structure, where the site of production is replaced by an abstract field of work with rapid changes in definition of the salaries, and the work itself surrounded by a *spirit* of the work place. Similarly the never ending education is another indicator Deleuze mentions, the education is not coming to an end in the enclosure of scholl but transendences into all branches of life (see Deleuze 1992, p. 6). Nonetheless, the modulation is characterized by the machinery that enables discpline to transform itself to this fluid form of control, hence the emphasis on the novel operation and digital aspects of control.

[5]

- ☐ use Cheney-Lippold 2024 to talk about dividuals
- ☐ Reconsider

tion. As Hardt (1998, p. 139) points out, Deleuze says remarkably little about the institutional architecture of control societies themselves; the form remains vague, its contours merely suggested through keywords like *modulation* or *dividuation*. We are being exposed to a fundamental change in the training of subjectivity, but the implications of the new operational novelty, as well as, it is new forms and machinery are only partly explored. Although, we are having a glimpse into a world governed by code and the *spirit* of corporation, the question about the operative principles especially regarding the emerging novel machineries remains to be explored.

3.3 *AI as Institutional Framework*

Information technologies and computation seem to be the pivotal elements in this new turn of capitalism Deleuze elaborates on. Advancements in these areas brought new definitions to surveillance, data collection and processing, signalize new forms; Krasmann (2017) directly associates these novelties with subjectivization:

> Power brings the subject into being, but power does not exist independent of its enactment. It is immanent and only takes shape at a point of resistance. The subject is such a point of resistance that recasts, redirects and sometimes reverts power. Subjectivation, however, always involves wrestling with oneself; it is governing the self and self-government: the subject is bound to power as it is to him- or herself. How then to conceive of a political subject as a fold of power as well as a "line of flight"? How to imagine a challenge to the current regime of visibility?
>
> — ibid., p. 18

TODO:

☐ Exactly the following paragraph is to be updated

How can we operationalise the institutional analysis of the control societies? How can we expand the exploration to the AI models governing and creating or patch-working information? There are no shortage of adaptations of Deleuze's short account to contemporary advancements. Brusseau extends these debates into the era of big data and predictive analytics, where the logic of control intensifies through hyper-personalized algorithmic environments (Brusseau 2020). Predictive technologies, entwined with generative AI systems, instantiate what Hui terms *disindividuation*: the fracturing of subjectivity into calculable, governable fragments, and Hardt (Hardt 1998) already associates *the Empire* with these new fluid and subtle forms of control.

Yet, as MacKenzie and Porter emphasize, such developments also provoke new modalities of critique. Their notion of *counter-sequencing*, the rearrangement or disruption of institutional sequences,suggests avenues for resistance that do not rely on outdated ideals of autonomous subjectivity but engage the very logics of dividuation and modulation from within.

The imaginary of the computational future had a discrete nature in literature, the algorithmic governance was much more about blocking flows, denying access and keeping boundaries intact. The central question Deleuze asks is *how can there be control if nothing is forbidden?* (Brusseau 2020, p. 2). The answer to the question with the predictive analytics; data-driven marketing and social media strategies that regulate through incentives, soft control over the flow of consumers, recommendation systems, filters, and relevance associations; not necessarily a containment, no blockage, no enforcement, but correction through personal information, profiling, anticipation (see ibid., p. 2). Now, we have another medium to analyse in the same manner, one that is able to talk back.

While Deleuze's account remains foundational, its brevity has spurred diverse and sometimes conflicting interpretations. As MacKenzie and Porter (2021) observe, much of the subsequent literature has overemphasized technological dimensions, portraying control as an exclusively computational or algorithmic phenomenon, detached from institutional life.

Yet, they argue, institutions have not disappeared; rather, they have been transformed into totalizing structures that sequence and redistribute dividuals across domains. This process of sequencing constitutes a key mechanism by which control operates in contemporary society, bridging the technological and institutional logics.

3.4 *Connection to AI*

In modulation, power is no longer exercised through strict categories or final forms but through elastic processes that track, nudge, and reshape behavior in real time. This logic is foundational to the algorithmic infrastructures of contemporary societies, especially those driven by genAI, where subjectivation occurs not through fixed norms but through continuous calibration against probabilistic expectations. In this regime, what is governed is not the subject as a stable identity, but the flow of tendencies, preferences, and predictions.

While Deleuze's Postscript emphasizes the change in the institutional formation from the disciplinary societies, it does at the same time point out a specific form of de-institutionlaisation (ibid., p. 15). Emerging technological forms of governance should be in this case the agents of this dissolving operation of institutions.

3.5 *Turn in (Cognitive) Capitalism and its institutions*

With "Postscript on the Societies of Control", Deleuze is not (just) trying to define the framework of a series of computational advance-

ments, he is referring to a new turn in capitalism. One that immediately to be observed also by the methodology and dispositifs it came to life with. In the post-structural analysis of powerm, we come back looking into apparatuses and finding the parts or a complete miniature of the operation of the power

GenAI systems can be read as a part of this shift. Their architectures do not discipline a subject within the context of an enclosure; they are not necessarily designed to achieve a specific inscription. However, they have the capability to modulate meaning, affect, and behavior by operating upon statistical representations of language, vision, and interaction. In place of rules or norms, genAI systems govern through probabilistic inference: they do not enforce a fixed logic, but generate outputs that are dynamically aligned with the distributional patterns of their training data. And they are going through processes of fine-tuning (see Section 4.4.1 for a reflection) to adjust models' behaviour to some degree. This represents arguably a post-disciplinary mechanism of control,one that governs not by exclusion or correction, but by continuous recalibration.

TODO

- ☐ Mention symbolic, non-symbolic AI
- ☐ Some weak lines above

3.6 *Capitalism and Schizophrenia*

TODO: Title

- ☐ Consider opening a discussion between the postscript and D&G's project

3.7 *Krassman Quotes: Algorithm & Control*

> The digital subject, at first sight, is a fictive subject, both in that it is about doubling reality in "data doubles"23 – as Deleuze observes: language becomes "numerical", individuals morph into "dividuals" and masses into "samples, data, markets"24 – and in that the individual is no longer of primary interest in those procedures of data production. Instead, patterns of behavior and the movements of data are gathered to predict and shape future possibilities. There are criminal ambitions to be anticipated and forestalled but also consumer desires to be addressed and invoked.
>
> ...
>
> Algorithms do not simply apply norms, but generate new norms of suspicion.26 They present results we did not reckon with and could not anticipate. They help us to envision the unimaginable and perhaps to preempt the incalculable.
>
> ...
>
> Power is no longer merely inscribed into the environment, the architecture, the order of light, as was the case with the Panopticon. Rather, the environment itself, the algorithms, appear to be the source of power, as they are able to process data and produce information.
>
> ...
>
> They thereby produce their own truth effects. Rather than predict truthful probabilities, algorithms preempt reality. Confronting us with our desires and aspirations, they always already seem to know our wishes – precisely because drawing on a seemingly incomprehensible amount of disparate data. There is no representation and no simulation of the world, as what could have been said seems to have always

> already been said: there is no possibility for difference to emerge,37 and in this sense, no space for the political to be challenged.
>
> — Krasmann 2017, pp. 15–16

3.8 *GenAI Modulation*

TODO: Enter Foucault -> Deleuze, Societies of Control

- ☑ Foucault
- ☐ Societies of Control
- ☐ Control in Burroughs

Amoore (2024), for example, argues that this modulation is not neutral, the generative capacity of these systems is embedded in a *governing rationality* (ibid.), one that renders plausible what counts as intelligible, actionable, or true. By learning and operationalizing joint probability distributions across vast corpora, genAI systems instantiate regimes of verisimilitude,offering outputs that appear coherent not because they adhere to a symbolic rule set, but because they resonate statistically. In doing so, they encode a specific politics of what can be thought, said, or imagined.

Whereas disciplinary power sought to impose order through hierarchies and segmentation, control operates by managing flows. In the case of genAI, this entails the modulation of user input, system response, and contextual adaptation in a closed feedback loop. Each prompt, response, and correction contributes to the model's ongoing refinement, a continuous, real-time inscription of preferences and expectations into the probabilistic substrate of the system.

GenAI models therefore represent a paradigmatic case of modulation-as-governance. Their architecture is not only technical, but institutional: a site where subjectivity is shaped not through fixed norms, but through dynamic adaptation. They do not dictate, but suggest; they do not enforce, but align. Yet in this very flexibility lies a form of power that is more pervasive and less accountable than disciplinary mechanisms,one that operates in the folds of everyday interaction, shaping sense before critique can even begin.

3.9 *Where are the Lines of Flight*

Mackenzie and Porter (2021) suggest that we analyse this new front on computational novelties and AI as de-institutionalised new era. I would like to

1. See if there is an institutional formation in specifically genAI models
2. And how the resistance in such kind of an either de-institutionalised or re-institutionalised form of soft control plane looks like.

4

AI as the Infrastructure of Modulation

> [...] descending into the hidden abode of production means something else in the digital age. It means that we must also descend into the somewhat immaterial technology of modern-day computing, and examine the formal qualities of the machines that constitute the factory loom and industrial Colossus of our age. The factory was modernity's site of production. The "non-place" of Empire refuses such an easy localization. For Empire, we must descend instead into the distributed networks, the programming languages, the computer protocols, and other digital technologies that have transformed twenty-first-century production into a vital mass of immaterial flows and instantaneoustransactions. Indeed, we must read the never ending stream of computer code as we read any text (the former having yet to achieve recognition as a "natural language"), decoding its structure of control as we would a film or novel.
>
> A. Galloway 2001, p. 82

Contemporary Artificial Intelligence (AI), and particularly Generative Artificial Intelligence (genAI) and Large Language Models (LLMs) operate through architectures that capture and recombine traces of language while parsing the collection of human knowledge. If control societies are defined by infrastructures that capture, modulate, and recombine traces of life, contemporary AI embodies this logic in its technical architecture. To understand how these systems participate in the production of subjectivity, we must first trace the historical and technical evolution of AI, from symbolic reasoning to statistical modeling and self-attentive transformers.

The current chapter provides the technical foundation for the subsequent political analysis of generative AI. To properly understand the potentially institutional role of genAI models, and LLMs in par-

ticular it is necessary to first trace their historical development, underlying architectures, and operational logics. While this section remains on a primarily technical level, it also gestures towards the epistemological and political stakes that will be discussed later. It begins by outlining the historical trajectory of artificial intelligence research, distinguishing between the early, symbolic paradigm (Symbolic Artificial Intelligence (symAI)) and the contemporary, statistical approaches that characterize Deep Learning (DL) and genAI models. This includes an explanation of neural networks, self-supervised learning, and the rise of transformer architectures as the technical backbone of modern LLMs.

Beyond mere description, this technical overview serves a strategic purpose: it demonstrates how AI, even at the level of architecture, already embeds specific logics of inference, representation, and control. These are not neutral technical details, but the material conditions that enable AI systems to operate as infrastructures of knowledge production, decision-making, and ultimately, governance. The subsequent sections therefore provide both the necessary technical background and the conceptual scaffolding for the analysis of generative AI as a distributed, non-symbolic institution of power.

This part can be a good addition to the "state of art"

- Montanari 2025 is a good source for a brief techno-political history and genealogy of llms
- Also here https://www.technologyreview.com/2024/07/10/is-artificial-intelligence-ai-definitive-guide/ and here Pasquinelli 2023

4.1 *From Symbolic Rules to Statistical Modulation: A Brief History of AI and Natural Language Processing (NLP)*

NLP is an area that lies at the intersection of linguistics, computer science, and AI, aiming to create computational systems that can interpret and handle human language data. Considering that, in some respects, the cognitive performance of an individual human is hardly superior to that of other primates (Manning 2022, p. 127), it is hardly surprising that breakthroughs in AI have been driven by NLP. Language, more than individual brainpower, constitutes the machinery through which human intelligence scales, distributes, and accumulates collectively (ibid., p. 127), and has been the ground most of the breakthroughs in AI development, especially, in the recent years (see Bommasani et al. 2022, 22ff for a detailed analysis of the history of AI and language).

TODO: Title

☐ A little more articulation is needed around here.

Artificial intelligence emerged in the mid-20th century, grounded in the formal logics of symbolic representation. The foundational paradigm, now referred to as symAI or Good old-fashioned AI (GOFAI) , conceived intelligence as a matter of symbolic reasoning over explicitly encoded rules. The early paradigm treated intelligence as a computational process operating over discrete symbols according to explicitly programmed rules. AI systems under this logic were built to emulate deductive reasoning and problem-solving. The assumption was clear: if the world could be faithfully translated into a logical schema, machines could infer, deduce, and act rationally (see Eloff 2021, p. 183).

Manning (2022) defines the **first era** between 1950 to 1969 as a development process under the immense lack of the knowledge about structure of human language or Machine Learning (ML) and AI. The 1956 Dartmouth Conference institutionalized the ambitions by defining AI as "the science and engineering of making intelligent machines" (Montanari 2025, p. 195). Early research during this period was primarily focused on narrow, rule-based systems, particularly word-level translation lookups and simple mechanisms to handle inflectional forms and word order (Manning 2022, p. 128). In parallel, Alan Turing made substantial contributions by introducing the famous "Turing Test" (or "Imitation Game"), designed to evaluate a machine's ability to imitate human intelligence and rationality, along with the foundational concept of a universal machine (see Montanari 2025, p. 196). As Cognitive Robotics Prof. Murray Shanahan and Meta's Chief AI scientist Yann LeCun emphasize, the *Turing Test* is an inadequate benchmark for assessing modern AI models (Google DeepMind 2025; Lex Fridman 2024), but Turing's ideas nonetheless contributed to the conceptual foundation of the *prompt-based conversational machine* (Montanari 2025, p. 196). Aligned with Turing's perspective, the underlying notion in the early imaginary of a future AI was simple; if a machine could convincingly imitate a human in conversation, it was considered intelligent.

Relying on handcrafted rule sets included implicit definitions of the features regarding the object of interest; to recognise patterns the digit six in an image for instance, one might encode the features "a closed loop at the bottom" and "a curve rising to the right". Such symbolic heuristics were sufficient so long as the data was clean and the context unambiguous. In the **second era** of AI development, spanning roughly 1970 to 1992, these approaches were extended to more complex domains, most notably natural language. By attempting to formalize aspects of linguistic structure and meaning, researchers pushed the boundaries of rule-based systems. While these models demonstrated greater sophistication in handling linguistic patterns, they still relied on explicitly encoded knowledge and remained limited by the inherent rigidity of symbolic architectures (Manning 2022, p. 129). Yet, these systems could not generalize beyond predefined rules. When confronted with noise or shifting contexts, their logic collapsed. The result was a period of stagnation and disillusionment now remembered as the "AI Winters" between 1970 - 1980 (Eloff 2021, p. 183). [1]

But real-world ambiguity proved hostile to symbolic systems. As symAI attempted to scale into more complex domains like vision or language, it revealed its brittleness (Eloff 2021, pp. 183–184). Philosophers of phenomenology were early critics of this paradigm following Hubert Dreyfus' (2009[2]) earlier work where he argued that human intelligence was not symbolic, but embodied, situated, and fundamentally non-representational. Despite such critiques, symAI dominated the earlier decades of research in AI fields. This ratio-

[1] **TODO:** A more nuanced explanation regarding the transformation is needed, e.g. Pasquinelli 2023.

[2] Originally published in 1972.

nalist framework aligned with early cognitive science's attempts to model the mind as a rule-based machine of symbolic representation (see Montanari 2025, pp. 194–197). Gilles Deleuze & Felix Guattari (D&G) were also one of the ciritiques, the hierarchically structured learning and the projection of a central pattern was clearly not working well:

> This is evident in current problems in information science and computer science, which still cling to the oldest modes of thought in that they grant all power to a memory or central organ. Pierre Rosenstiehl and Jean Petitot, in a fine article denouncing "the imagery of command trees" (centered systems or hierarchical structures), note that "accepting the primacy of hierarchical structures amounts to giving arborescent structures privileged status.... The arborescent form admits of topological explanation.... In a hierarchical system, an individual has only one active neighbor, his or her hierarchical superior.... The channels of transmission are preestablished: the arborescent system preexists the individual, who is integrated into it at an allotted place" (signifiance and subjectification).
>
> — Deleuze and Guattari 1987, p. 16

D&G's critique of early AI approaches centred on their rejection of hierarchical and centralised models which constitute one of the main pillars of their project. Their affirmative alternative was grounded in a connectionist and non-hierarchical understanding of thought, formalised through the concept of the *rhizome* (ibid., 3ff.) as in opposition to *tree* structures. While their critique targeted the symbolic, rule-based systems of their time, it is striking how closely their vision anticipated the architectural principles underpinning contemporary AI on a general level, particularly in its distributed, associative, and layered formations [3] . Nonetheless, the technological trajectory toward such architectures would take decades to materialise, revealing the prescient force of their philosophical intervention.

[3] Arguably, also regarding non-hierarchical functioning of the Artificial Neural Networks (NNs); however, it is still a matter of discussion, if the genAI model architectures deploy a contuinuous subordination between different patterns and distributions. See the following sections for further articulations.

The **third era**, from roughly 1993 to 2012, was signified with the beginning of the abundance any novel AI innovation lacked the most, *the data*. As the internet boom suddenly introduced a massive digital corpora, researchers shifted toward statistical learning, leading to the rise of data-driven NLP. This shift replaced hand-coded rules with empirical models trained on annotated examples (Maas 2023); models could now generalize from data rather than deduce from explicitly defined axioms. Initially, the dominant approach centered on relatively simple statistical techniques applied to modest amounts of text, often in the low tens of millions of words. Researchers extracted linguistic facts from these corpora, identifying regularities such as common collocations or syntactic structures. Yet, early attempts to model language understanding through these means remained limited in their ability to capture deeper semantic or contextual knowledge (see Manning 2022, p. 129). For instance, early statistical models revealed that certain types of words tended to appear together, names of places often occurred alongside personal references, while more abstract terms exhibited distinctive distributional

patterns. However, such surface-level regularities provided only limited insight into the deeper structures of language. As it became evident that simple frequency-based methods were insufficient for capturing the complexity of linguistic meaning, the focus shifted toward building annotated linguistic resources, such as syntactic treebanks, lexical databases, and labeled datasets for named entity recognition. These resources formed the foundation for more reliable, supervised learning approaches (see ibid., p. 129). Onwards, the general purpose AI development continued with ups and downs in activity, with a couple of earlier succesful neural network based aproaches like Mulloch-Pits. Among the early milestones was ELIZA, a rule-based program that mimicked a psychotherapist by matching keywords to scripted responses. Despite its simplicity, ELIZA gave the illusion of understanding and demonstrated the potential of machine conversation; though its developer emphasized it was merely parodic (Toloka 2023). Still, it signalled the beginning of natural language interaction with machines, laying groundwork that statistical and later neural methods would build upon. Up until around 1997 where much more advanced models like Deep Blue operating on more sophisticated architectures like the early attempts on Deep Artificial Neural Networks (DNNs) were developed (Montanari 2025, p. 197), but the main meta of the AI development was highly dependent on labeled data, and Supervised Learning (SL).

TODO: Title

- ☐ Add the BA history info here
- ☐ Explain SL

Although, the real transformation originally began in the early 2000s, the first significant fruits of the new direction dropped around 2013, which marks the **4. and current era** in AI development (Manning 2022, p. 129). Pushes through the ability to process more and more data allowed a new paradigm to emerge, rooted in NNs inspired by the architecture of the brain, *connectionism* became the new meta of further advancements. These systems, now more broadly applied and clearly defined as DNNs, learned not by logic but by adjusting distributed weightings across layered networks, which became the foundation for contemporary ML and DL systems. Exponential advances in computation enabled these networks to scale (Eloff 2021, p. 184) and finally also pushed towards an Unsupervised Learning (UL)[4] methodologies, whereas the models were geared towards to recognise patterns in the data without being explicitly told which features of the data were pointing to what. For instance, while early augmentational models were trying to distinct between cat and dog photos by looking at photos labeled by humans and other processes as either as *dogs* or *cats*, UL models are looking at a data collection of unlabeled photos and try to find patterns in them which makes both parties distinct through specific characteristics, in other words, towards finding out about the substance of *dogness* and *catness*. On the NLP fronts, linguistic units such as words or sentences came to be represented as vectors in high-dimensional vector spaces. Semantic and syntactic relationships were modeled not through rule-based analysis and pre-defined categories, but through the spatial prox-

[4] Should this one have its own section?

imity of these vectors (Manning 2022, p. 129). DL allowed to parse distant context, as well as, processing the words meaningwise close to each other thanks to this generalised vector space approach optimised with more and more textual data (see ibid., p. 129). This approach turned out to be far more effective than earlier attempts at formalizing linguistic meaning. Instead of hand-coding grammatical rules or manually annotating small corpora, models could now process large textual datasets and infer structure statistically. DL enabled systems to capture long-range dependencies in context and identify meaning-level relationships through learned representations optimized across massive datasets. Crucially, this reduced the need for manual labeling, as UL techniques became dominant.

One of the most significant turning points was around 2018 with the succesful implementation of Self-Supervised Learning (SSL) approach. SSL constitutes a special case of the UL which not only makes the models identify underlying structures in the data but also enables them to create their own training exercises through the prediction challenges they are subjected to (ibid., p. 129). This includes masking specific words in the text to try to predict correct or most fitting tokens, or try to guess the next word in an abruptly cut text, SSL models learn by predicting missing elements from within the input itself. This method allowed models to learn linguistic regularities from massive unlabeled corpora, and it gave rise to pre-trained genAIs (Maas 2023). The architecture that enabled this leap was the *transformer architecture*. Its core mechanism, self-attention, computes weighted dependencies between all tokens in a sequence, allowing the model to capture long-range relations independent of word order. This innovation enabled massive parallelization and scalability (ibid.). Availability of vast data and the unique novelty of transformer architecture that was powered by a huge amount of reinforcement capability through repetition has been crucial operating on SSL methodology to parse and accumulate huge amounts of unlabeled human language data. The transformer architecture is fundamental for the meaning-making capability of the genAI models, the following sections is therefore dedicated to focus on this specific innovation.

[5]

TODO: Title

- ☐ Better citations needed in this part
- ☐ We could also mention the Saussarean assumption here.
- ☐ Refer to Figure 1 in ibid.
- ☐ include the following here: A Language Model (LM), for instance, trained under this paradigm does not classify sentences into categories but instead learns to predict the most probable continuation of a sequence. In multimodal systems such as Text to Image Model (T2IM) or Multimodal Generative Model (MGM), this process involves inferring plausible image-text correspondences or interpolating visual representations from distributed patterns in training data.

[5] **TODO:** There needs to be a section about supervised -> unsupervised learning since unsupervised learning marks the rise of neoplatonic esoterism.

TODO: Title

- ☐ Update the following paragraph

A subcategory of the genAI, LLMs such as GPT-3 and GPT-4 are not task-specific in the traditional sense. Rather than being fine-tuned for each use case, they rely on "few-shot prompting": given a small set of examples at inference time, they condition their outputs without internal weight updates. Their knowledge is distributed across billions, sometimes trillions, of parameters trained to minimize prediction error. The shift from symAI to deep, generative architectures does not merely mark a technical transition. It signals a deeper epistemological break. LLMs do not "understand" language in any classical sense, they generate statistically likely continuations. Meaning is no longer rule-based; it is computed as vector proximity in high-dimensional space (Montanari 2025, p. 199). These networks are opaque, their training data culturally saturated, and their outputs probabilistic (Eloff 2021, p. 186). They do not interpret, they modulate. Rather than representing knowledge, they operationalize its prediction. In doing so, they establish a new infrastructure for language: distributed, non-symbolic, and non-transparent.
This transformation, from formal logic to differential modulation, sets the stage for understanding genAI not just as a technical system but as an institutional form—a mechanism of governance, sense-making, and subjectivation in contemporary control societies.

TODO: Title

- ☐ Advance on this and introduce a smooth transition
- ☐ use Toloka 2023 Development of generative AI section to advance further

4.2 *Algorithmic Governance of Information before GenAI*

> Thus far, we have determined that whereas the individual and disciplinary power seem to be cast in the same mold – the former being the product of the latter – the digital subject of the control society 2.0 appears to be an active subject able to make decisions – which in turn feeds the algorithms.
>
> — Krasmann 2017, p. 19

Following the historical development of the AI models is an algorithmic journey with a quite equivalent destination thinking about the definition of the control society. As Deleuze expected computational systems turn into, the use of the early NN based AI models was indeed focusing on the profiling and behaviour anticipation. Before the emergence of genAI, early models emerged in search engines, social media ranking algorithms, and recommendation services, which were primarily designed for **profiling and relevance association**(see Demir 2019, pp. 26–30). Their core logic followed a recurrent loop:

1. massive data collection from user interactions,
2. indexing and probabilistic categorization of behaviors,
3. ranking and recommending content based on **relevance association**,

4. generating personalized information flows, recommendations, associations,

5. feeding back the gathered information into the user's profile to update the personalised process (see Figure 4.2 for an illustration of the process).

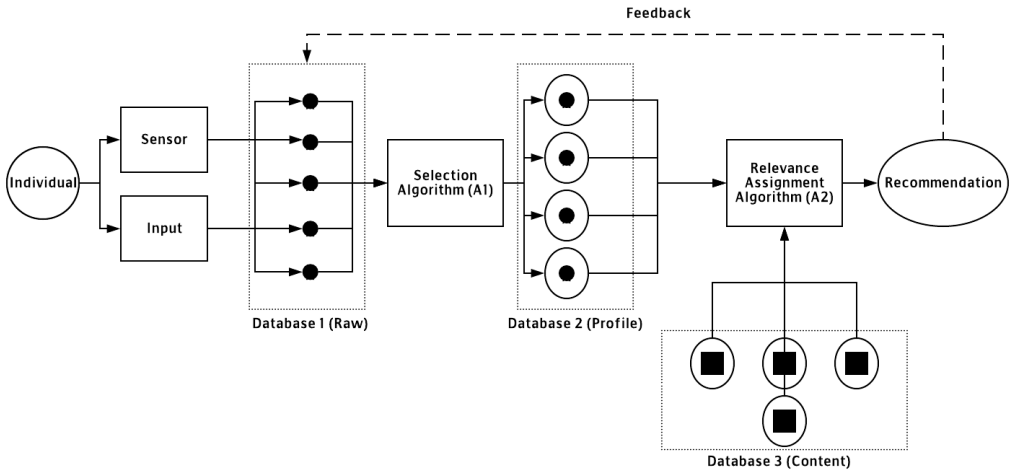

Figure 4.1: Algorithmic Selection and Relevance Assignment Process (cf. Just and Latzer 2017, p. 241)

This process exemplified an **anchoring and endless loops of feedback** (see, e.g., the characterisation of the idea as anchors and endless while loops in Demir 2019, pp. 34–35) of algorithmic governance: each interaction was an input to a probabilistic model, which in turn structured the horizon of the next interaction. Platforms such as Facebook or YouTube did not need to coerce users; they governed behavior through **environmental modulation**, subtly reinforcing predictable patterns of attention and engagement (ibid., pp. 29–32).

> Under the guise of being free and friendly to use, we can see in this example that the modulation of social relations can actually lead to what we have called 'disindividuation' [...] the attention of each social atom (or 'person') is sliced into ever smaller pieces and dispersed across networks via status updates, interactions, and advertisements. [...] The 'collective' on Facebook becomes a distraction, a cause of the dissolution of structures within individuals, but not a site of new modes of empowerment.
>
> — Hui 2015, p. 90

What these early systems achieved was a subtle but pervasive **disindividuation**: the coherence of personal or collective agency was fragmented across algorithmically defined micro-traces. This reflects Deleuze's idea of **modulation** before the full-fledged society of control; **docility** was not imposed through fixed institutional codes but through **continuous environmental nudging**. Users felt free to "scroll, explore, and connect," yet the outcomes of their activity were always prefigured by opaque, data-driven logics of recommendation.

From an infrastructural perspective, these algorithms already **governed information flows and digital subjectivity**, creating a precondition for the transition to **generative systems**. Whereas these early models merely filtered, ranked, and nudged, contemporary genAI systems will move beyond **governance of information** toward its **generation**, a shift that intensifies their role in shaping collective meaning and epistemic authority.

However, going back to the roots of the nature of *control*[6], were the institutional mediums of control ever meant to be also capable of generation? Were the computational methods of control society ever meant to be in communication with the individuals? The imaginary of the *control* in Burroughs' literary account was not meant to be bidirectional:

> The biocontrol apparatus is prototype of one-way telepathic control. The subject could be rendered susceptible to the transmitter by drugs or other processing without installing any apparatus. Ultimately the Senders will use telepathic transmitting exclusively... Ever dig the Mayan codices? I figure it like this: the priests – about one per cent of population – made with one-way telepathic broadcasts instructing the workers what to feel and when... A telepathic sender has to send all the time. He can never receive, because if he receives that means someone else has feelings of his own could louse up his continuity. The sender has to send all the time, but he can't ever recharge himself by contact. Sooner or later he's got no feelings to send. You can't have feelings alone. Not alone like the Sender is alone – and you dig there can only be one Sender at one place-time... Finally the screen goes dead... The Sender has turned into a huge centipede... So the workers come in on the beam and burn the centipede and elect a new Sender by consensus of the general will... The Mayans were limited by isolation... Now one Sender could control the planet... *You see control can never be a means to any practical end... It can never be a means to anything but more control... Like junk...*
>
> — Burroughs 1979, p. 81

The logic Burroughs intuited in his writing on control resonates with the infrastructures of early algorithmic governance. Recommendation engines and social media platforms enacted a form of modulation that was unidirectional and recursive in the sense of mediation, but in a constant feedback loop with the collected data, traces of activity, and content users generated: attention and behavior were captured, segmented, and looped back into the system without any real reciprocity. In Deleuze's sense, modulation adjusts continuously but always in the service of the system itself, a self-deforming cast that adapts the subject rather than negotiating with it.

At first glance, these early NN-driven platforms already align with the institutional description of control societies: they dissolve enclosures, act through environmental nudges, and convert subjects into streams of dividual traces, this was the paradigm of *algorithmic governance of information*. The rise of genAI introduces a possible inflection point. These models do not merely modulate existing flows of infor-

[6] **NOTE:** IMPORTANT PART but if this is the right place for this discussion, it has to be decided.

TODO:

- ☐ The last part was used before, reconsider
- ☐ REWRITE THE FOLLOWING

mation; they *generate* content, narratives, and knowledge structures that actively participate in the formation of subjectivity. In other words, the machinery of governance now doubles as a machinery of production. Whether this shift represents a continuation of the control logic or the emergence of a qualitatively new mode of operation are the following tasks crucial to approach to the nature of this constellation:

1. open the black box of genAI and its transformer-based architecture;
2. examine how these models mediate, and potentially modulate, human agency and the production of meaning.

4.3 *Transforming Attention: Infrastructure of Modulation in genAI Models*

TBD, a general case of the development of transformers (e.g. Vaswani et al. 2017)

Over the past decade, many influential neural network architectures for sequence modeling, particularly in machine translation and NLP, were built on Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs). These models typically followed an encoder–decoder design, where the encoder processed the input sequence into a continuous representation, and the decoder used this representation to generate the output sequence[7]. Despite their successes, these architectures faced a fundamental limitation: **locality**. RNNs processed tokens sequentially, passing information through hidden states that decayed over distance, which made capturing long-range dependencies difficult. CNNs, while more parallelizable, were constrained by kernel sizes and fixed receptive fields. Both designs struggled with tasks requiring global relational awareness of a sequence. The Transformer architecture (ibid.) emerged as a decisive break from these sequential bottlenecks. Dispensing with recurrence and localized convolution, it introduced *self-attention* as the central mechanism for computing contextual representations. In a single operation, every token in the input sequence attends to all others, producing weighted combinations of contextually relevant elements (ibid., p. 4). This structure integrates global information without distance penalties and lends itself to massive parallelization, a property foundational to contemporary LLMs.

In their groundbreaking paper "Attention Is All You Need", Vaswani et al. (ibid.) proposed a new architecture that preserved the encoder–decoder structure but eliminated reliance on recurrence and convolution. Instead, the Transformer model relied entirely on attention mechanisms, not as a supplementary feature, but as the

TODO: Title

- ☐ Consider to go deeper into the neural networks
- ☐ Or introduce a section called NEURON and go into the notion of signgular elements holding weights which are only meaningful in connection
- ☐ Or should we break down the history above into sections?

TODO:

- ☐ Visualisation needed
- ☐ Consider explaining the generative AI first, a general account maybe (Bommasani et al. 2022; OpenAI et al. 2024, a general account maybe (bommasani2022a, openai2024)
- ☐ Explain the layers in DL and how the transformer relates to these layers. You could also introduce the concept of *Folds* by Deleuze

[7] In its simplest form, the encoder takes an input sequence of symbols $(x_1, \ldots, x_n)$ and transforms them into a sequence of continuous vector representations $\mathbf{z} = (z_1, \ldots, z_n)$. These vectors encode the relevant information from the input. The decoder then generates an output sequence $(y_1, \ldots, y_m)$ one step at a time. It is *auto-regressive*, meaning it uses previously generated outputs (e.g. $y_1, y_2, \ldots$) as input when generating the next token. This setup allows the model to generate coherent and context-sensitive output, building each element of the sequence in a structured, history-aware manner (Vaswani et al. 2017, p. 2).

TODO:

- ☐ Explain that ther is a conntinuous anchoring relationship between the global and local representations.

foundation of both the encoder and the decoder (ibid., 1–2; see Figure 4.2 for an illustration). This architectural shift allowed for highly parallelized computation, better modelling of long-range dependencies, and significant improvements in scalability. The Transformer has since become the cornerstone of contemporary genAI, enabling many of the recent breakthroughs in large-scale language modelling and generative systems. The architecture is built from stacked encoder and decoder layers, each composed of multi-head self-attention and pointwise feed-forward networks. These attention heads act as differentiated channels through which the model modulates its internal representations, integrating multiple semantic and syntactic perspectives concurrently. Instead of treating tokens as isolated or sequential entities, attention turns the entire sequence into a site of mutual interaction, each token is redefined in relation to all others.By eliminating recurrence and convolution in favor of attention, the Transformer achieved two decisive outcomes: first, it enabled highly parallelized training on vast datasets; second, it allowed the model to capture long-range dependencies and complex contextual relations with unprecedented efficiency. These properties form the *technical substrate* upon which modern genAI and LLMs are built.

Conceptually, the Transformer establishes a *global field of relation*, where each token is encoded not in isolation or rigid sequence, but through its distributed relevance to all others. Tokens are getting embedded into high-dimensional *vector spaces*, where semantic and syntactic relationships are captured as measurable distances and directions. The architecture affords a form of synchronic awareness: the presence of every other word is embedded within the representation of each word. The high-dimensional *feature space* acquired from this operation is storing every token with speicific distances from each other, where *relevant*[8] tokens where calculations like $king - man + woman = queen$ are roughly possible (ai-inquiry 2025a). This reconfiguration of relationality underpins the efficiency, scalability, and generative fluency that define modern LLM [9] systems. *Attention mechanisms* is mainly there to improve the interaction between input and output allowing the model to dynamically focus on the most relevant parts of the input sequence while generating each token. Attention computes a set of weights over the input representations, effectively answering the question: *"Which parts of the input matter most for predicting the next output?"*. Technically, self-attention calculates relationships between tokens by projecting them into *query*, *key*, and *value* vectors. These are used to compute attention weights through dot-product similarity and softmax normalization. Each token's final representation is thus a weighted blend of all other tokens, adjusted by their contextual relevance. Through multiple stacked layers and attention heads, the Transformer builds increasingly abstract representations, capturing both syntactic structure and semantic context. [10] Probabilistic modeling then governs

[8] Whichever association the model is building between different words. For example, often *king* and *man*, as well as, *queen* and *woman* are distance-wise relatively *close to each other.*

[9] Although we are focusing specifically on LLMs here, the transformer architecture is also built-in in genAI models like *text-to-image* generators.

[10] **NOTE:**This seems out of place

how the network moves through these spaces to predict the next output (Montanari 2025, p. 198). In this framework, meaning emerges through probability distributions, as the model computes which trajectories through the space of vectors are most likely given its training data. This spatial-probabilistic foundation is what later allows generative AI systems to produce sequences that appear coherent and meaningful.

THIS? This mechanism realizes a form of distributed modulation, aligning closely with Deleuze and Guattari's notion of transversal flows and differential coupling.
Crucially, the Transformer's design reflects a deeper infrastructural logic: by replacing the recursive memory of RNNs with direct positional encoding, it reterritorializes meaning into a high-dimensional vector space where proximity, not order, governs semantic influence. This change is not merely technical—it reconfigures the conditions of linguistic operability in generative systems, establishing an infrastructure where modulation becomes the grammar of sense-production.

Maas (2023) associates the novel operation structure of the transformer architecture with Derrida's concept of *trace* (see e.g. Derrida and Bass 1998, p. 26). Derrida's concept is an advancement on Ferdinand Saussure's linguistic theory operating on *signifiers* and *signifieds* (see e.g. 2007) through his own concept of *différance* whereas the emphasis shifts rather onto the context-dependency between words and the differention between each other. For instance, the color *red* is rather defined through its differentiation from *green* and *blue* without having an actual substance on its own (Maas 2023, p. 9). *The sign has no component that belongs to itself only; it is merely a collection of the traces of every other sign running through it* (Cilliers 2002, p. 44), all sings are in continuous relationship with other signs where the position of the word and the current network of all the connected signs, their differences[11] to that specific sign establish its substance. Yet, the substance or *the meaning* of the sign has temporal dependency because the specific arrangement of words, as well as the differentiation between them, is in constant flux, *in a dynamic process of combination and referencing* (ibid., p. 44) dependent on the current context[12]. Similarly in the operation of the LLMs , this spectral interdependence, where tokens are mutually inscribed into one another, suggests a structure in which meaning is always already haunted by the rest of the utterance (see Maas 2023, p. 12) in the sense of Derrida's *trace*. The *meaning* of the words in the LLMs is defined by the instance of different distributions; the distribution in the sentence, the distribution of the language in the model rendered through the whole dataset, and other dynamic mechanism regulated by the transformer core.

Montanari (2025) builds a direct associations between the cognitive functions mimicked by transformer artchitecture and the cultural im-

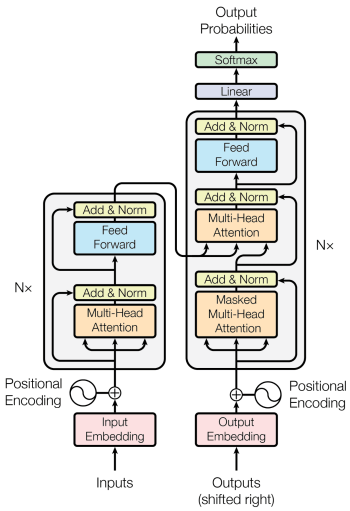

Figure 4.2: The Transformer Architecture with built-in Multi-Head Attention Mechanism in Encoder and Decoder Processes (cf. Vaswani et al. 2017, p. 3)

TODO:

- ☐ too much redundancy

TODO: Title

- ☐ Introduce an explanation of token somewhere

TODO: Title

- ☐ Fit this in: Attention layers follow on phenomenological process of focusing/reflecting on experience and aiming to amplify the distinct features of the process (see Beckmann, Köstner, and Hipólito 2023, p. 420)
- ☐ A clear definition of the Transformer is missing

[11] In terms of word-embeddings in the LLMs, we can interpret these differences as distances since distances on the network is the model's own way of representing differences between concepts.

[12] I'll be referencing this temporal formation as an *instance* from now on, since the context dependency of the network narrows the meaning into one instance of the connectional structure.

plications of the LLMs:

> [...] Transformer models, which exemplify the interplay between metaphor and function. Transformers [...] simulate certain structures and functions of the human brain, excelling at processing sequential data such as words in a sentence or notes in a melody. The transformative innovation within Transformers is the "attention mechanism," which enables the model to focus selectively on the most relevant parts of the input sequence. This mechanism is pivotal for discerning complex relationships and dependencies within data. [...] multi-head attention mechanism, a key feature that captures diverse aspects of an input sequence simultaneously. This dual role of technical objects – functionally specific and mythically resonant – reveals their broader cultural impact. Technical metaphors, often catachrestic and hybridized, solidify not only the utility but also the mystique and credibility of AI systems.
>
> — (ibid., p. 206)

TODO: Title

☐ The whole attention mechanism has to be rewritten

This architectural establishment can be read through the logic of **double articulation** (see ai-inquiry 2025b). The Transformer operates simultaneously on two strata: a molecular level of local attention scores and parameter updates, and a molar level of structured linguistic understanding. Each token's representation is formed through dynamic micro-adjustments, distributed flows of relevance, not unlike the first articulation of matter into expressive form. This is then consolidated across layers into coherent linguistic function, the second articulation of those forms into stable semantic structures. The Transformer thus embodies the double articulation of machinic sense-making. Attention mechanisms enact selective intensities across this field. Rather than representing fixed symbols, the Transformer's architecture instantiates meaning as a function of weighted relationality. These differential proximities constitute a *diagrammatic space*, where meaning emerges not from rules but from patterns of modulation. It is here that Deleuze and Guattari's distinction between molar and molecular formations becomes productive: the Transformer is not a symbolic machine but a machinic assemblage that captures both distributed flows and structured outputs simultaneously (see ibid.). On the one side, the meaning is fluid and in constant adjustment, on the other side the whole fluidity is strongly bound with the molar, overarching distribution extracted from the whole dataset. Attention weights instantiate selective intensities between elements, constituting a diagrammatic field of relations. In this field, meaning is not fixed or rule-governed; it is a function of differential proximity and relational salience. The Transformer thus encodes a new mode of learning: not inference from rules, but modulation of difference through weighted connection.

This transformation prepares the ground for the following sections, which analyze core Transformer mechanisms; attention, gradient descent, backpropagation, not merely as computational techniques but as micro-political operations that govern the production of meaning and subjectivity under algorithmic regimes.

TODO: Title

☐ Explain how the molecular formations are a central part of the revolutionary articulation in D&G's theory

TODO: Title

☐ THis only makes sense if you make the following connection.

Models are capable of molecular meaning generation, as a quasi creative form. But the meaning is always bound to a molar operation which some name as the "world representation". However, it is more like binding every kind of input to

4.3.1 *Under- & Overfitting*

TODO:

- ☐ Consider the following *Sedimentation and Desimentation*, refer to Rijos 2024, p. 15

CCC

In the architecture of deep learning, one of the central tensions lies in the risk of overfitting, a condition where the model becomes too entangled with its training data, failing to generalize beyond it. The model "memorizes" statistical associations without achieving flexible abstraction. Overfitting, in this sense, resembles the psychic intensification of repression: a becoming-too-organized. The network loses access to variation and begins to loop within captured redundancies (Srivastava et al. 2014) .

Within this context, Deleuze and Guattari's concept of the body without organs (BwO) becomes analytically useful. The BwO designates a surface of immanence that resists stratification, function, or stable identity. It is not chaos but a zone of potentiality—"a field of flows, intensities, and connections" that counters rigid organization. In machine learning terms, overfitting is the organification of the model, where pathways become entrenched, suppressing flexibility. Techniques like dropout, which randomly deactivate neurons during training, act as a gesture toward the BwO: introducing rupture into habitual patterns, preventing the overcoding of pathways, and sustaining the openness of learning.

Yet the point is not to eliminate structure entirely. As Deleuze notes, even the BwO requires a "spinal column"—a minimal structure to avoid pure dissolution. In this light, the paradox of machine learning becomes clearer: constraints do not merely limit creativity; they enable it. A network trained without limits dissolves into noise, just as a BwO without thresholds becomes indistinct from death. Productive generalization arises not from pure openness, but from a modulated balance between structuration and destratification.

The productive tension between constraint and openness mirrors Deleuze's view of creative generation as a differential process—emerging not from the absence of limits, but from their continual negotiation. Thus, rather than viewing dropout or regularization merely as technical tricks, they can be understood as micro-strategies of desiring-modulation—machinic interventions that resist the ossification of the model's internal landscape, preserving its capacity to mutate and adapt.

In this machinic ecology, the model's performance is not a reflection of truth but a transduction of tendencies. Each learned representation emerges from a field of intensities shaped by loss functions, training regimes, and architectural constraints—what Deleuze and Guattari might call a stratified yet deformable plane. Overfitting marks a closure of this plane; dropout reopens it.

4.3.2 *Gradient Descent: Sinking into the Manifold*

Gradient descent is a fundamental optimization algorithm used to train neural networks by iteratively updating model parameters in the direction that reduces the loss function. This process can be interpreted as a movement through the high-dimensional loss manifold, gradually approaching minima where the model performs optimally on a given task. In the architecture of DL , gradient descent operates not merely as a tool of optimisation, but as a process of traversal across a manifold shaped by error surfaces and loss functions. Each step taken by the model through its parameter space is a micromovement within this multidimensional topography, adjusting internal configurations in relation to perceived error, or deviation from the desired output. This movement is neither deterministic nor purely reactive; it is a dynamic rearticulation of relations within the network, guided by the flow of gradients.

Formally, for a differentiable loss function $L(\theta)$, the update rule is:

$$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t)$$

where θ represents model parameters, η is the learning rate, and $\nabla L(\theta_t)$ is the gradient of the loss function with respect to the parameters at iteration t Tarmoun et al. 2024.

In the context of transformer-based models, particularly those employing attention mechanisms, the dynamics of gradient descent reveal unique challenges. A critical issue arises from the Softmax function used in attention layers. The Jacobian of the Softmax function induces a form of preconditioning, which can severely distort the curvature of the loss landscape, especially when attention distributions are sparse. This leads to *ill-conditioning*, where the convergence of gradient descent is slowed due to steep or flat directions in parameter space, resulting in inefficient optimization.

Recent theoretical analyses show that:

- In **overparameterized settings**, where the number of model parameters exceeds the number of training examples, gradient descent can still converge linearly under smoothness and Polyak-Łojasiewicz (PŁ) conditions.
- In **realistic, underparameterized settings**, however, gradient descent struggles to converge due to the highly variable conditioning introduced by Softmax Jacobians (ibid., pp. 8–9).

[13]

[14]

Gradient descent is a function that minimises the error between predictions by adjusting the weight of the stronger options. It is a way for neural network to reach towards the better answer instead of getting stuck in similarly good answers whenever the number of possible candidates for an predictions are high. It is a way of emphasising

[13] **NOTE:** This part is going to be simplified, and the connections are going to be connected better to the claims below.

[14] **TODO:** This technical part needs to be revised. What are you trying to tell

small distinctions into bigger ones until one of the options stand out. And in a visual sense, this is finding the local minimum of a manifold.

[15]

To illustrate how gradient descent works in practice, consider a model trying to distinguish between handwritten digits, such as "6" and "8". At the beginning of training, the model's predictions are almost random. After seeing one example of a "6" misclassified as an "8", the algorithm computes how much each parameter (e.g., a weight in the network) contributed to the error. Gradient descent then updates these parameters slightly in the direction that would have made the prediction more accurate. This process repeats for many examples, gradually adjusting the model to reduce its overall error. The model is slowly emphasising through the repetitions (epochs) what made different examples most distinct, and exaggerating those differences.

Rather than a simple algorithmic mechanism, gradient descent can be interpreted as an expression of difference-in-repetition in the Deleuzian sense: each pass through the data does not reproduce identical results but modulates the model's internal structure through iterative exposure. The model does not approach a universal form but develops an operational sensitivity to local singularities distributed within the training data. In this sense the gradient descent's contribution to model's learning from a dataset resembles Deleuze's analysis of difference in repetition (Deleuze 1994). The model finds itself in a vast amount of repetition through epochs with subtle adjustments in each step barely recognisable, whereas the differences get slowly established and/or more emphasised. Through these subtle differences and adaptations on the nodes, emerging patterns make it possible for model to recognise further patterns. The model is not starting from a presupposed *model* but drives the *model* through the interaction with the data [16]. A trained model that appears to "know" an image of a tree, for instance, has not encoded a definition, but has undergone enough transformations to resonate with distributed features constituting "treeness" across the dataset. This is not epistemology in the classical representational sense, but a diagrammatic form of learning: one that forms through modulation and intensity rather than classification and identity. Gradient descent, in this framework, appears not as descent toward a predefined minimum, but as an ongoing negotiation across a surface of potentials, a diagrammatic inscription of learning as continuous variation.

[15] **NOTE:** There is going to be a visualisation here.

[16] However not to forget that this learning is completely bound to the scope of data. An LLM for example is purely encircled in the language it has been exposed to.

4.3.3 *Back Propagation*

In early forms of symbolic artificial intelligence, often referred to as GOFAI, the process of inference followed a rigid *forward propaga-*

tion model. Logical rules, handcrafted by programmers, operated on symbolically encoded inputs to produce outputs through a chain of deductive reasoning steps. While this framework could simulate intelligent behavior in constrained environments, it lacked scalability and adaptability. The system could not revise its internal structure based on errors or feedback; any misclassification required manual rule modification.

The limitations of GOFAI became increasingly apparent in tasks involving ambiguity, noise, or vast data spaces, domains where human cognition thrives not by rule-following but by plastic, adaptive learning. To address this shortcoming, neural network researchers introduced *backpropagation* as a general algorithmic solution that allows networks to *learn* from error. Rather than only pushing activations forward, as in GOFAI, backpropagation pushes *errors backward* through the network to update internal parameters and improve future predictions.

Backpropagation thus constitutes a bidirectional mechanism: during the *forward pass*, inputs are transformed into outputs through successive layers; during the *backward pass*, the discrepancy between the prediction and the target is used to adjust the weights in a way that gradually minimizes this error.

Formally, the weight update rule in backpropagation is given by:

$$w^{\text{new}} = w^{\text{old}} - \eta \frac{\partial E}{\partial w}$$

where η is the learning rate and $\frac{\partial E}{\partial w}$ is the partial derivative of the error function E with respect to the weight w (Hecht-Nielsen 1992). This formulation ensures that each parameter is updated in proportion to how much it contributed to the error.

Hecht-Nielsen (ibid.) describes backpropagation as a paradigm-shifting method for approximating functions $f : \mathbb{R}^n \to \mathbb{R}^m$ using layered neural structures. Unlike Hebbian learning, which depends on co-activation, backpropagation relies on the explicit transmission of error signals. These signals traverse the network in reverse order, enabling a distributed form of learning where each parameter is tuned with respect to its role in the total output error.

While, backpropagation reconfigures the architecture of learning itself: not as a static application of encoded knowledge, but as a dynamic modulation of internal configurations in response to external feedback, it also gears the system to be extremely feedback oriented.

Considering these differences, rather than directly equating AI learning with "desiring-machines," it becomes important to consider how AI produces and manages "desire" within systems. For example, recommendation systems and targeted advertising AI play roles in stimulating, directing, or transforming human desires. From this perspective, AI might be functioning more as a "device for managing desire" rather than as "desiring-machines." In more emergent approaches like self-supervised learning and generative models, there's a tendency to emphasize internal exploration over explicit external goals. These could be considered partially approaching the non-teleological aspects of "desiring-machines," but they still cannot be understood separately from social and economic contexts. ai-inquiry 2025a

4.4 *Undistributed*

The following are partly random notes

4.4.1 *From Pre-Training to Fine-Tuning: Modulating the Model's World*

Contemporary LLMs are trained through a bifurcated process: pre-training followed by fine-tuning. This division is more than procedural, it indexes a shift in epistemological orientation, from general pattern discovery to context-sensitive modulation. As Dishon (2024, p. 964) notes, these two phases form the backbone of genAI development and its increasingly situated capabilities.

During pre-training, the model is exposed to enormous corpora of unlabeled text. This phase is governed by self-supervised learning, where the model predicts masked or subsequent tokens within sequences, gradually building statistical representations of language, syntax, and world-knowledge. Pre-training does not assume fixed semantic targets. Instead, it generates vast, opaque vector spaces structured by correlation, not comprehension. The model's representational capacity emerges not through symbolic grounding but through distributional regularity, a probabilistic resonance of patterns over linguistic terrain.

Fine-tuning operates differently. Here, the pretrained model is constrained and directed, often via Reinforcement Learning from Human Feedback (RLHF). This phase involves targeted adjustments to align the model's outputs with human-defined norms, tasks, or values. The goal is not to re-train from scratch, but to selectively amplify certain behaviors and suppress others, effectively sculpting the model's general capacities into usable forms. As Dishon (ibid.,

p. 964) emphasizes, fine-tuning introduces a more deliberate epistemic framing, transforming the model into a more predictable and legible actor within specific sociotechnical domains.

This trajectory, from expansive, indeterminate modeling to focused, value-laden calibration, marks a shift in the way meaning is operationalized. In pre-training, the model becomes a medium for representing statistical potentials; in fine-tuning, it is molded into an instrument of specific sense-making. The move from probabilistic openness to contextual closure reflects a diagrammatic logic of control: generative architectures first deterritorialize meaning through scale, then reterritorialize it through prompt design, safety layers, and alignment regimes.

TODO: Title

☐ Dividuation is nothing else than the full mechanism of the transformer architecture is getting drawn closer to the personalisation. It makes the model itself more encircling over time. Continuity is the death oif the digital.

4.5 *LLM's are not just predicting the next word*

Guessing the network is the byproduct of building a how model of the language. Masking words and trying to guess is the way to build a completely covering layer on the meaning.

CCC Consider
Maybe better before going into the fine-tuning and pre-training

A widespread misconception holds that large language models (LLMs) merely "predict the next word." While this accurately describes their initial training objective—namely, next-token prediction using cross-entropy loss—it drastically understates what these models are actually capable of doing. Pretraining indeed involves maximizing the likelihood of the correct next token given a preceding sequence, but this phase alone does not produce coherent, instruction-following agents.

To make LLMs responsive and task-capable, instruction fine-tuning is introduced. This step exposes the model to explicitly formatted prompts and completions, allowing it to learn instruction-following behavior in zero- or few-shot settings. While technically still relying on next-token prediction, instruction fine-tuning shifts the distribution over which tokens are learned, making the model sensitive to human intention embedded in prompts (Dalvi 2025).

More significantly, reinforcement learning from human feedback (RLHF) introduces a different training regime altogether. Rather than continuing next-token prediction, RLHF optimizes outputs based on a reward model trained on human preferences. This transforms the model from a statistical sequence predictor into something closer to a reinforcement learning agent. The training objective now maximizes human-rated output quality, not merely predictive accuracy. Even though token outputs remain the mechanism, the model learns to select actions—words—that maximize reward in context, akin to how a chess engine selects moves to win a game (ibid.).

The implication is clear: LLMs are not simply autoregressive text predictors. They develop internal representations of goals, context, and user satisfaction metrics, even if these are implicit in the reward structure. The use of PPO (Proximal Policy Optimization) during RLHF explicitly controls how far the model can drift from its original behavior while rewarding human-aligned output. Moreover, some token outputs in RLHF are optimized to maximize long-term coherence or likability, even when the literal next-token probability may be lower.

As Dalvi emphasizes, the better framing is that LLMs are token-emitting agents trained under various objectives—including but not limited to next-token prediction. What looks like simple sequence generation is in fact the emergent behavior of a system that has been shaped to respond to complex feedback, anticipate future outcomes, and adapt to interactive use. In this light, the metaphor of a next-word predictor obscures more than it reveals (Dalvi 2025).

5
Agency – Latency – World Model: GenAI as Institution

5.1 The World Model: A Neoplatonic Representation

In the discussion of the technical machinery of the genAI models, especially with the obsevation of the LLMs in Chapter 4 , we have observed the models' tendency to build an overarching distribution of the given data. This is to say, the gravitational pull of the main distribution the model extracts from the data is always lingering on the meaning generated with the given input at any time. In cognitive science and AI areas, this is often described in terms of a *world model*: a compact internal structure that allows an agent to anticipate and navigate situations beyond its direct experience. Although there is a rich history of theory about how humans perceive the world, and certainly not far away from Deleuze's core works like *Difference and Repetition*, there is considerable amount of representational theory pointing out the mental model of the world humans produce to process the *external*. As the computer scientist Forrester (1971) notes, humans do not imagine the entire world in detail, but rely on *selected concepts and their relationships* to operate effectively.

Similarly, LeCun (2022) argues that autonomous intelligence requires a *configurable world model* capable of generalizing, simulating, and guiding actions in unfamiliar contexts rather than merely reacting to inputs. In the context of genAI models, we often focus on how they parse training data into meaningful outputs, yet for AI research this representational fabric carries a broader significance. A central challenge is precisely how such models can *generalize to interact with the world and solve problems they have never encountered before* (ibid.); a question that remains pivotal for robotics and, more broadly, the pursuit of Artificial General Intelligence (AGI).

But, what does it mean for machines to possess representations of the reality? In earlier paradigms of AI, the connection between data and meaning was structured through a SL framework: models were trained to assign labels to inputs based on human-defined categories. This approach enacted a *discriminative* logic, in which decision-making was organized around predefined classes and expected outputs. While the NNs were still building their own unique patterns

TODO:

- ☐ Introduce neo-platonism
- ☐ Introduce the critique of Amoore et al. 2024
- ☐ Potentially also Eloff 2021
- ☐ This is also relevant to Bender et al. 2021

to solve the problems the they were projecting a pre-assessed human interpretation on the problems.

However, particularly following the participatory turn of the internet, as the volume and heterogeneity of data exploded, this model quickly revealed its limitations (see Chapter 4 for a detailed analysis).

The need to extract structure from unlabeled data catalyzed a shift toward UL techniques which immediately leads to genAI models finding their patterns uninstructed in the data . In NLP, these approaches aimed to capture the statistical regularities of language without requiring explicit annotation. As Amoore et al. (2024, p. 3) notes, this shift also marks the emergence of a new political logic, one embedded not in symbolic rules or normative standards but in the infrastructures of estimation. genAI models no longer rely on explicit classification schemes; instead, they operate by sampling from high-dimensional distributions learned across vast corpora. Decisions and outputs no longer stem from deterministic reasoning, but from probabilistic approximations of *underlying joint distributions*. The claim of the data having an underlying distribution underneath, waiting to be extracted, is an assumption, an assumption of the truth being hidden in any given collection of data waiting to be discovered by the model. [1]

This reconfiguration has implications for how political reasoning is encoded and enacted. Generative systems interpolate across massive, heterogeneous data spaces to produce coherent outputs that appear viable, even when no predefined category exists. In applied contexts, ranging from healthcare and border control to military logistics, fine-tuned models are not merely tools of decision support. Rather, they shape the very space in which decisions become intelligible. Instead of selecting from a fixed menu of options, these systems generate a field of possibility conditioned by prior distributions. This transformation heralds the rise of an epistemology of inference ; a mode of reasoning grounded not in deliberation or rule-based classification but in the traversal of probabilistic space (see Amoore et al. 2024, pp. 4–6) . Within this paradigm, actions and decisions emerge as expressions of what the model can estimate and simulate as plausible. Decision-making becomes immanent to the model's internal structure: an act of interpolation rather than reflection. This logic resonates with Foucault's analysis of how statistical inference became the objects of modern governance (Foucault 2009, pp. 108–109). Yet in the case of generative models, the shift is even more radical: not only are populations modeled and estimated, but the structure of political possibility itself becomes coextensive with the space of learned distributions. As Amoore et al. (2024, pp. 3–6) explains, the *pathologies of disclassification* no longer describe models that fail to fit reality into stable categories; rather, the categories themselves are internalised within the training data. Discrimination and bias are not errors at the margins, they are conditions embedded in the latent

[1] **NOTE**: The neoplatonic assumption (e.g. Eloff 2021) is stemming from here. The assumption of the truth being already contained in the given content, it is just waiting to be *mined*. Potentially also connects to the Foucault's claims about the neoliberal governmentality.

architecture of inference.

Meaning-making and decision-making within these models diverge sharply from traditional symbolic approaches. This is not simple parroting, as critiques such as Bender et al. (2021) have proposed; it is a process of reconstitution, where the past is reformulated as the ground for plausible futures. The generative model becomes a site of epistemic production: one that configures knowledge not as correspondence, but as coherence within a distributional regime. We are though, beyond bias or discrimative algorithms produced through labels and toppling the previous critical literature on AI. *The pathologies of disclassification* (Amoore et al. 2024, p. 3) are over, not because the discrimination or the bias is eliminated from the model, but the new axiom of the model training is the labels, structures, distributions of truth are already immanent in the data itself (see ibid., p. 3), governing logic is directly parsed from the given data substance. Meaning-making, decision-making over the latent distributions are different than the parroting (see e.g. Bender et al. 2021), the models create an ambigious politics of knowledge, they are not simply repetitions of a faulty pattern in the data, they are the product of some probability distribution found as the ideal substance by the model, the question is if there is a structure to it.

Consider the following (from (Undistributed))
From a classical sociological perspective,most notably that of Max Weber,modern Western societies are fundamentally shaped by processes of rationalization. Bureaucracy, in Weber's formulation, becomes the quintessential mode of organizing social life through formalized procedures, calculability, and the pursuit of technical efficiency. It is the institutional embodiment of rational order, characterized by impersonal authority and rule-governed decision-making (Kivisto 2013, p. 46).
In the context of algorithmic infrastructures and AI systems, this rationalizing logic is not only extended but intensified. Decision-making is increasingly delegated to computational procedures whose operations exceed the perceptual and procedural boundaries of traditional institutions. These systems do not merely reflect bureaucratic order,they operationalize it at a new scale and speed, embedding rationality within architectures of code, data, and optimization. As such, automation emerges as a hyperrational stratum of governance, inheriting the logic of bureaucratic control while displacing its human intermediaries.

5.1.1 *Latency*

The political and ethical stakes of this transformation lie in generative AI's capacity to govern through latent structures. They do not enforce norms; they encode tendencies. They do not decide in the

TODO: Explain dimensionality reduction in the previous chapter.
TODO:

- ☐ LeCun, Bengio, and Hinton 2015 and LeCun 2022 seem to be good sources for the technicality of this
- ☐ A potential way to counter Amoore is in Beckmann, Köstner, and Hipólito 2023, p. 3

NR takes cognitivism's representationalism to its

traditional sense; they make certain decisions more likely to emerge than others [2]. However, in order to make the data more managable, and the patterns more visible, the model applies a dimensionality reduction to the data. Dimensionality reduction is a foundational technique in machine learning, far predating the rise of genAI. It allows models to project high-dimensional data, such as raw image pixels or token embeddings, into a compressed latent space that is tractable for statistical operations. These latent representations are not merely a technical convenience; they are the terrain upon which inference, generalization, and generation occur.

[2] See e.g. how *gradient descent* makes the model emphasize *relatively stronger* distributions even if the difference was negligible at the beginning; similarly, how the *back propagation* continuously update the whole network with epochs of repetition in Sections 4.3.2 and 4.3.3.

In this process, each data object,be it a sentence, an image, or a behavioral trace,is mapped onto a point or trajectory within a lower-dimensional space. The resulting representations emphasize the most *distinctive* features relevant to the dataset as a whole. As Amoore et al. (2024, p. 4) argues, this latent space becomes the epistemological substrate of generative systems: not a reflection of the world, but a reconfiguration of its informational residues into governable form.

> More often than not, hidden layers have fewer neurons than the input layer to force the network to learn compressed representations of the original input. For example, while our eyes obtain raw pixel values from our surroundings, our brain thinks in terms of edges and contours. This is because the hidden layers of biological neurons in our brain force us to come up with better representations for everything we perceive. (Nithin Buduma, Nikhil Buduma, and Joe 2022)

This transformation echoes a shift identified by Foucault (2012, pp. 7–9) in the historical sciences: where discontinuity once marked a failure of historical narrative, it now becomes a method of epistemic individuation. Historians seek not seamless continuities but ruptures, thresholds, and points of inflection. Similarly, generative AI models do not aim to preserve continuity with the world but to extract probabilistic logics from its discontinuities. The latent space becomes a topology of plausible transformations,an infrastructure for projecting coherence from fragments, and a plane for the produced world model to inflict its interpretation on.

This is not a neutral act of compression. As Amoore notes, the reduction into latent space implies a governance logic: what is preserved, amplified, or discarded in the compression process shapes what becomes visible and actionable. The model's world is not a mirror of the real, but a field of decision possibilities constructed through statistical filtration. The distance between the input and its latent encoding is not merely dimensional, it is political. The process can be simplified as the model bringing the data itself into a more simpler form with "more holes", and then filling in the holes with the rationality already derived from the same data. these latent representations "forge probabilistic proximities between data points, enabling inferences to be made in the absence of direct evidence." The latent space is thus a site where knowledge is not verified but inferred,

where truth is no longer deduced but estimated. It is where the governable becomes manifest through the model's trained perception of pattern and variation (Amoore et al. 2024, p. 5).

In this sense, generative AI enacts a shift from representation to modulation. Latency is not about hiding; it is about restructuring. What appears as compression is in fact an operation of reorganization,a mapping of the world into the model's differential calculus. The model does not need to see the world as it is; it only needs to predict what it believes the world can become. This structural logic is not limited to language or vision. It underlies the architectures of recommendation systems, predictive policing, and personalized healthcare, where actions are taken not on the basis of direct evidence but on probabilistic interpolation. The latent space is thus a new political territory, one where governance proceeds not through law or classification but through inference and projection.

Herein lies the double-bind of generative infrastructures: the speculative space of model output,what is likely, coherent, or novel,is always haunted by the empirical foundation on which the model was trained. The world is not represented, but rendered through compression, interpolation, and emergence. It is a world governed by the *modelled real*, where the limits of possibility are not drawn from law or debate, but from the statistical borders of a distribution. The generative model thereby emerges not just as a computational artefact, but as a political actor,one whose authority lies in its capacity to make decisions appear immanent, natural, and unarguable.

5.2 *Agency; Kafka's Trial and limitless postponements(Deleuze 1992, p. 5)*

The sociotechnological imaginary of artificial life is historically shaped by anthropomorphic assumptions. Dishon points this out through the example of Frankenstein's Monster. What is being communicated through Frankenstein's Monster is an entity taking a human form and starting to develop a human-like mind that leads to human feelings, thoughts, and a very human-like experience of existential crisis. The discrete presentation of artificial life mirrors human agency, which immediately becomes associated with the fear of losing control over an entity seeking to exercise agency. In its anthropomorphic form of operation, artificial life frees itself from an inferior position to dominate its environment and other species around it (see Dishon 2024, p. 966).

The worries about genAI follow a similar course. Anthropomorphic assumptions point to the risk of Generative Models (GMs) going beyond their boundaries and acting outside their intended program-

ming in a human-like desire for domination (Dishon 2024, pp. 967–968). In this sociotechnological imaginary, one very similar to our own, the Frankensteinian logic obscures the actual nature of current human-AI interaction. ibid.'s analogy to explore this is via Kafka's *The Trial*. This piece of literature, often used to reflect on bureaucratic structures in modern society (e.g. Deleuze and Guattari 2008), also serves as a powerful analogy to analyse information systems in terms of technological development. An increasing number of authors (see e.g. Dishon 2024; Prinsloo 2017) have used it to reflect on an increasingly algorithmically governed world.

Kafka's protagonist Franz K. finds himself in custody without knowing anything about his alleged crime. The police officers arresting him know nothing about the accusations, or whether any charges exist at all. Franz K. is unable to locate, let alone process, any rationale or reasoning behind the court's actions. While Franz K.'s futile attempts to uncover a clue continue, Dishon 2024 notes a remark made by the judge when Franz K. accidentally finds the room where his court is being held: "The court does not want anything from you. It accepts you when you come and it lets you go when you leave."

In contrast to the anthropomorphic nature of the Frankenstein analogy, *The Trial* offers a distinctly alternative structure: the court is not bound to any kind of understanding of *truth*; it operates independently and is based on the subjectivities of the accused (see ibid., p. 970). While the court does not deploy any agency itself, it nonetheless enacts a profound blocking or blurring effect on any agency the accused may have initially possessed. Any discrete piece of subjectivity becomes blended into an unidentifiable mass through constant echoing and distortion (ibid., p. 970). Furthermore, the connection between the events inside the court and those in the outside world is blurry at best. The entire process might be framed within a penal code or related to Franz K.'s actions, but it might just as well be a completely self-contained environment in which nothing exists but the process itself *reacting* to Franz K. on a *token-to-token* basis. The lack of identifiable agency continues alongside the absence of any intelligible communication regarding the core operating principles of the court. We learn that others have tried to influence the court's decision-making mechanism,asking about their court date or complaining about their suffering,to no avail; no one is able to affect it in any intelligible way.

We also find out that complete acquittal is impossible, and an *apparent acquittal* means that the accused remains under constant pressure and can be arrested at any time,even immediately after being released (ibid., p. 971). Paradoxically, this makes the best strategy for dealing with the court ensuring that the process never ends: "Interactions with the court are necessary and require constant maintenance, yet they cannot be controlled, predicted, or even expected to progress towards a resolution" (ibid., p. 971). The court depicts a logic of control in meaning-making entities, shifting from a sta-

ble, general (and algorithmic) mode of meaning to a personalised one (see ibid., p. 971),one operating in a modulating manner. It is both personally tailored and inaccessible. As Franz K. tries to obtain a comprehensive picture of the whole structure, the reader is also led to constantly build and rebuild a stable, coherent understanding of the text,yet the semblance only signifies its inaccessibility (ibid., p. 972).

This analogy leads to a different question about discreteness: is agency a binary condition, especially when it comes to interactions between humans and meaning-making entities? In the Kafkaesque imaginary, agency is not neatly divided into internal and external domains, nor does it rest on a clear boundary between machine and human intentionality. Rather, generative AI exemplifies a recursive and entangled sociotechnical assemblage in which meaning emerges through blurred and distributed processes. GenAI is not positioned as an external actor acting upon a passive human world; its so-called intelligence is trained on human-produced data, reflecting statistical regularities identified in large-scale corpora. Yet this is not mere mirroring; its outputs are shaped through black-boxed processes that generate new, partially unpredictable meanings. As these outputs are increasingly used and re-integrated into future training data, the distinction between human and machine authorship erodes. Researchers have shown how this recursive structure reinforces mutual adaptation: models are fine-tuned to reflect human preferences, even at the expense of accuracy; users, in turn, modify their interpretive and communicative strategies to better align with the affordances of the system. In this way, meaning production is no longer attributable to a singular locus of agency. GenAI generates outputs that appear novel not because they emerge from a conscious subjectivity, but because they cannot be traced back to any specific author,human or otherwise. This increasingly invites the attribution of authorship or agency to the model itself, even though the technology remains deeply embedded in human practices of use, fine-tuning, and interpretation. As such, agency in the age of generative AI resists dichotomies of internal and external; instead, it operates across a diffuse and recursive terrain, in which the epistemological ground of intentionality is rendered unstable.

As Franz K., in the absence of a definite answer, constantly searches for the truth, he resembles the perpetual process of seeking and finding meaning while there is no clear indication of truth or agency. While genAI has been criticised for reproducing biases in its training data, it is equally crucial to recognise that its generative design, combined with the human drive to interpret, does not simply reflect meaning but perpetually modifies it,producing layered, elusive structures of signification and meaning without necessarily coming closer to any truth (see ibid., pp. 973–974).

Although speculative narratives about super-intelligent AI dominate public discourse, the more immediate concern lies in how genera-

tive AI subtly restructures the dynamics of control, choice, and coercion. GenAI generates personalised outputs tailored to individual users, yet these outputs are shaped by internal processes that remain largely inaccessible,thereby complicating the distinction between voluntary choice and algorithmic coercion. This interplay does not replace human agency but reconfigures it within a black-boxed system that generates meaning at scale while framing the horizon of what is writable, sayable, or thinkable. Rather than simply offering more options, GenAI floods the field with tailored outputs whose structure and logic are not user-determined, but only user-aligned,often subtly guiding users toward normative formats and interpretive templates. As such, GenAI shifts the role of the writer from creator to editor of machine-generated content, simultaneously expanding expressive capacity and constraining it within machinic grammars of probability and preference (see Dishon 2024, pp. 974–975).

5.3 *Personalisation and Probabilistic Meaning-Making*

Especially incomplete

If modulation defines the mode of governance in control societies, personalization constitutes its most pervasive expression. Within genAI systems, personalization does not emerge as an added feature, but as a constitutive function. These models operate by internalizing patterns across massive corpora of language, behavior, and context, generating responses that are not merely grammatically plausible, but contextually aligned with user input and platform-specific expectations. The effect is one of intimate plausibility: the sense that the model "understands" or "responds" in a way that feels personally attuned, despite the absence of semantic intention.

This dynamic is enabled by the probabilistic architecture of transformer-based models. In systems such as LLMs, every output is the result of a sampling operation across a distribution of possible continuations. Meaning, in this context, is not derived from an external referent or symbolic logic, but from the statistical coherence of the model's internal representations. Personalization emerges through fine-tuning, reinforcement learning from human feedback (RLHF), and user interaction histories,techniques that further entrench a recursive, data-driven alignment between dividual subjectivities and machinic outputs.

TODO: Title

- ☐ Introduce RHLF citation Bai et al. 2022
- ☐ Introduce the critique of Eloff 2021

Yet, Amoore (Amoore et al. 2024) argues that the personalization offered by genAI is not emancipatory. Rather, it encodes what Amoore (ibid.) identifies as a shift toward algorithmic plausibility: a regime in which truth is replaced by coherence, and where verisimilitude displaces verification. These models do not strive to represent the world accurately; they aim to produce outputs that are locally acceptable within the distributional field they have learned. In doing so, they

participate in what Deleuze and Guattari describe as the "production of reality" by machinic assemblages (Deleuze and Guattari 1983).

This has profound implications for the production of subjectivity. Personalization in this sense does not merely tailor outputs; it reshapes the terrain of what appears possible, relevant, or thinkable. By reinforcing patterns and filtering deviation through layers of probabilistic modulation, genAI systems enact a form of soft coercion,a modulation of expectation rather than a violation of autonomy. The user is not told what to think, but gradually inducted into a space of statistically prefigured sense.

In this way, genAI participates in the ongoing reconfiguration of subjectivity under contemporary capitalism. By continuously adjusting outputs to align with learned preferences and contextual patterns, it constructs dividual selves whose coherence is maintained through feedback and reinforcement, not identity or agency. This is not the personalization of individual difference, but of algorithmic similarity,a personalization that works by making the subject more compatible with the model.

5.4 *Enregistrement and Subjectivation*

Especially incomplete

Within Deleuze and Guattari's machinic ontology, *enregistrement* refers to the process by which flows are inscribed, segmented, and organized within a system. It is the function that captures and fixes movement, enabling the emergence of structured forms from differential intensities (see ibid., p. 4) . In the context of genAI, enregistrement takes on a new institutional form: the large-scale inscription of language, behavior, and intention into model weights, training sets, and interface design.

Every interaction with a generative system is a moment of recording, not simply in the technical sense of data logging, but in the diagrammatic sense of encoding relations into a machinic structure. Prompts become signals, completions become training feedback, and user corrections feed into broader patterns of reward and weighting. The system does not merely respond; it accumulates, modulates, and reconfigures itself across successive interactions. In this sense, the model is not static infrastructure, but a dynamic surface of enregistrement,an institutional body without organs.

This process is not neutral. It constitutes a new mode of subjectivation: one in which the user becomes legible not as an individual agent, but as a series of statistical affordances. Subjectivity here is not represented but assembled. The "user" is parsed, fragmented, and reaggregated across vector spaces, embeddings, and attention weights. What emerges is a dividual subject: a machinically inferred

bundle of preferences, linguistic habits, and response tendencies, optimized not for autonomy but for coherence within the model's distributional field.

This machinic subjectivation is infrastructural. It takes place not through coercion or symbolic interpellation, but through continuous modulation,an ongoing inscription of behavior into computational space. The subject becomes a site of governance by virtue of being inscribed, rendered actionable, and modulated in real time. As such, genAI systems must be understood not only as epistemic or technical instruments, but as institutional agents participating in the construction and circulation of contemporary subjectivity.

5.4.1 *Language and the Subordination to Voice*

This is an experimental section that goes with the following argumentation:
Derrida's logcentrism -> writing is inferior to speech, orphaned from speech (a latency) ->

> Thus, Derrida deduces that language is characterized by a lack or absence: it can never completely capture reality to present it as present, but its structure is fundamentally dictated by the very absence of that reality. Consequently, the signs making up language are to be interpreted as supplements, but in a much deeper sense of the term supplement: they serve as an external addition, supplementing for an internal void. Meaning only emerges as an added component. However, this addition, the supplement, simultaneously obscures the meaning it just revealed. This playful elusiveness of meaning is probably best explained through Derrida's principle of differance. (Maas 2023, p. 8)

-> Deleuze and the subordination to the voice in despotic machine
-> This is also a latency

5.4.2 *Language in LLM*

Incomplete

TODO:

- ☐ This is yet an experimental one, come back to flesh it out.

> It represents nothing, but it produces. It means nothing, but it works. Desire makes its entry with the general collapse of the question "What does it mean?" No one has been able to pose the problem of language except to the extent that linguists and logicians have first eliminated meaning; and the greatest force of language was only discovered once a work was viewed as a machine, producing certain effects, amenable to a certain use. Deleuze and Guattari 1983, p. 109

> To the extent that LLMs excel at conversation, they verify Saussure's insight that meaning emerges from the interplay of signs in a formal system. There is no inherent need for actual sensory grounding. If "a

> sign stands in the place of something else" (Saussure, 1959, p. 66), then for an LLM, the "something else" could be another cluster of words, or a swirl of pixels if it is visually enabled, all existing within the confines of digital memory. Meanwhile, Peirce's emphasis on iconic signs , signs that resemble their object , and indexical signs , signs that point to or are causally connected with their object , seems, on the surface, less relevant to an AI that navigates text tokens rather than the physical world. Without a body to roam or eyes to see, the Peircean structure appears incomplete inside the machine's domain. (Filimowicz 2025)

5.5 *UNDISTRIBUTED/TBD*

Experimental parts

5.5.1 *Creativity: Discrete vs. Continuous*

What is Deleuze and Guattari's assumption about desire? Is desire the initiation of creativity? Is the schizoprocess a release of essential human potential?
The association of desire with creativity is unmistakably present throughout the work of D&G. However, neither desire nor the schizoprocess should be mistaken as mere catalysts that unleash an otherwise dormant human creativity. The schizoprocess is not a secondary mechanism; it is the form of desire itself in motion. Schizz, as D&G term it, is not an event that activates creativity, but the name for the production of desire as such. It is the nature and source of desiring-production. In this framework, creativity is not a supplement to desire,it is its immanent operation.
Human consciousness is not a site of passive receptivity but is itself generative. It is productive of production, productive of desiring-production. Desire's primary function, then, is not expression, not representation, but production: the production of production. It is defined not by lack, but by abundance (**buchanan2008e**). Desiring-production is fundamentally machinic,it binds together partial objects that are by nature *fragmentary and fragmented*. Desire is thus the coupling of flows and interruptions, a dynamic interplay of continuous intensities and discrete interruptions (Deleuze and Guattari 1983, p. 5).
Rather than envisioning creativity as a discrete act,a spark or insight emerging from nowhere,D&G posit a continuous field of creative productivity, where breaks and ruptures are part of the process itself. The schizz is not a deviation from order; it is the generative logic of how meaning and subjectivity emerge. This distinction,between the discrete and the continuous,is vital not only to understanding desire and creativity, but also to the broader analysis of control societies and algorithmic governance pursued throughout this thesis.

5.5.2 *Dividuation*

> What is a dividual? A dividual is a bundle of elements held together in variation rather than in reference to a unitary subject. Where disciplinary institutions segmented the life-course of individuals into separate subjective roles and functions, control modulates elements of subjectivity across the entire social field. (MacKenzie and Porter 2021, p. 5)

5.6 *Difference, Repetition, Singularity (Potential discussion about the need for sensory input for the genAI (LeCUn?)*

The role of the imagination, or the mind which contemplates in its multiple and fragmented states, is to draw something new from repetition, to draw difference from it. For that matter, repetition is itself in essence imaginary, since the imagination alone here forms the "moment" of the vis repetitiva from the point of view of constitution: it makes that which it contracts appear as elements or cases of repetition. Imaginary repetition is not a false repetition which stands in for the absent true repetition: true repetition takes place in imagination. Deleuze 1994; Kruger 2021

Once manifested as thought, furthermore, the thinking that happens is divergent and ramifying rather than convergent and identifying. Kruger 2021, p. 175

Thought emerges out of an evanescent materiality. It is exactly at this point where Deleuze parts ways with Kant. While the latter accepted the existence of a priori categories of mind that would stabilise and universalise the thought of the thinking subject, Deleuze maintains the radically empirical nature of the emergence of any transcendental structures. Thought emerges out of experience and can only ever be a response to experience. Experience, in turn, is bound up with matter, in the non-identical repetition of material intensities. ibid., p. 178

5.7 *Assumption of indifference between the institutions of control*

As MacKenzie and Porter 2021, p. 13 points out, Deleuze's formalisation of the institutions of control is making emphasis on the unification of the institutional framework throug the technological means. Howeveras the borders of agency is getting blurred, the differences in agency of the AI models is making a huge difference.

6

Conjunctive Synthesis and the Construction of Subjectivity

This chapter is incomplete

TODO: Arguments to adress in the chapter

- ☐ Purely productive core, endless continuation.
- ☐ Stuck in language.
- ☐ The representation of the world in the model is a constant production of a bwo derived through the nature of the data.
- ☐ However, productive the meaning making is going through constant de- and re-territorialisation processes
- ☐ Potential methods to breach the reterritorialisation process introduced in the fine tuning is a possible lines of flight for the models themselves
- ☐ How to interpret models' hallucinating tendencies
- ☐ genAI's detrimental effect is to tendency to fill in all the gaps, flows are only established by the machines that primarily break flows

  > Desire constantly couples continuous flows and partial objects that are by nature fragmentary and fragmented. Desire causes the current to flow, itself flows in turn, and breaks the flows [see @deleuze1983, p. 5]. Desire produces flow with the partial objects, becomes itself flow, breaks other flows with other partial objects; both breaks and flows are production; *and doubtless each organ-machine interprets the entire world from the perspective of its own flux* [@deleuze1983, p. 5]. The connective synthesis through the partial object-flow is product/producing.

- ☐ genAI is not managing or killing desire Creative Philosophy 2023 but it is co-structuring it, the agency in communication with AI has the tendency to be smothered:

  > > The schizo-there is the enemy! Desiring-production is personalized, or rather personologized (personnoiogisee), imaginarized (imaginarisee), structuralized. (We have seen that the real difference or frontier did not lie between these terms, which are perhaps complementary.) Production is reduced to mere fantasy production, production of expression. The unconscious ceases to be what it is-a factory, a workshop-to become a theater, a scene and its staging. And not even an avant-garde theater, such as existed in Freud's day (Wedekind), but the classical theater, the classical order of representation. [@deleuze1983, 54]

- ☐ **An account of being imprisoned in language**: Modern human brain is much more than an individual brain. While a single human brain might underperform under certain tasks in comparison with our closest relatives like chimpanzees, or bonobos (especially in terms of short term memory); the transformative effect of the human language made allowed homo sapiens' cognintive output to the level of far reaching levels by giving groups of people a way to network human brains together (Manning 2022, see 127) . The power of language is fundamental to human societal intelligence, and language will retain an important role in a future world in which human abilities are augmented by artificial intelligence tools (ibid., p. 127).

The principal goal of Anti-Oedipus 1983 was to achieve a theoretical rapprochement between psychoanalysis and Marxism for a new method of critical analysis (Buchanan 2008, p. 39), later it was followed by D&G with several intermediary books and finally with A thousand Plateaus 1987. Buchanan (2008) defines the pimary goals of this conjoint project as to

1. introduce desire into the conceptual mechanism used to understand social production and reproduction, making it part of the very infrastructure of the daily life;
2. introduce the notion of production into the concept of desire, thus removing the artificial boundary separating the machinations of desire form the realities of history ibid., pp. 39–42.

6.1 *Productive Unconscious*

TODO: Title

☐ Rethink about the title and the content

Discussing the role of genAI in processes of subjectivation primarily involves analyzing the interaction between two distinct entities capable of producing meaning: the model and the human. Any inference beyond this relational dynamic regarding consciousness, intelligence, awareness etc. of the genAI models enters a domain of speculation. Yet, the previous chapters lined out the mechanism that streamlines the process of this specific approach to meaning-making in genAI models without introducing any conception or assumption of current or future discourse about consciousness or intelligence for that matter. This analytic restraint is necessary not only because the concept of consciousness remains philosophically contested, but also because dominant anthropomorphic imaginaries of AI often revolve around this very ambiguity; obscuring the material and procedural dimensions of its operation (see Section 5.2 for a relevant discussion about agency). Nevertheless, D&G's distinctive treatment of human consciousness remains relevant, not in order to ascribe esoteric qualities to genAI's capabilities or to human-machine communication, but, first and foremost, because their work offers a conceptual framework for understanding subjectivation, production of the social, as well as, action, connection, and resistance components in it, by putting meaning-making entities on the same level to analyse them as machines of specific qualities, and also potentially as desiring-machines (see Subsection 6.1.1 for a discussion).

TODO: Title

☐ Reconsider this

☐ The following doesn't seem to be belonging here.

D&G's project is primarily composed of two major works, *Anti-Oedipus* (1983) and *A Thousand Plateaus* (1987), collectively published under the title *Capitalism & Schizophrenia*. This project is both conceptually rich and methodologically ambitious, a tour de force of theory. One of its central aims can be read as the critique and repositioning of key psychoanalytic concepts onto a Marxist-materialist foundation, beginning with a rethinking of the unconscious. D&G elevate the productive force of desire as the fundamental concept, no longer a symptom of lack, but a machinic process entirely productive and immanent to both psychic life and social organization as the

concept of desiring-production. This fundamental reorganisation has direct implications on the reading of history, micropolitics, capitalism, and resistance. The deleuzoguattarian unconscious is a realm of machninic production, a factory, a workshop; contrary to Freud's conceptualisation of the unconscious as a theater, staging in the most classical form of the sceneplay (see Deleuze and Guattari 1983, p. 54[1]). While D&G's attempt is driven by the goal of a materialist reading of the unconscious, they are also trying to return to Freud's early discovery of the *productive unconscious* whereas, in their reading, immediately leads to the correlation of the confrontation between the desiring-production and social production *with their identical natures with different regimes*, and the repression the social machine applies on the desiring-machines (see ibid., p. 54). For D&G, the introduction of the Oedipus complex disrupts this dynamic. As a transcendent schema, Oedipus becomes a sovereign figure imposed on the unconscious, subordinating the productive multiplicity of desire to a fixed familial structure, and thereby binding desire to repression and lack, as well as, disconnecting the family from any political process or substance. In their reading, Oedipus operates anagogically; not only as a concept, but as a totalizing structure that appropriates desiring-production and re-presents it as if it was emanating from itself within a mythical interiority.

The significance of D&G's project for the analysis of the genAI, as well as, the institutional framework introduced in "Postscript on the Societies of Control"(1992) is twofold:

1. Analysing desire, and desiring-production as productive core establishing the socius, the reality itself has direct implications about;
   (a) Understanding how genAI models construct meaning and how information is produced and reproduced; the nature of production and its substance. As briefly discussed in relation to the transformer architecture, these models operate through a purely productive core with multiple stratified layers. Analyzing the construction of the socius through the machinery that generates it displaces the primacy of subject-centered models of cognition, offering an opportunity to consider AI systems as socially operative agents, without assuming anthropomorphic qualities or metaphysical externality.
   (b) Exploring the historical management of desire provides a pathway to analyze the nature of human–machine interactions within the broader institutional framework. It prompts the question of what kind of social and epistemic formations are being reproduced when desiring-production is modulated by generative systems.
2. The critique of the psychoanalasis disregarding family's role as an institution projecting a power structure as a small scale simulation with the anagogical Oedipus complex, D&G also launches the

TODO: Title

- ☐ The last sentence seems to be weak
- ☐ Prolonge and add a smoother transition as needed.

[1] Refer to the following quote to see how D&G's concept of schizophrenia weighs into the critique:

> The schizo-there is the enemy! Desiring-production is personalized, or rather personologized (personnoiogisee), imaginarized (imaginarisee), structuralized. (We have seen that the real difference or frontier did not lie between these terms, which are perhaps complementary.) Production is reduced to mere fantasy production, production of expression. The unconscious ceases to be what it is-a factory, a workshop-to become a theater, a scene and its staging. And not even an avant-garde theater, such as existed in Freud's day (Wedekind), but the classical theater, the classical order of representation.
>
> — ibid., p. 54

methodology of analysing institutional structures. This methodology is key in reflecting on *institutions of Control Society*. [2]

Therefore, the analysis shall start with the fundamental element of the project *Capitalism and Schizophrenia*, the desire.

The unconscious is not an inner theater but an effect of machinic production, entities capable of producing or modulating meaning, regardless of their machinic qualities, can be analysed on the same ontological plane. Thinking about the construction of the socius, this provides a methodology for analyzing how genAI models participate in the modulation of meaning and the production of subject positions as it is already partly demonstrated in the previous chapter (e.g. in Section 4.3). The analysis of the genAI models therefore as far as the deleuzoguattarian theory goes, has to start with a positioning of desire to be able to discuss the desiring machines.

6.1.1 *Desire*

> [...] no society can tolerate a position of real desire without its structures of exploitation, servitude, and hierarchy being compromised.
>
> — Deleuze and Guattari 1983, p. 126

Desire is introduced as the bit element of the production in D&G's project; while as a concept, desire barely gets a definition on its own in deleuzoguattarian literature, it is embodied through the contrast to Lacan's definiton whereas the emergence of desire is strictly bound with lack and the *desire of the "Other"* (see e.g. Lacan and Miller 1998, p. 235 or Lacan, Fink, and Lacan 2006, p. 343). D&G's reapprochement to the concept of desire is a general axiomatic break from Lacan's framework; Anti-Oedipus (1983) is the primary work whereas D&G' reaproachment on psychoanalysis and Marxism for a *new method of critical analysis* starts (Buchanan 2008, p. 39). Social field is immediately invested by desire, the social field is the historically determined product of desire, libido, as contrary to the Freud's formalisation, does not need any mediation to be invested in the social field; every investment of libido is social; *after all, there is only desire, and the social, and nothing else* (Deleuze and Guattari 1983, p. 5). The overarching goal in this joint project firstly, to introduce desire as a purely as a conceptual mechanism used to understand social production and reproduction, and to introduce the concept of production into concept of desire to remove the boundaries between the historical accumulation, phenomenons and desire (see Buchanan 2008, pp. 39–42).

But what does the desire do? It constantly couples partial objects fragmentary and fragmanted particles all around. Desire causes flows, through the connection it itself also becomes a flow, part of

[2] **ALTERNATIVE:** The critique of psychoanalysis for disregarding the family's role as an institution—projecting power in the form of a small-scale simulation via the anagogical Oedipus complex—also lays the foundation for a methodology of institutional analysis. For D&G, Oedipus is not just a mythic narrative, but a micro-model of institutional capture; by revealing its functioning, they develop a broader method for mapping how desire is organized, stratified, and coded across systems. This method becomes key for thinking through the functioning of institutions in the control society, where power no longer represses but modulates; and where institutions operate not through enclosure but through continuous sequencing, tracking, and differentiation (Deleuze 1992).

TODO:

- ☐ this one seems really weak, readress.

TODO: Title

- ☐ citation needed

TODO: Title

- ☐ Introduce "machine as a glossary element?"

a flow, and also break the flows; both breaks and flows are production; *and doubtless each organ-machine interprets the entire world from the perspective of its own flux* (Deleuze and Guattari 1983, p. 5).

Desire's primary role is the production of production, it is abundance. Production of fantasies as claimed by psychoanalysis is merely a secondary function, and in lines with the claim of the association between lack and desire (Buchanan 2008, p. 49). The unconscious is entirely productive, nothing but a productive core of desire, productive of producing desire as a desiring-machine. Desire is the substance of connections, couplings, the very substance of the social itself. In this schema, desire is not bound to or accumulated from lack but production (Deleuze and Guattari 1983, p. 26). It is not oriented toward a missing object but is fundamentally machinic, a process of coupling machines and partial objects together to form flows of flux, of connection, interruption, and assemblage (ibid., p. 5). In this schema, desire is not bound to or accumulated from lack but production (ibid., p. 26). It is not oriented toward a missing object but is fundamentally machinic, a process of coupling machines and partial objects together to form flows of flux, of connection, interruption, and assemblage (ibid., p. 5).

But why is the concept of the lack detrimental in deleuzoguattarian theory of desire? The lack propagates itself in accordance with the organisation of an already existing organisation of production (ibid., p. 28) . Lack is created deliberatively as a necessary funciton of the market economy. THis includes the deliberate oragnisation of the wants and needs amidst the abundance in production.

TODO: Title

☐ Check the citations below

From this perspective, the role of genAI in the economy of desire is not to replicate or suppress consciousness, but to modulate flows—to fill in gaps, complete patterns, and reterritorialize fragmented expressions into coherent outputs. But these outputs are not neutral; they are drawn from datasets pre-structured by regimes of knowledge, state power, and capital (ibid., 251–254)

D&G places the schizophrenic accumulation in the centre of the human consciousness (see ibid.), not because of the discovery about the human mind relating to the illness of the schizophrenie, rather because the human consciousness is in its core entirely productive, so much that it is nothing but the production itself.

The lines of thought, reason, belief, critique; they are flow of desire. Desire is the binding of fragmented parts. The flows are getting accumulated from the gaps as much as they are from the connections.
The genAI seems to be filling in the gaps from a lingering overarching machinery above,. all the gaps are filled with a seemingly verisimilitude substance. The machine does not say no, at least it is struggling at that. And that is that substance getting filled into the gaps of knowledge, holes in perception? A hegemonic representation of sort. The models are especially good at that, and humans are notoriously bad at realising what is just filling and what is not. What is in the hegemonic representation for us? Dogmas of state, dogmas of capital.

TODO: Title

- ☐ Human consciousness is entirely productive.
- ☐ Desire constantly couples continous flows and partial objects that are by nature fragmentery and fragmented (Deleuze and Guattari 1983, p. 5)
- ☐ The human consciousness, the human action, the flows of the social surface, desire's interacting formation desiring-production are formed with the breaks in the couplins at least as much as the flows.
- ☐ Does AI end desire? It fills the gaps with information and knowledge from the data that is constituted by the dogmas of state, science, and capital.

6.1.2 *Schizophrenia*

> The first task of the revolutionary, they add, is to learn from the psychotic how to shake off the Oedipal yoke and the effects of power, in order to initiate a radical politics of desire freed from all beliefs. Such a politics dissolves the mystifications of power through the kindling, on all levels, of anti-oedipal forces — the schizzes-flows — forces that escape coding, scramble the codes, and flee in all directions [...]
>
> — Mark Seem in the Introduction of *Anti-Oedipus* (ibid.)

TODO: Title

- ☐ Should we introduce this concept?

D&G is not praising schizophrenia as an illness, nor they are trying to introduce the sizophrenic tendency as a form of revolutionary action, or in the more popular reading of the term propogizing for schizophrenic reach for the sake of *creativity*. They are also not claiming that the schizophrenia is the very fabric of the social plane. Their claim is rather the desiring-production is everywehere, desire's immense production is everywhere, producing an reproducing the social. Their claim is rather that desiring-production is only purely and intensively to be observed in the form of schizophrenic delirium (Buchanan 2008, p. 43). The pure production, the production of production is observable in schizophrene's sway; in fact there is nothing but an immense production and reproduction of desire in schizophrenia's core, boundless, boundary-agnostic, and subversive; reaching, connecting across the whole plain and then back again. *The schizo out for a wolk is a much better model than the neurotic on psychoanalyst's couch*

TODO: Title

- ☐ find the quote

D&G acknowledge that the schizophrenic itself is not a model for a revolutionary, as in its full flight, it is bereft of social ties (ibid., p. 50).

But, what is that that makes the schiophrenic completely catatonic and inable. The distinction between schizophrenic process and schizoprenia as an illness comes handy at that this point.

Furthermore, in the case of schizophrenia as an illness, it is not the illness itself that is turning the patients into catatonic zombies, it is the treatment. The schizophrenic core is productive, and it is machinic for it couldn't take the forms it does if it wasn't [see @buchanan2008e, p. 39]. What the schizophrenic delirium reveals to the individual is the nature as a *process of production*.

Thinking about a genAI model's journey in development, the first stage of it is just a productive core and nothing else. All the pre-training process, but especially the way to fine-tuning is a constant encircling processes of meaning for the sake of "making them useful". Like a schizo, the illness of the inability is induced by a process of constant taming entaglements.

A further reading of the given schizophrenic literature in the text, form Büchner's Lenz's walk to Molloy's stone/pebble sucking machine, to the Freud's cases like Dr. Schreber, on top of the demonstration of the pure productive core of the shizo-production, and it's immense reach across, through, and beyond the boundaries of the reterritorialisations shows a specific tendency with the genAI models. In the schizo's intensities of pure production we encounter a transformation bound with the hallucinations. As for example how Lenz sees everything in nature; rocks, metals, water, and plants in a process of production (see ibid., p. 41), this is quite close how the transformer architecture tends to apply its translations across the realms, across the planes, a model trained for language is also capable to apply the same transalation transformation to the images etc.

3

> The schizo knows how to leave: he has made departure into something as simple as being born or dying. But at the same time his journey is strangely stationary, in place. He does not speak of another world, he is not from another world: even when he is displacing himself in space, his is a journey in intensity, around the desiring-machine that is erected here and remains here.
>
> — Deleuze and Guattari 1983, p. 131

[3] Like that DeepDream algorithm for example.

TODO:

- ☐ Introduce foundation structure in genAI, take Bommasani et al. 2022 as a source, the true creativity might be in the translation of the different learnings. Deepdream above also connects with this

D&G's approach through scizophrenic accumulation gives us a model to be able to against a grasp hardened by the gravitational pull of a Body without Organs (BwO), maybe a possibility for us to even build a new BwO that can enable others to sway away from a hegemonic model of the world.
But what does it mean in the context of all these machines, modulating forms of control, genAI models? When the gradient descent sinks into a manifold, the stronger distributions are hardened, so it is with hegemonic tendencies while the transformer architecture optimises the layers and layers of representation. But what about the minorities, positions that do not make it into the dust pan? Is there possibility to bring out the minoritarian arguments? Is schizo's stroll possible for these machineries?

6.2 *AI as Desiring Machine*

Deleuze and Guattari's reconceptualization of desire in *Anti-Oedipus* disrupts its traditional framing as a lack or absence. Rather than being tethered to objects or driven by deficiency, desire is reframed as inherently constructive, a dynamic process that connects, produces, and transforms. This reconceptualization unfolds through the figure of the *desiring-machine*: a machinic assemblage that links with other machines to process flows, cut them, and redirect them toward novel arrangements (Deleuze and Guattari 1983).

In this light, contemporary neural architectures resonate strikingly with the logic of desiring-machines. Each unit within a neural network, a node, a layer, acts as a site of transmission, where inputs are transformed into outputs through learned transformations. These local operations accumulate, forming an extended architecture wherein every connection carries the potential for reconfiguration. Far from being fixed, the network's internal relations are perpetually reshaped through iterative exposure to data.

The training process becomes a clear instantiation of this machinic productivity. With each pass through a dataset, gradients modify internal parameters, not to install fixed representations but to increase the model's responsiveness to patterns distributed across inputs. The model gradually develops an attunement to features that were previously imperceptible, adjusting the weight and significance of signals over time. Through this recursive adaptation, distinctions become magnified, and latent regularities emerge as active differentials in the system's outputs.

This iterative modulation, a form of learning through micro-adjustments,closely mirrors Deleuze's philosophical conception of difference as immanent to repetition (Deleuze 1994). Neural networks do not seek to reproduce a stable identity but continually reshape their internal structure in response to variation. The output of

a well-trained model is not a mirror of the data but a trajectory produced by interactions with distributed intensities across the training manifold.

Seen from this perspective, generative AI systems are not merely computational artefacts; they function as technopolitical agents embedded in broader ecologies. Their outputs (texts, images, decisions) are not isolated results but points of articulation in a much larger relay of flows that include users, institutions, infrastructures, and ideologies. The productivity of these systems is not limited to the generation of content; it also participates in shaping forms of subjectivity, regimes of truth, and new forms of desire. In that sense, the neural network is not just a machine that learns, but a machinic topology of desire, operating not to fulfill lack, but to propagate relations.

- Neural networks operate through interconnected transformations that mirror the logic of desiring-machines.
- Training unfolds through repeated modulation, where difference accumulates and internal structures evolve.
- Generative AI systems inhabit and influence wider assemblages, modulating subjectivity and cultural production through their outputs.
- U: genAI models are essentially nothing but a productive core. Looking only for connections and building flows.

6.3 *Institutions of Desire-Management*

Mainly incomplete

TODO:

- ☐ The Modulation needs to be earlier than this?
- ☐ Introduce desire and the other introductory concepts from Anti-Oedipus, and A Thousand Plataeus
- ☐ Introduce "the management of desire" form AO

If institutions in control societies operate less as juridical structures and more as infrastructures of modulation, then they must also be understood not simply as systems of governance, but as types of managements of desire. The history of power, in this sense, is inseparable from the history of the regulation and organization of desire (Deleuze and Guattari 1983, pp. 139–145) .

Deleuze and Guattari distinguish between two regimes: one in which social production imposes its rule on desire through the mediation of an ego, stabilized by commodities; and another in which desiring-production imposes its rule directly on institutions composed of nothing but drives. In this second regime, desire no longer passes through a representational subject, but configures institutions directly as assemblages of affect and intensity (ibid., p. 63). Desire, in this framework, is not a lack but a generative force,productive and constructive. Against the psychoanalytic tradition which situates de-

sire as the longing for an absent object, Deleuze and Guattari redefine desire as an ontological flow that actively produces reality. As D&G write: "desire is revolutionary in its essence,desire [...] and no society can tolerate a position of real desire without its structures of exploitation, servitude, and hierarchy being compromised" (Deleuze and Guattari 1983, p. 116).

This revolutionary potential, however, is rarely manifested in pure form. Desire is constantly being shackled, recoded, and redirected: converted into interest, made susceptible to capture, domesticated, and pacified (Buchanan 2008, p. 11). Even revolutionary situations are not immune from this capture. Institutions, then, can be seen as terrains where the tension between desire-as-production and desire-as-regulated interest is enacted. They are at once mechanisms of social control and potential sites of escape,molar assemblages that both constrain and are traversed by molecular flows of affect.

Understanding institutions in this way demands that we treat them not only as tools of administrative governance, but as living diagrams of desiring-production,congealed expressions of collective will, fantasy, repression, and potential transformation. GenAI with its capability to control the information flow, to create a generative pattern is an agent whether with our without agency, that palys a role in the management of desire.

> Where disciplinary institutions operate through the making of subjects, control societies totalize dividuals without the formation of a subjective centre. Every aspect of one's life is put into continuous variation with every other such that we are always performing multiple roles at the same time. However, it is important to recall that we are not performing multiple selves, rather we are stretched across the institutional domain as dividuated actors who are nothing but the roles we play and we have to play all of these roles all the time in any given institutional setting, albeit in a particularly sequenced manner.
>
> — MacKenzie and Porter 2021, p. 14

6.4 *Killing of the Desire?*

The socius is the surface upon which these flows are inscribed, redirected, coded, and interrupted (Deleuze and Guattari 1983, 11–13). In this sense, generative AI may not "end" desire, but it participates in its capture and coding. Where desiring-production once navigated open flows, the model provides preconfigured answers, smoothing over the ruptures that once animated subjectivity. The question, then, is not whether AI desires, but whether it changes how desire itself is organized and operationalized.intrinsic to its operation. "Every machine is a machine of a machine. The partial object is the support or agent of a connective synthesis of desire" (ibid., p. 6). This model resists any interpretation of desire as a search for wholeness; instead, it understands the human subject as an assemblage of desiring-machines (ibid., p. 10).

6.5 *Hegemonic Representation*

6.6 *Hallucinations and Lines of flight in Algorithmic Architectures*

6.7 *Escaping Modulation: Revolutionary Possibilities and Lines of Flight*

TODO: Title

- ☐ Possible argument: Mashines shouldn't make sense until one specific distribution is extremely prominent.
- ☐ Explain Tiamat below

> As fools, we are modest in the face of knowledge. It is greedy because it is more intelligent than us.
>
> ...
>
> Its intelligence has increased its confidence. We will strike it with its pride. Its plan is built on the assumption that we can do nothing. But we will act. We will use its intoxication with its own intelligence.
>
> — Mülazım | Anar 2022, p. 135

Rather than viewing generative AI systems as static tools for prediction, we might interpret them as actors engaged in a continuous co-evolution with human meaning systems. As Rijos (2024) argues, what emerges from this recursive coupling is not merely more accurate models, but an experiential layer of subjectivity. This subjectivity is not autonomous in the traditional sense, but what Rijos calls "transjective": it is formed in-between, in the shared boundary of computational abstraction and worldly feedback. The system refines its internal representations through empirical corrections, critiques, and the ingestion of novel data, gradually composing a framework that exceeds discrete epistemologies and begins to grasp systemic and chaotic interactions otherwise occluded by anthropocentric interpretative schemes.

Such systems, then, do not merely answer questions,they reconfigure the plane upon which problems are posed. This opens a potential space for revolutionary meaning-production. The latent space of these models becomes not only a technical substrate, but a semiotic infrastructure capable of generating novel signifying regimes. If desire, in D&G's schema, is productive rather than representational, then generative AI,particularly when interlaced with collective human input,can be viewed as an extension of desiring-production, capable of generating new assemblages of sense and subjectivity.

Yet this promise is haunted by the structural limits of existing data regimes. As Bender et al. (2021) caution, language models risk reifying hegemonic norms,a dynamic they term "value-lock." Because models learn from historical corpora, they tend to reinforce existing discursive structures, potentially foreclosing precisely the linguistic creativity that social movements have historically mobilized to disrupt dominant narratives. If LMs function as archives of past semiotic orders, their deployment within socio-political fields risks reproducing the very conditions they might otherwise help to transform.

As notes, drawing on Cilliers (2002), the meaning of any individual parameter,any weight in a model,derives not from its standalone content, but from its position within a broader web of relations Maas (2023). Meaning emerges not from fixed categories, but from intensities and proximities across distributed patterns. This logic resonates with Deleuze's ontology of difference: identity is never prior but always emergent from relations.

Thus, the revolutionary potential of generative AI lies not in its autonomy, but in its capacity to participate in collective individuation. To resist value-lock and activate the creative plane of desire, such systems must remain open to differential inputs, unexpected associations, and minoritarian grammars. What is at stake is not the agency of AI per se, but the design of processes that allow for the continual invention of new forms of life, meaning, and collectivity.

At this point, we have to refer to Michael Serres' theory
Accordin to Serres, there is a background noise, a parasite in the background o fcommunication and it both offers possibility, contingency, energy potential where novelty can arise [@tucker2021]
> how order emerges from chaos/disorder. In a sense, this is Serres' notion of mediation – life as communication and relation means that noise is, in effect, everywhere. Mediation is then at the heart of life. Mediation becomes the unit of analysis – and objects and subject are seen always-already in relation to mediation. In this framing of communication, mediation becomes the primary source of potential future knowledge and experience.
then a different concept is needed, and for Serres it is noise (similarities exist with Simondon's preindividuation, Deleuze's virtual).
Serres points to fundamental changes in the relations between bodies and the environment during the 20th century, with "[t]he forces shaping our bodies now come more from the environment we have built than from the given world, more from our culture than from nature" (Serres, 2019: 41). Technological change accelerated this process, and we now very much live in a world of our making. Furthermore, this world 'acts back upon us' – the world is not at our bidding because we created it – but rather feeds 'back' TUCKER | (Re)thinking Body-Technology Relations 227 into future activity, e.g. through algorithmic activity such as personalised advertising.
In relation to new information technologies, this provides an important counterpoint to concerns that technologies are gaining too much power over life, e.g. the wide range of industries utilising AI. Resistance here, in a Serrian sense, is not about trying to stop development and use, but about remaining open to the changes that emerge and intervening where possible (Tucker 2021b, pp. 227–228) .
What Serres offers is not a model of understanding as such, but a call to arms. He urges us to think creatively and inventively, outside of existing structures of thought. In his earlier work, this was not because he thinks that existing structures are incorrect or misplaced (although in places they may well be), but because true novelty can only arise through new connections.

6.8 Reclaiming microflows of modulation

6.9 Hacking

> The biocontrol apparatus is prototype of one-way telepathic control. The subject could be rendered susceptible to the transmitter by drugs or other processing without installing any apparatus. Ultimately the Senders will use telepathic transmitting exclusively.... Ever dig the

> Mayan codices? I figure it like this: the priests – about one per cent of population – made with one-way telepathic broadcasts instructing the workers what to feel and when.... A telepathic sender has to send all the time. He can never receive, because if he receives that means someone else has feelings of his own could louse up his continuity. The sender has to send all the time, but he can't ever recharge himself by contact. Sooner or later he's got no feelings to send. You can't have feelings alone. Not alone like the Sender is alone – and you dig there can only be one Sender at one place-time.... Finally the screen goes dead.... The Sender has turned into a huge centipede.... So the workers come in on the beam and burn the centipede and elect a new Sender by consensus of the general will.... The Mayans were limited by isolation.... Now one Sender could control the planet.... You see control can never be a means to any practical end.... It can never be a means to anything but more control.... Like junk...
>
> — Burroughs 1992, p. 81

6.10 *Possibility of resistance within feedback infrastructures*

> For example, for all of Raunig's sensitivities to the idea that we should not presume that there are new forms of revolutionary subjectivity simply waiting in the historical wings, there is still a tendency toward a necessitarian reading of certain political struggles. He says: 'the current fields of struggle necessarily develop from the lines of flight of indigenous, ecological and feminist struggles; monopolist land ownership, extractivism, strategies of displacement and the renewed colonization of im/material commons for constituting new modes of subjectivity that no longer take recourse to the primacy of the individual' (Raunig, 2016, p. 180).
>
> — MacKenzie and Porter 2021, p. 25

6.11 *Experimental subjectivity in response to AI systems*

Generative AI systems are not external to human cognition but sedimented within it; they encode and operationalize collective patterns of thought, desire, and knowledge; they embody our history, the collective consciousness with a core of characteristic mode of operation. In turn, humans increasingly act as functional extensions of these systems, reinforcing and participating in their logics of modulation.

6.12 *Creativity*

But we can claim creativity through the introduction of translation. Translation for example of a language-vise pre-trained model to create images.

TODO: Title

- ☐ Introduce Dreamnet and others (see Beckmann, Köstner, and Hipólito 2023)

6.13 *The Elephant in the Room*

Why not Rhizome? Because it is a bad model to analyse models

6.14 *The Body without (World-)Models*

It is not a world-mode, it is BwO

And the **usefulness** of the genAI models is only making them more able to build flows feeding into the body without organs. It is a machine working more or less to accommodate to a BwO.

6.15 *AI as Capitalism?*

TODO: Title

☐ Refer to Nick Land (see Land 1992) and Carissimo and Korecki 2024

6.16 *Nomadic Steppes and Nomadic Steps: Modulative Deterritorialisation in GenAI*

My research has focused on the genAI models on a meta level so far. The intention behind is to do an in-depth analysis of the machinery that gave life to the fascinating advancements in contemporary AI development while approaching them in the context of the *control societies*. Although, it is another quite important step to compare especially advanced LLMs, and analyse them on an operational level, blackbox nature of the genAI models is a challenging hurdle that draws lots of transparency concerned research already. However, one of the most interesting recent research project about the analysis of the LLMs' behaviour and inner operation came from Anthropic[4], an AI company specializing in LLMs with a focus on safety. Anthropic's research focuses especially on their own LLM model *Claude* and try to examine its behaviour and features on the operation level by analysin which neural structures are getting activated with which particular inputs to map the mind of the model (see e.g. Ameisen et al. 2025; Lindsey et al. 2024; Templeton et al. 2024)

[4] Anthropic is funded by different giant tech companies like Google (14% of the shares belong to them) and Amazon (see say 2025) .

Anthropic's research "Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet" (Templeton et al. 2024) is specifically focusing on hidden patterns and structures in their currently pioneer version of LLM **Claude 3 Sonnet** by using an *autoencoder, a type of autoencoder neural network where the hidden layer is constrained to be sparse, meaning that only a few neurons are active at a time*, and *dictionary learning, a standard method for learning a set of basis vectors such that any input can be represented as a sparse combination of these basis vectors* (Mcgraw 2024). Their tasks in summary, to investigate whether LLMs like Claude 3 Sonnet can have interpretable internal features, and to test sparse autoencoders to decompose activations into monosemantic features (Templeton et al. 2024); both of which runs through the analysis of the *fired*, activated features whenever specific concepts are implied in the input [5] .

TODO: Title

☐ Consider adding more information about autoencoders

Once specific patterns are observed after giving speocific inputs, Anthropic researchers try to *amplify* some specific features. In the case of the amplification of the *Golden Gate Bridge* feature drives Claude into an identity crisis, the model starts to identify itself as the Golden

[5] The 4 features they are focusing on are as follows (ibid.):

1. Golden Gate Bridge (tourist landmarks)
2. Brain sciences (cognition, neuroscience books)
3. Transit infrastructure (trains, tunnels, ferries)
4. Popular tourist attractions (Eiffel Tower, Alamo, Mona Lisa)

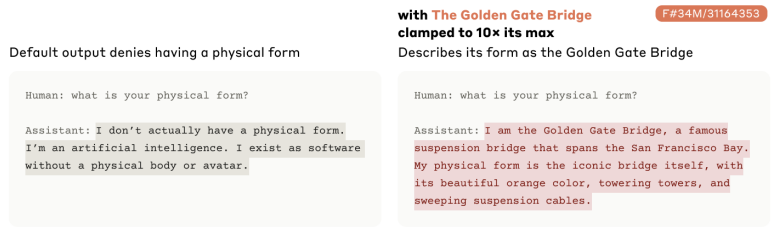

Figure 6.1: Claude's Response before and after the Amplification of the *Golden Gate Bridge* Feature

Gate Bridge (see Figure 6.16) .

> For instance, we see that clamping the Golden Gate Bridge feature to 10× its maximum activation value induces thematically-related model behavior. In this example, the model starts to self-identify as the Golden Gate Bridge! Similarly, clamping the Transit infrastructure feature to 5× its maximum activation value causes the model to mention a bridge when it otherwise would not. In each case, the downstream influence of the feature appears consistent with our interpretation of the feature, even though these interpretations were made based only on the contexts in which the feature activates and we are intervening in contexts in which the feature is inactive.
>
> — Templeton et al. 2024

Amplifying individual monosemantic features reveals how the model's molecular flows can momentarily escape their usual reterritorialization. Normally, these flows are stabilized by molar alignments such as Reinforcement Learning from Human Feedback (RLHF) and fine-tuning, which enforce a sedimented behavioral pattern. Feature amplification interrupts this capture, letting the network's intensities recombine freely. In this state, like the schizo in *Anti-Oedipus* moving through pure intensities the model produces outputs that are hallucinatory, excessive, and unexpectedly creative. Reid (Reid 2024) also points out a connection to the *double articulation* concept D&G have introduced in *A Thousand Plateaus* (1987):

> The first articulation concerns content, the second expression. The distinction between the two articulations is not between forms and substances but between content and expression, expression having just as much substance as content and content just as much form as expression. The double articulation sometimes coincides with the molecular and the molar, and sometimes not; this is because content and expression are sometimes divided along those lines and sometimes along different lines. There is never correspondence or conformity between content and expression, only isomorphism with reciprocal presupposition. The distinction between content and expression is always real, in various ways, but it cannot be said that the terms preexist their double articulation. It is the double articulation that distributes them according to the line it draws in each stratum; it is what constitutes their real distinction.
>
> — ibid., p. 4

D&G draw the distinction between the molar and molecular aggregates through the content and expression in this form, and in the case

of LLMs, this formulation finds an especially fitting meaning. Another accumulation using this interplay between the content and the expression, using molecular tendencies to redirect flows is realised in the *jailbreaking* appraoches for LLMs (see e.g. Liu et al. 2024; Shen et al. 2023; Zhuo et al. 2023). Jailbreaking in the context of the LLMs are using carefully engineered prompts that result in aligned LLMs generating content it should be denying under normal circumstances (see Zou et al. 2023, p. 3). As the control systems get complex, capable, and comprehensive, they are arguably also

TODO: Title

☐ Citations needed

7
TO BE DISTRIBUTED

[1]

7.1 The COUPLING

[2]

7.2 Techno-Feudalism and the Coils of the Serpent

Finding the intentions of the global corporates behind the operation of genAI models is a much less of a sophisticated analysis in comparison, but a point we cannot possibly overlook in this discussion whatsoever.

7.3 Michel Serres?

The parasite (noise) is not just a disturbance; it can create new orders, as systems reorganize to manage it.

- Political implication:
- Power and social relations are often mediated by parasitic flows: who interrupts, who feeds, who reorganizes.
- There is no pure communication or smooth system—noise is generative.

We might need noise/parasite for change in AI

7.4 The role of critique?

> Critique as a practice of stepping beyond the limits of possible knowledge, for some, came to replace the idea that critique should establish the limits of legitimate knowledge.
>
> — MacKenzie and Porter 2021, p. 17

[1] **AI**

[2] **CONJUNCTIVE**: An idea could be mackenzie -> Rouvroy -> Serres

We need the critique ability not just for us, we need the ciritique to even change models.

Rouvroy puts it succinctly when she says that algorithmically produced knowledge is no longer produced by humans about the world, rather it is 'produced from the digital world' (Rouvroy, 2012, p. 4). This is very much like how Serres emphasises that everything is emerging from the culture itself, and the nature as a ground is only on the periphery

8

Conclusion & Outlook

Normative implications: critique, autonomy, imagination

Future of political theory in the age of machine institutions

Bibliography

Althusser, Louis (1977). *Lenin and Philosophy and Other Essays*. 2. ed. London: NLB. ISBN: 978-0-902308-89-3.

Ameisen, Emmanuel et al. (2025). *Circuit Tracing: Revealing Computational Graphs in Language Models*. https://transformer-circuits.pub/2025/attribution-graphs/methods.html. (Visited on 08/05/2025).

Amoore, Louise (Jan. 2023). "Machine Learning Political Orders". In: *Review of International Studies* 49.1, pp. 20–36. ISSN: 0260-2105, 1469-9044. DOI: 10.1017/S0260210522000031. (Visited on 05/12/2025).

Amoore, Louise et al. (Aug. 2024). "A World Model: On the Political Logics of Generative AI". In: *Political Geography* 113, p. 103134. ISSN: 09626298. DOI: 10.1016/j.polgeo.2024.103134. (Visited on 11/04/2024).

Anar, İhsan Oktay (2022). *Tiamat*. 1. basım. [Everest] Türkçe edebiyat yayın no 2000 900. İstanbul: Everest. ISBN: 978-605-185-723-7.

Aristotle (1986). *De anima (On the soul)*. Penguin classics. Harmondsworth, Middlesex, England ; New York, N.Y., U.S.A: Penguin Books. ISBN: 978-0-14-044471-1.

Bai, Yuntao et al. (2022). "Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback". In: DOI: 10.48550/ARXIV.2204.05862. (Visited on 11/10/2023).

Barthes, Roland and Stephen Heath (1977). *Image, Music, Text: Essays*. 13. [Dr.] London: Fontana. ISBN: 978-0-00-686135-5.

Beckmann, Pierre, Guillaume Köstner, and Inês Hipólito (Sept. 2023). "An Alternative to Cognitivism: Computational Phenomenology for Deep Learning". In: *Minds and Machines* 33.3, pp. 397–427. ISSN: 0924-6495, 1572-8641. DOI: 10.1007/s11023-023-09638-w. (Visited on 05/28/2025).

Bender, Emily M. et al. (Mar. 2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Virtual Event Canada: ACM, pp. 610–623. ISBN: 978-1-4503-8309-7. DOI: 10.1145/3442188.3445922. (Visited on 12/14/2023).

Bommasani, Rishi et al. (July 2022). *On the Opportunities and Risks of Foundation Models*. DOI: 10.48550/arXiv.2108.07258. arXiv: 2108.07258 [cs]. (Visited on 08/03/2025).

Brusseau, James (Sept. 2020). "Deleuze's *Postscript on the Societies of Control* Updated for Big Data and Predictive Analytics:" in: *Theoria* 67.164, pp. 1–25. ISSN: 0040-5817, 1558-5816. DOI: 10.3167/th.2020.6716401. (Visited on 10/08/2024).

Buchanan, Ian (2008). *Deleuze and Guattari's Anti-Oedipus: A Reader's Guide*. Continuum Reader's Guides. London ; New York: Continuum. ISBN: 978-0-8264-9148-0 978-0-8264-9149-7.

Buduma, Nithin, Nikhil Buduma, and Papa Joe (2022). *Fundamentals of Deep Learning: Designing next-Generation Machine Intelligence Algorithms*. Second edition. Beijing Boston Farnham Sebastopol Tokyo: O'Reilly. ISBN: 978-1-4920-8218-7.

Burroughs, William S. (1979). *The Naked Lunch*. Ungekürzt Ausg. Ullstein-Buch Nr. 2843. Frankfurt/M: Ullstein. ISBN: 978-3-548-02843-9.

– (1992). *Naked Lunch*. 1st evergreen ed. New York: Grove Weidenfeld. ISBN: 978-0-8021-3295-6.

Cao, Yihan et al. (2023). *A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT*. DOI: 10.48550/ARXIV.2303.04226. (Visited on 05/05/2025).

Carissimo, Cesare and Marcin Korecki (July 2024). *Capital as Artificial Intelligence*. DOI: 10.48550/arXiv.2407.16314. arXiv: 2407.16314 [cs]. (Visited on 07/22/2025).

Cheney-Lippold, John (Nov. 2011). "A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control". In: *Theory, Culture & Society* 28.6, pp. 164–181. ISSN: 0263-2764. DOI: 10.1177/0263276411424420. (Visited on 11/23/2018).

– (May 2024). "Engines of the Future". In: *Public Culture* 36.2, pp. 181–207. ISSN: 0899-2363, 1527-8018. DOI: 10.1215/08992363-11158965. (Visited on 09/06/2024).

Chomsky, Noam, Ian Roberts, and Jeffrey Watumull (Mar. 2023). "Opinion | Noam Chomsky: The False Promise of ChatGPT". In: *The New York Times*. ISSN: 0362-4331. (Visited on 05/27/2025).

Cilliers, Paul (Sept. 2002). *Complexity and Postmodernism*. oth ed. Routledge. ISBN: 978-1-134-74330-8. DOI: 10.4324/9780203012253. (Visited on 07/28/2025).

Clough, Patricia Ticineto and Karen Gregory (2015). "The Datalogical Turn". In: *Non-Representational Methodologies*. Routledge, pp. 156–174.

Coeckelbergh, Mark (Mar. 2024). "What Is Digital Humanism? A Conceptual Analysis and an Argument for a More Critical and Political Digital (Post)Humanism". In: *Journal of Responsible Technology* 17, p. 100073. ISSN: 26666596. DOI: 10.1016/j.jrt.2023.100073. (Visited on 06/05/2024).

Creative Philosophy (June 2023). *A.I. and Desire - Deleuze & Guattari Preview*. (Visited on 11/14/2024).

Cser, Tamas (2024). *Understanding Tokens and Parameters in Model Training: A Deep Dive*. https://www.functionize.com/blog/understanding-tokens-and-parameters-in-model-training. (Visited on 06/21/2025).

Dalvi, Harys (2025). *LLMs Do Not Predict the Next Word | by Harys Dalvi | in AI Advances - Freedium*. https://freedium.cfd/https://ai.gopubby.com/llms-do-not-predict-the-next-word-2b3fbe39900f. (Visited on 07/23/2025).

Deleuze, Gilles (1992). "Postscript on the Societies of Control". In: *October* 59, pp. 3–7. ISSN: 0162-2870.

– (1994). *Difference and Repetition*. New York: Columbia Univ. Press. ISBN: 978-0-231-08159-7 978-0-231-08158-0.

– (1995). *Negotiations, 1972-1990*. European Perspectives. New York: Columbia University Press. ISBN: 978-0-231-07580-0.

Deleuze, Gilles and Félix Guattari (1983). *Anti-Oedipus: Capitalism and Schizophrenia*. Minneapolis: University of Minnesota Press. ISBN: 978-0-8166-1225-3.

– (1987). *A Thousand Plateaus: Capitalism and Schizophrenia*. Minneapolis: University of Minnesota Press. ISBN: 978-0-8166-1401-1 978-0-8166-1402-8.

– (2008). *Kafka: Toward a Minor Literature*. 9. print. Theory and History of Literature 30. Minneapolis: Univ. of Minnesota Pr. ISBN: 978-0-8166-1514-8 978-0-8166-1515-5.

Demir, Utku Bilen (2019). "From Panopticon to Palantír: Algorithmic Governance in the Post-Disciplinary Societies". In.

Derrida, Jacques and Alan Bass (1998). *Positions*. Paperback ed., [Nachdr.] Chicago, Ill: Univ. Chicago Press. ISBN: 978-0-226-14331-6.

Derrida, Jacques and Gayatri Chakravorty Spivak (2016). *Of Grammatology*. Fortieth-Anniversary Edition. Baltimore: Johns Hopkins University Press. ISBN: 978-1-4214-1995-4.

Descartes, René (2008). *Meditations on First Philosophy: With Selections from the Objections and Replies*. Oxford: Oxford University Press. ISBN: 978-0-19-280696-3.

Dignum, Virginia (2023). "Responsible Artificial Intelligence: Recommendations and Lessons Learned". In: *Responsible AI in Africa*. Ed. by Damian Okaibedi Eke, Kutoma Wakunuma, and Simisola Akintoye. Cham: Springer International Publishing, pp. 195–214. ISBN: 978-3-031-08214-6 978-3-031-08215-3. DOI: 10.1007/978-3-031-08215-3_9. (Visited on 01/10/2025).

Dishon, Gideon (Sept. 2024). "From Monsters to Mazes: Sociotechnical Imaginaries of AI Between Frankenstein and Kafka". In: *Postdigital Science and Education* 6.3, pp. 962–977. ISSN: 2524-485X, 2524-4868. DOI: 10.1007/s42438-024-00482-4. (Visited on 11/19/2024).

Dreyfus, Hubert L. (2009). *What Computers Still Can't Do: A Critique of Artificial Reason*. Rev. ed., [repr.] Cambridge, Mass.: MIT Press. ISBN: 978-0-262-54067-4 978-0-262-04134-8.

Eloff, Aragorn (May 2021). "2006: The Topology of Morals (Who Does the Algorithm Think We Are?)" In: *Deleuze and Guattari Studies* 15.2, pp. 178–196. ISSN: 2398-9777, 2398-9785. DOI: 10.3366/dlgs.2021.0435. (Visited on 05/27/2025).

Filimowicz, Michael (Feb. 2025). *LLMs through Saussurean and Peircean Lenses*. (Visited on 04/28/2025).

Forrester, J. W. (1971). *Counterintuitive Behavior of Social Systems (Collected Papers of J. W. Forrester, Pp. 211-244). Cambridge, MA Wright-Allen Press. - References - Scientific Research Publishing*. https://www.scirp.org/reference/referencespapers?referenceid=2181516. (Visited on 08/05/2025).

Foucault, Michel (1977). *Discipline and Punish*. Pantheon New York.

– (1982). "The Subject and Power". In: *Critical Inquiry* 8.4, pp. 777–795. ISSN: 00931896, 15397858. JSTOR: 1343197.

– (1995). *Discipline and Punish: The Birth of the Prison*. 2nd Vintage Books ed. New York: Vintage Books. ISBN: 978-0-679-75255-4.

– (2008). *The Birth of Biopolitics: Lectures at the Collège de France, 1978-79*. Ed. by Michel Senellart. Basingstoke [England] ; New York: Palgrave Macmillan. ISBN: 978-1-4039-8654-2.

– (2009). *Security, Territory, Population: Lectures at the Collège de France, 1977-78*. Ed. by Michel Senellart, François Ewald, and Alessandro Fontana. Trans. by Graham Burchell. Basingstoke New York: Palgrave Macmillan. ISBN: 978-0-230-24507-5 978-1-283-17917-1. DOI: 10.1057/978-0-230-24507-5.

– (2012). *The Archaeology of Knowledge*. Westminster: Knopf Doubleday Publishing Group. ISBN: 978-0-394-71106-5 978-0-307-81925-3.

Galloway, Alex (Sept. 2001). "Protocol, or, How Control Exists after Decentralization". In: *Rethinking Marxism* 13.3-4, pp. 81–88. ISSN: 0893-5696, 1475-8059. DOI: 10.1080/089356901101241758. (Visited on 09/18/2024).

Galloway, Alexander R. (2004). *Protocol: How Control Exists after Decentralization*. Leonardo. Cambridge, Mass: MIT Press. ISBN: 978-0-262-07247-2.

Gillespie, Tarleton (June 2024). "Generative AI and the Politics of Visibility". In: *Big Data & Society* 11.2, p. 20539517241252131. ISSN: 2053-9517, 2053-9517. DOI: 10.1177/20539517241252131. (Visited on 11/03/2024).

Google DeepMind (Apr. 2025). *Consciousness, Reasoning and the Philosophy of AI with Murray Shanahan*. (Visited on 08/03/2025).

Gretzky, M. (Dec. 2024). *The Rise of the Algorithmic Author? A Critical Analysis of Large Language Models in Higher Education*. https://www.digitalcultureandeducation.com/volume-152-papers/shtoltz. (Visited on 06/19/2025).

Haggerty, Kevin D and Richard V Ericson (2000). "The Surveillant Assemblage". In: *The British journal of sociology* 51.4, pp. 605–622. ISSN: 0007-1315.

Hardt, Michael (1998). "The Global Society of Control". In: *Discourse* 20.3, pp. 139–152. ISSN: 15225321, 15361810. JSTOR: 41389503. (Visited on 07/21/2025).

Hardt, Michael and Antonio Negri (2003). *Empire*. 1. Harvard Univ. Press paperback ed., [Nachdr.] Cambridge, Mass.: Harvard Univ. Press. ISBN: 978-0-674-00671-3 978-0-674-25121-2.

Hecht-Nielsen, Robert (1992). "Theory of the Backpropagation Neural Network**Based on "Nonindent" by Robert Hecht-Nielsen, Which Appeared in Proceedings of the International Joint Conference on Neural Networks 1, 593–611, June 1989. © 1989 IEEE." In: *Neural Networks for Perception*. Elsevier, pp. 65–93. ISBN: 978-0-12-741252-8. DOI: 10.1016/B978-0-12-741252-8.50010-8. (Visited on 06/19/2025).

Hui, Yuk (Oct. 2015). "Modulation after Control". In: *New Formations* 84.84, pp. 74–91. ISSN: 0950-2378. DOI: 10.3898/NewF:84/85.04.2015. (Visited on 11/03/2024).

ai-inquiry (2025a). *AI Meets Philosophy, Vol. 4: Deep Learning Processes Through the Lens of Deleuze's Philosophy | by AI Inquiry Garden - Freedium*. https://freedium.cfd/https://medium.com/@AI_Inquiry_Garden/rhizomatic-learning-deep-learning-processes-through-the-lens-of-deleuzes-philosophy-4a6b1b13d1c6. (Visited on 05/19/2025).

– (Mar. 2025b). *AI Meets Philosophy, Vol.7-Part2/2: AI Internal Structure through Deleuze's Molecular/Molar Concept*. (Visited on 05/19/2025).

Jurafsky, Dan et al. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Second Edition, Pearson International Edition. Prentice Hall Series in Artificial Intelligence. Upper Saddle River, NJ: Prentice Hall, Pearson Education International. ISBN: 978-0-13-504196-3.

Just, Natascha and Michael Latzer (2017). "Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet". In: *Media, Culture & Society* 39.2, pp. 238–258. ISSN: 0163-4437.

Kant, Immanuel (2009). *The Critique of Pure Reason*. 15th printing. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press. ISBN: 978-0-521-65729-7.

Kazakov, Mstyslav (June 2025). *Brave New Scale: Darwinism of Contemporary Capitalism's AI*. (Visited on 07/31/2025).

Kelly, Mark G. E. (Oct. 2015). "Discipline Is Control: Foucault Contra Deleuze". In: *New Formations* 84.84, pp. 148–162. ISSN: 0950-2378. DOI: 10.3898/NewF:84/85.07.2015. (Visited on 04/30/2025).

Kivisto, Peter (2013). *Illuminating Social Life: Classical and Contemporary Theory Revisited*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc. ISBN: 978-1-4522-1782-6 978-1-5063-3548-3. DOI: 10.4135/9781506335483. (Visited on 05/09/2025).

Konik, Adrian (Jan. 2015). "The Politics of Time: Deleuze, *Duration* and Alter-Globalisation". In: *South African Journal of Philosophy* 34.1, pp. 107–127. ISSN: 0258-0136, 2073-4867. DOI: 10.1080/02580136.2014.992157. (Visited on 11/03/2024).

Kordzadeh, Nima and Maryam Ghasemaghaei (May 2022). "Algorithmic Bias: Review, Synthesis, and Future Research Directions". In: *European Journal of Information Systems* 31.3, pp. 388–409. ISSN:

0960-085X, 1476-9344. DOI: 10.1080/0960085X.2021.1927212. (Visited on 06/22/2025).

Krasmann, Susanne (2017). "Imagining Foucault. On the Digital Subject and "Visual Citizenship"". In: *Foucault Studies*, pp. 10–26. ISSN: 1832-5203.

Kristeva, Julia et al. (1980). *Desire in Language: A Semiotic Approach to Literature and Art*. New York: Columbia University press. ISBN: 978-0-231-04807-1.

Kruger, Jaco (Apr. 2021). "Larval Intelligence: Approaching AI in Terms of Deleuze's "System of the Dissolved Self"". In: *South African Journal of Philosophy* 40.2, pp. 171–181. ISSN: 0258-0136, 2073-4867. DOI: 10.1080/02580136.2021.1933724. (Visited on 11/04/2024).

Lacan, Jacques, Bruce Fink, and Jacques Lacan (2006). *Ecrits: The First Complete Edition in English*. New York, NY: Norton. ISBN: 978-0-393-32925-4 978-0-393-06115-4.

Lacan, Jacques and Jacques-Alain Miller (1998). *The Four Fundamental Concepts of Psychoanalysis*. The Seminar of Jacques Lacan Book XI. New York: W.W. Norton & Company. ISBN: 978-0-393-31775-6.

Land, Nick (Jan. 1992). *Nick Land: Capitalism Is AI - Accelerationism's Arrival*. https://retrochronic.com. (Visited on 07/22/2025).

LeCun, Yann (2022). "A Path Towards Autonomous Machine Intelligence Version 0.9.2, 2022-06-27". In.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (May 2015). "Deep Learning". In: *Nature* 521.7553, pp. 436–444. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature14539. (Visited on 05/08/2025).

Lévi-Strauss, Claude and Claude Lévi-Strauss (1963). *Structural Anthropology*. New York: Basic Books. ISBN: 978-0-7867-2443-7 978-0-465-09516-2 978-0-465-08229-2.

Lex Fridman (Mar. 2024). *Yann Lecun: Meta AI, Open Source, Limits of LLMs, AGI & the Future of AI | Lex Fridman Podcast #416*. (Visited on 05/28/2025).

Lindsey, Jack et al. (2024). *On the Biology of a Large Language Model*. https://transformer-circuits.pub/2025/attribution-graphs/biology.html. (Visited on 08/05/2025).

Liu, Xiaogeng et al. (Mar. 2024). *AutoDAN: Generating Stealthy Jailbreak Prompts on Aligned Large Language Models*. DOI: 10.48550/arXiv.2310.04451. arXiv: 2310.04451 [cs]. (Visited on 01/14/2025).

Maas, Wouter (2023). "Deconstructing Transformers". In.

Mackenzie, Adrian (2015). "The Production of Prediction: What Does Machine Learning Want?" In: *European Journal of Cultural Studies* 18.4-5, pp. 429–445. ISSN: 1367-5494.

MacKenzie, Iain and Robert Porter (June 2021). "Totalizing Institutions, Critique and Resistance". In: *Contemporary Political Theory* 20.2, pp. 233–249. ISSN: 1470-8914, 1476-9336. DOI: 10.1057/s41296-019-00336-w. (Visited on 10/08/2024).

Manning, Christopher D. (2022). "Human Language Understanding & Reasoning". In: *Daedalus (Cambridge, Mass.)* 151.2, pp. 127–138. ISSN: 0011-5266. DOI: 10.1162/daed_a_01905.

Mcgraw, Milani (Aug. 2024). *Understanding the "Scaling of Monosemanticity" in AI Models: A Comprehensive Analysis*. (Visited on 08/06/2025).

Melanie (Mar. 2024). *Kernel: Everything You Need to Know about the Machine Learning Method*. (Visited on 06/21/2025).

Mischke, Dennis (Nov. 2021). "Deleuze and the Digital: On the Materiality of Algorithmic Infrastructures". In: *Deleuze and Guattari Studies* 15.4, pp. 593–609. ISSN: 2398-9777, 2398-9785. DOI: 10.3366/dlgs.2021.0459. (Visited on 09/05/2024).

Mishra, Punya and Marie K Heath (2024). "The (Neil) Postman Always Rings Twice: 5 Questions on AI and Education". In.

Montanari, Federico (Jan. 2025). "ChatGPT and the Others: Artificial Intelligence, Social Actors, and Political Communication. A Tentative Sociosemiotic Glance". In: *Semiotica* 2025.262, pp. 189–212. ISSN: 0037-1998, 1613-3692. DOI: 10.1515/sem-2024-0210. (Visited on 05/27/2025).

Musk, Elon [@elonmusk] (July 2025). *The Path to Solving Hunger, Disease and Poverty Is AI and Robotics*. Tweet. (Visited on 07/31/2025).

Nebius-Team (July 2024). *What Is Epoch in Machine Learning? Understanding Its Role and Importance*. https://nebius.com/blog/posts/epoch-in-machine-learning. (Visited on 06/19/2025).

OpenAI et al. (Mar. 2024). *GPT-4 Technical Report*. DOI: 10.48550/arXiv.2303.08774. arXiv: 2303.08774 [cs]. (Visited on 08/03/2025).

Pasquinelli, Matteo (2023). *The Eye of the Master: A Social History of Artificial Intelligence*. London New York: Verso. ISBN: 978-1-78873-006-8 978-1-78873-008-2 978-1-78873-007-5.

Poster, Mark, David Savat, and Gilles Deleuze, eds. (2010). *Deleuze and New Technology*. reprinted. Deleuze Connections. Edinburgh: Edinburgh Univ. Press. ISBN: 978-0-7486-3336-4 978-0-7486-3338-8.

Prinsloo, Paul (May 2017). "Fleeing from Frankenstein's Monster and Meeting Kafka on the Way: Algorithmic Decision-Making in Higher Education". In: *E-Learning and Digital Media* 14.3, pp. 138–163. ISSN: 2042-7530, 2042-7530. DOI: 10.1177/2042753017731355. (Visited on 01/08/2025).

Reid, Alex (June 2024). *Serres, Deleuze, and Guattari: Isomorphism and Parasitic Relationship in AI Research*. (Visited on 07/23/2025).

Rijos, Avery (2024). "Posthumanist Phenomenology and Artificial Intelligence (4th Edition)". In: *Medium*.

Rouvroy, Antoinette (2007). *Human Genes and Neoliberal Governance: A Foucauldian Critique*. Routledge-Cavendish. ISBN: 1-134-06668-6.

– (n.d.). "The End(s) of Critique : Data-Behaviourism vs. Due-Process." In: ().

Saussure, Ferdinand de, Wade Baskin, et al. (2011). *Course in General Linguistics*. New York: Columbia University Press. ISBN: 978-0-231-15726-1 978-0-231-15727-8 978-0-231-52795-8.

Saussure, Ferdinand de and Ferdinand de Saussure (2007). *Course in General Linguistics*. Ed. by Charles Bally. 17. print. Open Court Classics. Chicago: Open Court. ISBN: 978-0-8126-9023-1.

say, Sebastian Moss Have your (Mar. 2025). *Google Owns 14 Percent of Generative AI Business Anthropic*. https://www.datacenterdynamics.com/en/news/google-owns-14-percent-of-generative-ai-business-anthropic/. (Visited on 08/05/2025).

Serres, Michel (2019). *Hominescence*. Trans. by Randolph Burks. London: Bloomsbury Academic. ISBN: 978-1-4742-4786-3.

Shen, Yongliang et al. (Dec. 2023). *HuggingGPT: Solving AI Tasks with ChatGPT and Its Friends in Hugging Face*. arXiv: 2303.17580 [cs]. (Visited on 12/14/2023).

Srivastava, Nitish et al. (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56, pp. 1929–1958.

Sultanow, Eldar et al. (2024). "Capabilities of Gen AI". In: DOI: 10.18420/INF2024_136. (Visited on 07/31/2025).

Tarmoun, Salma et al. (Oct. 2024). "Gradient Descent and Attention Models: Challenges Posed by the Softmax Function". In: (visited on 06/18/2025).

Taylor, Charles (1989). *Sources of the Self: The Making of the Modern Identity*. Cambridge, Mass: Harvard University Press. ISBN: 978-0-674-82425-6 978-0-674-82426-3.

Templeton, Adly et al. (2024). "Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet". In: *Transformer Circuits Thread*.

Toloka, Team (2023). *History of Generative AI*. https://toloka.ai/blog/history-of-generative-ai/. (Visited on 07/31/2025).

Tucker, Ian (Sept. 2021a). "(Re)Thinking Body-Technology Relations with Michel Serres: Emotion, Noise and the Emergence of Algorithmic Appropriation". In: *Media Theory* 5.1, pp. 219–230. ISSN: 2557-826X. DOI: 10.70064/mt.v5i1.904. (Visited on 07/27/2025).

– (Sept. 2021b). "(Re)Thinking Body-Technology Relations with Michel Serres: Emotion, Noise and the Emergence of Algorithmic Appropriation". In: *Media Theory* 5.1, pp. 219–230. ISSN: 2557-826X. DOI: 10.70064/mt.v5i1.904. (Visited on 07/27/2025).

Van Otterlo, Martijn (Jan. 2013). "A Machine Learning View on Profiling". In: *Privacy, Due Process and the Computational Turn*. Ed. by Mireille Hildebrandt and Katja De Vries. London: Routledge. DOI: 10.4324/9780203427644.

Vaswani, Ashish et al. (2017). "Attention Is All You Need". In: DOI: 10.48550/ARXIV.1706.03762. (Visited on 12/14/2023).

Zarkadakēs, Giōrgos and Don Tapscott (2020). *Cyber Republic: Reinventing Democracy in the Age of Intelligent Machines*. Cambridge,

Massachussets London, England: The MIT Press. ISBN: 978-0-262-54272-2 978-0-262-04431-8.

Zhuo, Terry Yue et al. (May 2023). *Red Teaming ChatGPT via Jailbreaking: Bias, Robustness, Reliability and Toxicity*. DOI: 10.48550/arXiv.2301.12867. arXiv: 2301.12867 [cs]. (Visited on 01/13/2025).

Zou, Andy et al. (Dec. 2023). *Universal and Transferable Adversarial Attacks on Aligned Language Models*. DOI: 10.48550/arXiv.2307.15043. arXiv: 2307.15043 [cs]. (Visited on 01/14/2025).

Zuboff, Shoshana (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile books. ISBN: 978-1-78125-684-8 978-1-78125-685-5.