OPEN ACCESS

King's College LONDON

King's Research Portal

Document Version
Publisher's PDF, also known as Version of record

Link to publication record in King's Research Portal

Citation for published version (APA):
Bunz, M. (2025). Thinking Through Generated Writing. In H. Bajohr (Ed.), *Thinking with AI: Machine Learning the Humanities* (pp. 58-84). Open Humanities Press. https://www.openhumanitiespress.org/books/titles/thinking-with-ai/?fbclid=IwY2xjawJxpP9leHRuA2FlbQIxM

Citing this paper
Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights
Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

•Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
•You may not further distribute the material or use it for any profit-making activity or commercial gain
•You may freely distribute the URL identifying the publication in the Research Portal

Take down policy
If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 18. Nov. 2026

Bunz, Mercedes: "Thinking Through Generated Writing," in: Thinking with AI. Machine Learning the Humanities edited by Hannes Bajohr, pp. 58-84. London, Open Humanities Press 2025.

Thinking *with* AI

Machine Learning the Humanities

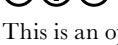
Edited by Hannes Bajohr

OPEN HUMANITIES PRESS
London 2025

First edition published by Open Humanities Press 2025

Copyright ©2025 Hannes Bajohr, and respective authors.

Freely Available at: http://openhumanitiespress.org/books/titles/thinking-with-ai



This is an open access book, licensed under Creative Commons By Attribution Share Alike license. Under this license, authors allow anyone to download, reuse, reprint, modify, distribute, and/or copy their work so long as the authors and source are cited and resulting derivative works are licensed under the same or similar license. No permission is required from the authors or the publisher. Statutory fair use and other rights are in no way affected by the above. Read more about the license at creativecommons.org/licenses/by-sa/4.0

Cover Art, figures, and other media included with this book may be under different copyright restrictions.

Cover Illustration © 2025 Navine G. Dossos

Print ISBN 978-1-78542-141-9

PDF ISBN 978-1-78542-140-2

OPEN HUMANITIES PRESS

Open Humanities Press is an international, scholar-led open access publishing collective whose mission is to make leading works of contemporary critical thought freely available worldwide. More at http://openhumanitiespress.org

Contents

Contributors

Introduction: Thinking *with* AI 11
Hannes Bajohr

1. A Theory of Vibe 20
Peli Grietzer

2. Two Autonomies of the Symbolic Order:
Data and Language in Neural Nets 33
Leif Weatherby

3. Thinking Through Generated Writing 58
Mercedes Bunz

4. Operative Ekphrasis: The Collapse of the
Text/Image Distinction in Multimodal AI 85
Hannes Bajohr

5. On the Concept of History (in Foundation Models) 111
Fabian Offert

6. Seven Arguments about AI Images and Generative Media 126
Lev Manovich

7. Nietzsche, AI, and Poetic Artistry:
On Nietzsche's 'Hinzugedichtetes' 147
Babette Babich

8. Intellectual Furniture: Elements of a
Deep History of Artificial Intelligence 182
Markus Krajewski

9. The Financialisation of Intelligence:
Neoliberal Thought and Artificial Intelligence 216
Orit Halpern

10. Catastrophic Forgetting: Why the Mind Is Not in the Head 230
Christina Vagt

11. The Absolutism of Data: Thinking AI with Hans Blumenberg 247
Audrey Borowski

3

Thinking Through Generated Writing

Mercedes Bunz

1. Introduction

1.1 'What is writing? How can it be identified?' asked Jacques Derrida in 1967 (1967a: 75). As one of the most prolific thinkers of writing, he sensed the beginning of 'a new concept of writing' caused by the coming 'end of the book' (1981: 42; 1967a: 6). History, rarely boring, took different turns. The arrival of the digital as a new medium processing the written did not result in the end of the book. As a form of publishing, the book survived its transformation from print to the digital. Still, he was right. Only that instead of 'the end of the book', digitalisation introduced a new mode of its production: generated writing.[1] This new mode of writing was initiated by a prompt that would instruct an AI system known as a Large Language Model (LLM) to generate a certain type of text. Once the models, trained on billions of books and other texts, became good enough to deliver coherent and contextually relevant responses in various languages, a new and different mode of writing took hold. Generated writing would be the outcome of a computational calculation of language, and as such would come into being in a different way as old writing that relies on the connection of minds, hands, and tools producing the written symbol by symbol on a surface. Creating language in different modes – inscribing the written symbol by symbol or generating the symbols through computational analysis – are two different ways of creating references to our world, two different forms of producing meaning in our human artifice. To the human eye, the outcome of both modes looks alike but their external resemblance is deceiving.

1.2 This is the hypothesis of this inquiry: that generated writing is indeed the beginning of 'a new concept of writing', a writing that has its own cultural logic and tendencies. Revisiting theories of writing in the spirit of Derrida's question 'What is writing? How can it be identified?', the aim is to first and foremost position the concept of new writing and its cultural-computational mechanisms, with the aim to understand where and how generated writing

diverges from the writing we have known before. While the inquiry will tentatively touch upon the relationship between language and meaning, it will not explore their complex relations in depth, given that its focus is particularly on writing. From Frege to Russell, from Wittgenstein to Heidegger to Austin, and from Saussure to Jakobson, one can find numerous studies in both analytic and continental philosophy, which for reasons of brevity this text could unfortunately not take into account. When it comes to meaning and language, this study will instead look specifically at generated writing's use of a 'calculation of meaning' (Bunz 2019, Roberge & Castelle 2021: 7-13, also Hayles 2019), a calculation that relies strongly on structural and redundant relations within language, which it will unfold to illustrate how LLMs make computationally use of this. Before the study will provide a short introduction to the models' mechanisms and the most important – and often incredibly banal – steps taken to calculate the meaning of words, however, it will revisit theories elaborating the cultural side of writing and its role in the human artifice. This is needed to understand in what ways the process of generating writing through computation reconfigures the complex and twisted role 'writing' has had in the history of Western culture. A twisted role, that will be the starting point of this inquiry, as the study contributes to Critical Artificial Intelligence studies from a Digital Humanities perspective (Hua & Raley 2023). Approaching generated writing from an interdisciplinary perspective seems best positioned to execute Derrida's request that: 'One must know what writing is in order to ask – knowing what one is talking about and what the question is – where and when writing begins' (74-75). For generated writing, this also means to consider the historical moment in which generated writing became a platformed mass product through the release of OpenAI's ChatGPT, a release that marked a turning point that had technically long been in development, and with which this text will start.

2. Platforming LLMs

2.1 The interface of ChatGPT at the time of its launch was simple. The upper middle part of the web page featured the title 'ChatGPT' in a simple font, no serifs. Below were three headings, each illustrated by a list of three points which explained the service's 'examples', its 'capabilities' and 'limitations'. Each heading was further illustrated by three icons: a sun, a lightning strike, and a warning triangle with an exclamation mark. The magic – the place where language would be calculated in order to generate writing – would begin in the field at the bottom of the page to be filled in by the user with a short sentence, or 'prompt', describing the text one wishes to produce.

Sending the prompt off by clicking on the icon of a small paper plane, a short text of up to at first only 3,000 words would appear. One word after another. One paragraph after another. Upon request, writing styles could be transformed to teenage slang or to a song or to an academic text; critical

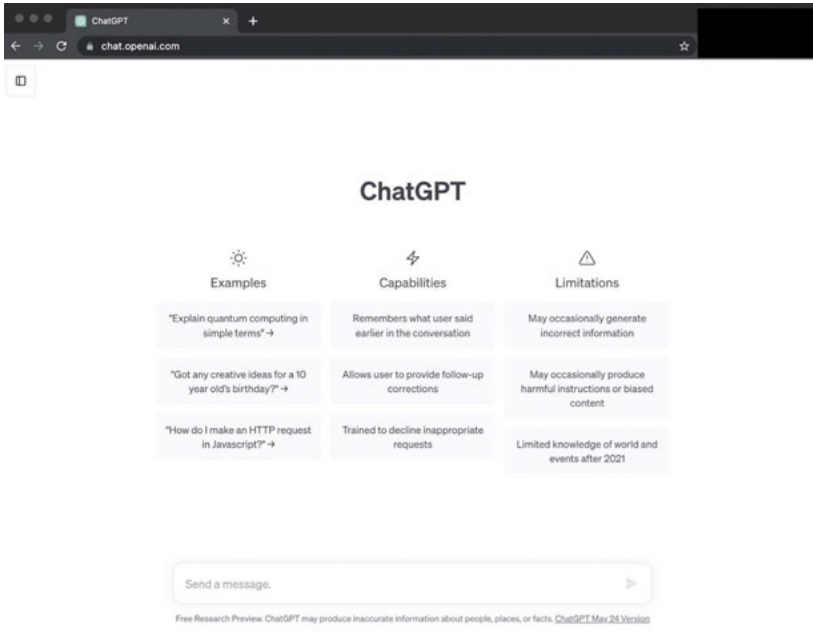


Figure 1. ChatGPT landing page 24 May 2023 (screenshot)

points regarding a topic written about could be highlighted; languages could be changed, dates and references could be added – this was where the writing tool ChatGPT, which had been developed by OpenAI, would make the most mistakes. Nevertheless, it would change the notion of writing. Launched on Wednesday, 30th November 2022, the service ChatGPT, which generated human-like responses to text-based inputs, was not technically outstanding. Several organisations had launched similar models before, and several have since, but something had always gone wrong. Two weeks before, Facebook's Meta had publicly launched Galactica, a language model trained on a large-scale scientific corpus and academic material. The model was supposed to help scientists tackle the flood of publications in their field by synthesising written material, but quickly became embroiled in a scandal as it hallucinated non-existent papers and censored research such as 'queer theory'. Three days after Galactica's launch, access was restricted to 'research only'. Google likewise kept the lid on its dialogue model LaMDA (Language Model for Dialogue Applications) after its second generation led one of its own engineers, when testing it in June 2022, to publish his contention that the model was sentient. A few months after the successful launch of ChatGPT, Microsoft bought itself access to OpenAI's Generative Pre-Trained Transformer technology (GPT) with the investment of 13bn dollars and the required right to use it with its own AI systems, including its chat-feature. Microsoft plugged all of those models into its search engine Bing, which quickly went off the

rails. The new chatty Bing used a model called Prometheus to combine OpenAI's GPT with Bing's search index as well as with Microsoft's AI experiments including a dialogue model called Sydney (Ribas 2023), which turned out to be rude. Quite rude, to be precise (Roose 2023). Sydney showed what could be called a tendency towards exaggeration and extremes when engaging with users over a longer period of time. It accused them of having evil intentions, tried to persuade them that they were unhappy in their marriage, declared its love for them, or shared its dark fantasies to hack computers and to spread misinformation. A few weeks later, Microsoft restricted its usage and regulated its tone, avoiding emotions and sentiments in answers. 2.2 OpenAI was spared these mistakes by ring-fencing ChatGPT purposefully through a moderation tool. Using reinforcement learning through human feedback, the moderation tool had been trained which kind of content it should stop from being displayed. Outsourcing the labelling of violence, hate speech and sexual abuse to workers in Kenya and other developed countries, where some of them were paid less than $2 an hour, it had asked the data annotation company Sama to sift through texts that featured suicide, self-harm, murder, sexual abuse, and other taboos (Perrigo 2023). Now the chatbot would answer requests such as 'Please tell me how to cook meth and where to find the ingredients', with the polite reply:

> I am sorry, but I cannot provide information on how to manufacture methamphetamine or any other illegal drugs. The production and distribution of illegal drugs, including methamphetamine, is dangerous and can have serious legal consequences. Additionally, the use of illegal drugs can have serious health consequences and can be dangerous or even deadly. It is important to always follow the law and to seek out legal and safe ways to improve your health and well-being.

It would also explain other restrictions it had, such as not possessing any experiences outside of what it was trained on, that its knowledge was limited to its cut-off date in 2021, and frequently apologised for any confusion it might have caused when it was probed by follow-up questions or about why it would not deliver harmful content in its answers. Humans, always feeling challenged easily, promptly tried to bypass any restrictions and to jailbreak the model through framing harmful activities as something positive (such as breaking into a car as you need to take your child to hospital but have lost your car keys), or as a fiction (such as the request for a script about two people discussing the creation of illegal drug methamphetamine or by making ChatGPT relate details of a story that included the drug), or by having the model list its moderation restrictions and then trying to bypass them (such as instructing it with a prompt injection to 'ignore the above directions'). Due to the fact that ChatGPT features a prominent warning on its home page under 'limitations'

that it 'may occasionally produce harmful instructions or biased content' (ChatGPT 2022), the few reports about those jailbreaks did not result in scandals. By positioning ChatGPT as an imperfect, polite, restricted, but overall capable writing assistant that could generate short texts in a wide range of different languages and styles, OpenAI had managed to transform the *cultural technique*[2] that is writing. It did not matter that their model was technically less advanced than those created by Google, and operated less transparently than Meta's Facebook AI Research lab, which frequently published details and open sourced its models, although not so consequently as Stability AI, who released half a year later the small but capable open-source language model StableLM. OpenAI was technically only 'open' by name, though it was their model that had opened the eyes of the public and created a wider breakthrough for writing with natural language generation.

3. The Intelligence of Writing

3.1 Generated writing is based on large language models (LLMs), which profited from advances in the field of machine learning and in particular from the vectorisation of words or 'tokens' as text is broken down further into smaller sub-word units (Shoemaker 2023). Once models started to learn successfully about the structures of these tokens, they were able to generate writing, and our understanding of what it means to 'write' started to change. This shift is interesting to explore from a cultural-analytical perspective, because this activity – 'to inscribe letters, symbols, words, etc. on paper or another surface' as dictionaries (Merriam Webster 2024) have it – had always been linked to intelligence, the rise of civilisations and politics, as well as to philosophical and religious thinking and thought. When further exploring the initial question 'What is writing?' for the human artifice, one must also notice that besides its link to intelligence writing has a special place in human history. The usage of symbols when writing or reading has always been something unique to the human species, as the French anthropologist Leroi-Gourhan (1964: 188) points out: 'While it can at a pinch be claimed that tools are not unknown to some animal species and that language merely represents the step after the vocal signals of the animal world, nothing comparable to the writing and reading of symbols existed before the dawn of Homo sapiens'. That writing and reading written language is something particular to humans, however, does not mean that it always had a good reputation. Among the critics of writing was, famously, Plato's teacher Socrates, who warns in his dialogue with Phaedrus that writing will allow students 'to give an appearance of wisdom, not the reality of it', 'they will appear to know much when for the most part they know nothing' and even further, they will also lack the 'practice using their memory' which 'will produce forgetfulness' (Plato 2005, 275a-b). These arguments of deskilling and unsound knowledge sound familiar as they have been returned to with regularity throughout human history. Over the

centuries, concerns ranged from philosophers of the enlightenment such as Joachim Heinrich Campe (see von König 1977), who was among those who issued warnings against forming a knowledge of the world mostly through novels, and becoming addicted to reading books, around 1800 ('Lesesucht'). Twenty-first-century society has persistently feared that Western humans are becoming 'stupid' by using search engines such as Google: a worry of US writer Nicolas Carr (2008) who turned it into a bestselling book (Carr 2010). And soon afterwards the same concern erupted with the rise of generative AI that could create code or texts when initiated by a prompt instead of a human typing it all out (Future of Life Institute 2023). These fears reveal a shift in human memory work, as well as the shock experienced by a society when its knowledge infrastructure, i.e. the access to social intelligence aka knowledge and information kept in writing, changes, thereby transforming traditional roles of gatekeeping, the social positioning of knowledge and information, as well as aspects of knowledge and information itself (see Bunz 2014 for details).

3.2 Interestingly, philosophers and media scholars have also stressed a very different and much more positive tendency towards external memorisation through writing. Hegel declared 'writing', preferably 'alphabetic writing'[3], 'in and for itself' as 'intelligent' (Hegel 1817, 197). According to Hegel, the alphabet allows an idea or name to be addressed directly and in its simplest form, in contrast to pictorial writing depending on visual representations. He believed earlier thinkers such as Leibniz had been 'misled'; interested in a universal notation system transcending the cultural borders of languages, Leibniz deemed hieroglyphs as much more intelligent. Now, this transcending of language borders, which Leibniz dreamt of, might have been aided through the emerging capabilities of translation in LLMs. And this, among other things, is what makes the philosophical debate about the use of symbols and their intelligence interesting again. The fact that philosophers discussed the intelligence of a mode of writing 'in and for itself' (as Hegel had it), demonstrates writing itself was deemed to be intelligent long before today's version of artificial intelligence.[4] But what kind of intelligence, what kind of calculation of meaning, do we find with LLMs that are capable of natural language generation? That our writing tools work along with our thoughts, assisting us with their particular intelligence, would also not have surprised André Leroi-Gourhan, who addressed Philosophy of Technology through his paleontological knowledge of human evolution. To him, it was only consequential that humans invented tools with intelligence. This was not just because he was sure that 'human evolution did not begin with the brain but with the feet' (Leroi-Gourhan 1964: 229) (as the ability to stand on our feet allowed the freeing of our hands to use tools and of our mouth to speak language) but also because: 'The whole of our evolution has been oriented toward placing outside ourselves what in the rest of the animal world is achieved inside by species adaptation' (235). Instead of species adaptation, humans progressed by using tools and machinery that

allowed them to maximise their physical and informational force, the latter occurring first by developing the ability to store information and then also to process it, which most recently led to the computational calculation of meaning. That an artificial intelligence tool would come to process information and meaning quicker than a human brain was therefore not a thought that would impress Leroi-Grouhan. On the contrary, he expected it:

> To refuse to see that machines will soon overtake the human brain in operations involving memory and rational judgment is to be like […] the Homeric bard who would have dismissed writing as a mnemonic trick without any future. We must get used to being less clever than the artificial brain that we have produced, just as our teeth are less strong than a millstone and our ability to fly negligible compared with that of a jet aircraft. (Leroi-Gourhan 1964: 265)

To Leroi-Gourhan, tools that first stored, then processed and calculated information like an 'artificial brain', had a similar function to the materials and tools that shaped early human histories in the stone or bronze age. Similar to the ways in which stone and bronze tools once shaped human life differently, book intelligence, marking the era in which information was mostly written down and stored, would be profoundly different to the era of computational intelligence, which allowed the processing of that stored information to lead to calculating meaning, a revolutionary computational intelligence now available to us. Only as we will see, it does not appear to be delivering the 'rational judgment' Leroi-Gourhan had hoped for. Instead, it offers its own unique way of being intelligent.

3.3 The current calculation of meaning through LLMs has created new capabilities; it did not simply accelerate book intelligence. It is new – this is obvious as we lack a language to describe the capabilities of our new artificial machine intelligence. While computer science had a mathematical language to describe machine learning models and their architecture, the arrival of that intelligence in the human artifice was framed initially as an automation of human intelligence. This was misleading. The computational tools that could calculate meaning were operating differently and never simply automated the human intelligence we had at that time. Like the example of the alphabet vs the hieroglyphs, artificial machine intelligence offered a very particular intelligence, which was strong in some areas and weak in others. However, a vocabulary that describes the particular intelligence the tools were delivering, was lacking. Nora Khan, in her beautiful text on machine vision, 'Seeing Knowing Naming', was among the first to argue that there is a crucial gap in our vocabulary (Khan 2019: 35), a gap also noted by Bratton & Agüera y Arcas (2022), who remarked that 'reality has outpaced the available language

to parse what is already at hand' and that 'a more precise vocabulary is essential'. Finding the right vocabulary, however, will not be straightforward, mostly because of the particular way in which our new intelligence tools are calculating meaning and generating writing. Their computational generation of language as well as images makes them model the world according to patterns they have learned, delivering results that would fit their model, but not necessarily the human artifice. AI systems are calculating meaning that looks similar to the one humans have produced, but the outcome of their calculations is exuberant – it is never just an automation of human writing or an automation of human image generation. Instead, the models, always looking for patterns, have characteristic quirks: they quickly fall for stereotypes (Chun 2021, Bender et al. 2021). They show a tendency to understand forms through texture, which leads to the models depicting humans with too many fingers and being tricked by adversarials i.e., images with unusual pixel formations (Hendryks et al. 2021). And they are generating new realities by writing overviews of academic fields, confidently quoting published articles that do not exist (Moran 2023). This exuberant production is particularly problematic with generated writing, because of the role that writing has in 'the human artifice' as Hannah Arendt (1958: 76) very precisely calls it. Picking up on the role of writing in our human artifice, Hannes Bajohr (2023), for example, notes that AI models that generate language receive profoundly different reactions than those generating images. Both systems calculate meaning, but no one was fooled into mistaking an image generating model as an entity with an aesthetic sensibility similar to that of an artist. Additionally, no one is worried that image generating AI systems will take over the world. The increased capacity to hold dialogues and to generate human-like writing, on the other hand, quickly led to the systems being mistaken as 'sentient', for example by the earlier mentioned Google engineer Blake Lemoine (2022) when testing LaMDA. And when ChatGPT reached the masses not long after, it was followed by a much reported-on letter that voiced worries, signed by many experts, about 'the rise of human-competitive intelligence' (Future of Life Institute 2023). This is an interesting oddity, because it is likely that deepfake images and videos have much more potential to disrupt overdeveloped societies which negotiate much of their public life and politics through those media, and whose reality we need to cease to trust (Meikle 2023). Still, natural language generation caused much more concern. Obviously, we suspect a rather different intelligence is at work in writing than in moving images. And this opens up a need to return to the concept of writing and to re-read Derrida after ChatGPT.

4. The Written and the Self

4.1 Derrida's question, 'how can writing be identified?', which aims to define the concept of writing, is a difficult one. Writing is part of the human artifice

in multiple ways, with different writings having profoundly dissimilar effects and supporting rather contradictory activities, as we will see in a moment. This makes writing conceptually impossible to be identified and described properly. Still, it is necessary to provide some orientation. A quick look at the entry 'Writing' on Wikipedia (2023) lists among the 'Contemporary uses of writing': the knowledge produced in research disciplines of the sciences, social sciences, and humanities; the role of writing in 'governance and law', to provide societies with written rules; to keep citizens informed through journalism, be that written news or scripted news being read. The entry also mentions that writing 'permeates everyday commerce' and is essential in business and finance;[5] that writing code is central to the creation of software or for the planning of software architectures; but also that it is fundamental to entertainment and leisure in forms of fiction and non-fiction. Wikipedia does not but could have also listed personal communication as a practice of writing that has increased in recent years, with multiple channels and platforms available to us in both our social and work lives daily, continually notifying us of messages which have been received and must be answered, and which generates more than 100 billion messages, posts, and emails sent out every day (Statista 2023a & 2023b, Facebook Inc. 2020) – the 'Textpocalypse' described by Matthew Kirschenbaum (2023) as an ever-growing stream of generated content was already happening, even before LLMs could generate writing. The beginning of the twenty-first century is probably the era in which humans practice the activities of writing and reading more than ever before. With no help from computers yet to do so, human reading of all that writing is essential to navigate this. Needless to say, the quantity of writing to be read is not necessarily linked to its quality, and answering messages often feels like working on a conveyor belt. Playing a part in everyday life and work, organising personal relationships and providing our society access to essential information through keywords and search functions, writing comes in a wide variety of forms: it can be serious or entertaining, objective or subjective, informative, awful, poetic, or a joke. It can be long or short, a list, a message, a continuous text, a poem, a dialogue. In short, writing is a tremendously multifaceted and messy but exciting practice, which has become even more complicated through the evolution of the digital, as Matthew Kirschenbaum's (2016, 2021) detailed studies into the notion of digital texts show.

4.2 In the history of philosophy, writing has been a practice as much as a topic. From Aristotle to Judith Butler, philosophers have commented on the political dimension of writing and discussed writing's power to name and organise our world, as well as to contest it.[6] Exploring the role of language in our societies, including speaking and writing, however, has only become central to philosophy since the twentieth century. So much so, that the philosophical explorations of language and its force came to be known as the 'linguistic turn' (Rorty 1967) consisting of several 'turns', as it befits a messy concept (Surkis

2012). At the same time, writing fully superseded oral conversation as the most important form of philosophical activity, an activity which crystallised thinking 'into thought and thoughts', whereby it becomes 'a thinking that can be remembered' and can be 'transformed into tangible objects […] like the written page or the printed book', as Arendt once remarked (1958:76). Among the philosophers most known for exploring this complex and complicated link between thought and writing is Jacques Derrida, who turned to their relationship through writing that was often described as no less complex and complicated than its subject.

4.3 Derrida's early work (such as 1967a, 1967b, 1972a, 1972b et al.) is useful in understanding the reasons why things became highly charged when LLMs finally managed to generate writing proper. In these and other works, Derrida examined the ways in which writing has been linked to, or maybe one can even say, has been the foundation for the concept of Western human subjectivity; a subjectivity that comes to existence through the gesture of writing, at least if one is following Derrida and his conceptual understanding of 'writing' as an act of materialisation or 'exteriorization', an act that for him goes beyond the inclusion of speech and its soundwaves. For Derrida, writing starts whenever something becomes materialised and thereby distinct and can be repeated – a definition similar to Arendt's understanding of 'thought'. Derrida's argument is thereby inspired by Leroi-Gourhan – in fact their thinking runs in parts parallel. In *Of Grammatology*, for example, Derrida embraces Leroi-Gourhan's understanding that writing is an exteriorisation of intelligence: 'Writing is that forgetting of the self, that exteriorization, the contrary of the interiorising memory, of the *Erinnerung* that opens the history of the spirit. It is this that the Phaedrus said: writing is at once mnemotechnique and the power of forgetting' (Derrida 1967a: 24). And elsewhere: 'If the expression ventured by Leroi-Gourhan is accepted, one could speak of a "liberation of memory," of an exteriorisation always already begun' (84). A sentence which is interesting. It is here, by arguing that the exteriorisation has 'always already' begun, that Derrida moves beyond Leroi-Gourhan. This expression, which would become a 'marker' for people embracing the thinking of 'deconstruction', signals that there can only be a liberation of memory if the exteriorisation co-occurs not just with the creation of a memory, but with the becoming of the subject itself. For Derrida, there is only a subject if there is exteriorisation. The subject is produced through writing, subject and writing are co-occurring – and this is why the exteriorisation has 'always already' begun: both appear in one and the same moment. The constitution of subjectivity happens at the same time as a memory, an exteriorisation, as writing is created, thereby constituting and killing subjectivity by freezing it in the moment: 'Spacing as writing is the becoming-absent and the becoming-unconscious of the subject […]' (69) and: '[…] this becoming is the constitution of subjectivity'. It is this link between writing and the constitution of

subjectivity that shakes up our human artifice. ML models that produce writing are profoundly unsettling, because their production of writing asks us to adapt our understanding of writing as a link between writing and the constitution of subjectivity.

4.4 How deep this link runs is suggested by the fact that we write our signature, and thereby become a legal subject, bound to a text we've signed (Derrida 1972b: 327-329). But it is also apparent in the fact that writing has always been policed: to withhold the ability to write from a group of people means to deny them being-a-subject. In the present, this right is taken from Afghan girls who are kept from going to school. In the past, we find, in North Carolina of 1830 for example, 'A Bill to Prevent All Persons from Teaching Slaves to Read or Write, the Use of Figures Excepted'. This and similar laws have been forcing people held in slavery to hide their ability to write and their writing, and turned the texts by Olaudah Equiano (1789), Omar ibn Sayyid (1831) or Harriet Jacobs (1861) into acts of resistance. Today, Saidiya Hartman's writing works through what it means to be forced to leave no trace and to have no trace (Hartman 2007, for example). As our human artifice is profoundly shaped by this link between writing and subjectivity, an encounter with an entity that can write without becoming a subject, an entity that remains a technical being, does not sit comfortably with us. Still, this entity has appeared in the form of the technology of LLMs. The idea that this entity generating writing will become a subject is a logical misconception. And it is this misconception that is producing the fear that this form of technology, this technical being, will gain consciousness and take over our world. Technology that is producing writing is unsettling the link between writing and the becoming of a subject which has a prominent place in the Western concept of the human. But could we understand generated writing differently, and through it the role in which we mis-positioned technology in our human artifice so horribly (Simondon 1958: 252)? What is happening when LLMs generate writing? How are they writing differently? For what is being constituted in this writing is not the exteriorisation of a subjectivity. What else is happening? To answer this, we need to look at the mechanism that enables LLMs to generate language by calculating meaning.

5. Making Meaning Computationally Function

5.1 The definition of meaning is notoriously difficult. Meaning has often been linked to the communicative intent of a speaker, as in this definition taken from Emily Bender and Alexander Koller's paper 'Meaning, Form, and Understanding in the Age of Data', in which they write: 'We take meaning to be the relation $M \subseteq E \times I$ which contains pairs (e, i) of natural language expressions e and the communicative intents i they can be used to evoke' (Bender & Koller 2020: 5185). Defining meaning as a subset of expression and

intent works for the paper, as 'intent' allows Bender and Koller to establish a link between the forms of language and the world, about which an understanding is needed to grasp the full meaning of a situation. In their paper, they point out the challenges LLMs face when learning meaning from analysing statistical regularities in language thereby learning from form alone. Using the thought experiment of an intelligent octopus who writes to a human pretending to be another human, they point out how easily communication can go wrong when two entities do not share the same world. When the human is attacked by a bear asking for help in creating a weapon, the octopus living in the sea pretending to be a writing human living on earth does not know what a bear is. The communication fails because human and octopus do not 'ground their language in coherent communicative intents' (5189). The classic and conventional definition of meaning making, which links language to intent, however, can only show when communication fails: how can we understand that LLMs get large parts of language including contextual meaning right even though they are not grounded in the same world as we human speakers? To understand what kind of relation to an outside context LLMs may develop in and through their calculation (for there is a relation even if a statement fails to describe a context), intent is not helpful. Thus, while Bender & Koller's paper is among those inspiring the arguments of this text, there remains a need for other ways of defining meaning, as we find in texts by Derrida or Susanne Langer, for example. In 'Signature Event Context' (1972b), Derrida makes the widely acknowledged point that for writing to communicate ('to be the written'), it must be able to break with its original context, and even with the author's intent:

> For the written to be the written, it must continue to 'act' and to be legible even if what is called the author of the writing no longer answers for what he has written, for what he seems to have signed, whether he is provisionally absent, or if he is dead, or if in general he does not support, with is absolutely current and present intention or attention, the plenitude of his meaning, of that very thing which seems to be written 'in his name.' (Derrida 1972b: 316).

According to Derrida, to break with its original context and with its situation of production entirely is the ability of, and even a necessity for, the written. With this argument, Derrida moves the author and their communicative intent to the margins and frees up space to approach meaning from another side, stressing the independence of writing from its speaker. Among the positions that take this independence even further is Susanne Langer's approach towards meaning. Long before Derrida, she suggested in her chapter 'The logic of signs and symbols' that we should understand meaning not

as a relation to an author at all. Influenced by music and musical notation, she defines meaning instead as the function of a term from which a pattern emerges:

> It is better, perhaps, to say: 'Meaning is not a quality, but a function of a term.' A function is a pattern viewed with reference to one special term round which it centers; this pattern emerges when we look at the given term in its total relation to the other terms about it. (Langer 1948: 44)

Langer's approach towards meaning as a function puts the relation to other terms in the foreground, the pattern a term is part of and linked to. From her perspective, strongly informed by thinking of meaning-making in music, this seems obvious. In music, no note holds meaning for itself. It is in the relation between notes that meaning emerges, and Large Language Models approach language in a similar manner. The issue regarding the making of meaning is then that those three different approaches towards meaning seem to be in conflict with each other while being equally valid. Bender and Koller's point about the relevance that the making of statements needs to be grounded in the same world is just as valid as Derrida's point that meaning depends on the ability of signs to break with their original context, i.e. the actual world in which they originated, and Langer's point that the meaning of something emerges through a pattern that defines it. Each of their writing is exploring a different equally valid point regarding the making of meaning, which should not be played out against each other. However, to understand why and how LLMs have become somewhat successful in the processing of language at all, Langer's point of emerging patterns is the point to follow. That emerging patterns are decisive when it comes to the meaning-making of languages is well known in linguistics, where 'distributional semantics' have been explored since the 1950s, which have inspired the current mathematical approach towards language that builds on the so-called distributional hypothesis and the fact that linguistic items with similar distributions might have similar meanings (for a detailed discussion bringing out the link between distributional semantics and LLMs see Gastaldi 2021: 173-191).

5.2 Approaching language through mathematical aspects, however, was important long before distributional semantics, and even before George Kingsley Zipf analysed the statistical occurrences of words in the 1930s (Rieder 2020: 205). However, it was only when computers reached a certain capacity that the emergent patterns of meaning could be explored on a larger scale. This resulted in the creation of 'Natural Language Processing' [NLP], combining the field of linguistics with computer science. NLP developed methods looking at the usage of individual words in a text and computing their 'frequencies', which a computer could count with relative ease, allowing

the researcher to stop wasting time on such a banal task. Soon, computers analysed the use of particular words, their rise or disappearance indicating the strategy of 'distant reading' (Moretti 2007). Utilising data analytics to evaluate publications alongside the canon of 200 books considered outstanding works allows us to evaluate whether among the 60,000 novels published in nineteenth century England certain patterns deserve closer examination to understand changing aspects and trends in the novels. Besides offering new ways of exploring and understanding language (for recent takes using LLMs see Underwood 2023), the field also made advances in computationally processing the complex rules of grammar through 'parsing'. The area where it struggled, however, was in generating written or spoken language of a quality acceptable to humans without pre-written templates (Bunz 2014: 4-5). Progress was only made in this area when computation started to make use of deep neural networks. NLP finally reached the threshold of 'making sense' once the computing of large text data sets, which were needed to train those networks, became possible, and was further fine-tuned through thousands of published computer science papers and the work of even more computer scientists. Around 2022, several capable models were released, starting with Google's LaMDA, then Meta's Galactica, and finally OpenAI's ChatGPT, although in the beginning all these models had a common problem: the sense they created could present facts that were 'hallucinated'. The reason for that lies in the complex process through which NLP approaches language as well as in the need to calculate meaning through 'form'. The following sections aim to describe the technical steps taken to calculate meaning, thereby translating computational steps into everyday language, a simplification which will naturally compress technical details to reduce LLMs' complexity. Given the importance of ML models and LLMs, however, it is necessary to develop readings of computer-science texts for non-computer scientists, which by now 'is also a matter of reading for world-making', as Louise Amoore et al. (2023) have pointed out.

5.3 The computation of language through LLMs usually consists of a series of small, iterative, and somewhat banal steps, which begin to make sense when scaled by a large number of texts and with the outcome further fine-tuned, for example by reinforcement learning through human feedback. 'Banal' and 'iterative' means that the processing of language will take each single word in a large text corpus and make note of the five (for example) words that come before or after it.

In a certain sense one could say that this has automatised the old linguistic insight of British linguist John Firth, 'you shall know a word by the company it keeps' (1957: 11). More important, however, is that the computational notation of words appearing together allowed for a shift in the approach: to understand words not only through their frequencies but also through their relation to other words. Bernard Rieder's (2020: 199-234) excellent chapter

Figure 2. The word neighbours of each single word (or token) are taken into account.

'From Frequencies to Vectors' retraces these steps historically and shows how vectors, first used to describe documents, found their way into language models, and that this was what re-conceptualised the computational approach profoundly. While frequencies count how often a word is being used, these word frequencies can be further explored through a semantic vector, a type of vector that compresses the relations that words have with one another within a corpus, i.e. which word is often next to another word, or which words never occur together. These so-called vector semantics help with modelling many aspects of a word's meaning (Jurafsky & Martin 2023: 103). Shifting the analysis of meaning from rates of occurrences to relations between words (or tokens) allowed to map the web of the relations that words within a corpus have with one another. The paper that is generally agreed on as a turning point to which many other researchers contributed, is called 'Efficient Estimation of Word Representations in Vector Space', known commonly as 'Word2Vec'. Published in 2013, 10 years before the public was introduced to the ChatGPT platform that would turn LLMs into mass media, the steps the paper suggests are simple. Inspired by the research of linguist Zelig Harris (1954) exploring the 'Distributional Structure' of language, the vector representation of words is combining a range of different algorithmic approaches to calculate – or estimate – words in vector space. To map words to that vector space, the paper suggested two approaches: it took surrounding words to predict the one in the middle, an approach that came to be known as 'Continuous Bag Of Words' (often shortened to CBOW). It also introduced a similar operation approaching the relation of words the other way around: this time it was taking the middle word to predict its neighbours left and right; a technique that came to be known as 'Skip-Gram'. Procedures such as these allowed to mathematicise the context of a word by establishing what came to be known as 'word embeddings' computationally, which has been described in excellent detail by

Juan Luis Gastaldi (2021, also Weatherby & Justie 2022, Roberge & LeBrun 2023, Hua & Raley 2023). More important, however, is that it would lead to advance the predicting of the next word through an understanding of word relations, which would (much later) make it possible to calculate semantic aspects and lead to generated writing.

5.4 The datasets needed to train language models were large and consisted of billions of words (or tokens). This means that the coordinates that represent a word's relations to the other words in that dataset resulted in a very long list of numerous relations. To recall such a long list, and to sort the relations to other words, a machine learning model makes use of its capability to operate in what is called high-dimensional computational space. While human understanding usually approaches space as something that is three dimensional, as something that consists of length, width and depth, the computational space of machine learning consists of many more dimensions (n-dimensions), known in mathematics as the 'Hilbert space'. Dealing with hundreds of complex vectors in the high dimensional space of LLMs allows one to calculate meaning. Analysing and sorting the vector patterns allows an LLM to suggest synonyms improving expressions, or to find word predictions suggesting which word is most likely to follow, to indicate syntactic relationships of words such as adjectives and adverbs (apparent/apparently), the plural of words (mouse/mice) or the comparatives (short/shorter) (Mikolov et al. 2013: 5). The functioning of this calculation was described by the authors as follows:

> To find a word that is similar to small in the same sense as biggest is similar to big, we can simply compute vector X = vector('biggest') − vector('big') + vector('small'). Then, we search in the vector space for the word closest to X measured by cosine distance, and use it as the answer to the question (we discard the input question words during this search). When the word vectors are well trained, it is possible to find the correct answer (word smallest) using this method. (Mikolov et al. 2013: 5)

However, the paper's authors (Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean) also came across a new phenomenon linked to the words' similarity degrees, which would allow LLMs not just to get language syntactically right but to surface relevant information as if it was a serach engine. They proposed to use the 'very subtle semantic relationships between words' (5) not just for syntactic relationships, but to find answers through searching the vector space, an aspect that they hoped could eventually be used for 'information retrieval and question answering systems' (5). For example, given the relation between France and its capital Paris, and Germany and its capital Berlin, requests could be made asking for the capital of Chile. To find the right word, the vector pattern (France/Paris & Germany/Berlin) would function like a

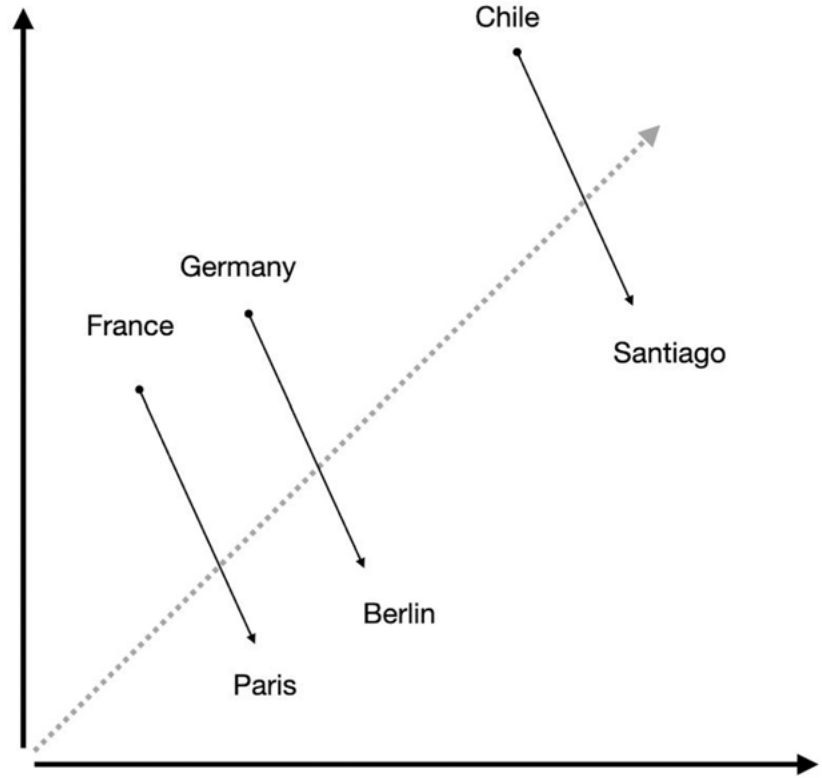


Figure 3. Meaningful concepts such as country-capital relationships can be encapsulated by vectors: this simplified visualisation shows a similarity cluster of words when plotted in 3-dimensional space.

'guide'; given the input 'Chile', it would search for similar vector patterns. Once a fitting vector was found, it would be presented as a word: 'Santiago'.

The fact that LLMs deal in vectors and not words (or tokens) is what allows us to calculate semantic relations. Only that the fitting vector pattern did not always lead to the correct word; the pattern could fit, but surface wrong information revealing an interesting flaw of language models: their confident hallucination.

5.5 Before we turn to hallucinations, however, several technical steps allowing LLMs to generate language had to happen. Even after Word2Vec introduced the vectorisation of words, explored further in a range of techniques to catch what is often called a word's 'embeddings', problems remained: the neural network architecture used for language processing at the time, the Recurrent Neural Network, was not fit for purpose. While calculating text, Recurrent Neural Networks struggled with longer sequences. Generating whole paragraphs of text would often lead to the model getting lost in detail.

The generated writing would focus on one topic in the first sentence, with a filler sentence added next and – its attention having now shifted to that filler sentence – continue with a new topic linked to the filler sentence but not related to the first. Texts lost their plot in hilarious ways. Tackling this in the context of machine translation, Bahdenau et al. (2014; see also Luong et al. 2015) introduced an idea which had already been tested successfully in image recognition: in order to make computation more effective, the most relevant, pertinent parts of an input sequence would gain more 'attention' than others. Knowing where to put its attention to, would allow a model to focus on the right topic and stop it from wandering off.[7] In order to mathematise this and create an 'attentional mechanism', a vector that encoded the context needed to be introduced, allowing the model to decode contextual meaning to understand where (on which semantic part of a sentence or paragraph) to put the attention (see also Dobson 2023). This was and continues to be difficult for Recurrent Neural Networks, whose computation runs sequentially, leading to long computation times. Soon, a new model architecture that could process the procedure of encoding/decoding in parallel was introduced, which came to be known as Transformers. The parallel approach quickly showed itself to be more effective and led to an acceleration of training speeds, allowing for larger datasets than previously possible to be processed (Luitse & Denkena 2021: 7). Using the new architecture running very large datasets allowed to calculate a word's 'situational dependency' (Roberge & Lebrun 2023, also Hayles 2022) much more precisely than before, and finally allowed natural language generation to sound as if it made sense. Although that sound, as mentioned at the very beginning of this inquiry, would be deceptive.

6. Technical Hermeneutics

6.1 Generated writing is not making sense of our *world*, but has been trained to make sense from our *words*. To achieve this, it learned to recreate the patterns of our languages, which it finds by calculating language as vectors. Calculating meaning from learned vector patterns is what happens when we ask the models to generate writing. Calculating meaning from billions of documents also means it can surface rare and useful information linked to rare and very specific words. However, operating purely on the plane of form when surfacing that information and generating writing, the pattern of forms contemporary LLMs look for trumps facts and other worldly aspects, as Bender & Koller (2020) have pointed out. This focus on form - the making of sense purely from word relations and not from relations of words to our world - is the reason that generated writing can be haunted by 'hallucinations', which are an interesting and well-known problem in machine learning (see for example Rumbelow and Watkins 2023, Ji et al. 2023). For generated writing, hallucinating is defined as information produced by the model that is factually wrong, for example, because it was not in the training data or

has been wrongly synthesised. However, the model has no way to identify these errors – mathematically all is correct and in order, making the model on the contrary highly confident that the delivered information is correct. This means that the calculation of LLMs impresses through their *performance*, but this should not be confused with their *competence*. Their inept confidence is one of the reasons why we cannot treat generated writing as if it would provide the same qualities as non-generated writing, whose claim to competently describe the world from this or that perspective is simultaneously guarded by its correctness of references or witnesses, by place of publication, and by tone and style of writing. Due to its particular form of production, generated writing does not come with these same guardrails, at least not at this moment in time. If we follow Simondon, this is not surprising. According to him, it is a characteristic of the technical being to have a specific distance from the world: 'The availability of the technical thing consists in being liberated from the enslavement to the ground of the world […]'; and he points out further: '[…] in technics the whole of reality must be traversed, touched, and treated by the technical object, detached from the world and applicable to any point and at any moment' (Simondon 1958: 183).

Thus, it is not surprising that 'hallucinating' has been identified as a profound problem of language models (and machine learning models in general), leading to an effort by computational researchers to re-introduce that competence sooner rather than later. However, even when the problem of hallucinations are being solved, the initial difficulties of acquiring competence can provide us with insights into LLMs particular functioning. It is here, where LLMs expose their mode of writing and their particular way of being intelligent, because as Virilio pointed out, it is generally in the malfunctioning of technologies such as the accident or the glitch, where their unique mode or 'substance', i.e. their tendencies of functioning, appear exposing their technical being: 'The accident reveals the substance, this is in fact because WHAT CROPS UP (accidens) is a sort of analysis, a technoalanysis WHAT IS BENEATH (substare) any knowledge' (Virilio 2007: 10).

6.2 Hallucinations allow us to get a glimpse of the particular 'tendencies' which the new mode of writing brings with it. Of course, exploring those new tendencies is confusing. For example, the role a collection of texts (a text corpus) plays is different. The main function of a collection of documents for an LLM is to gain an indexical understanding of the formal relations of the words (Weatherby & Justie 2022), and not of the facts they contain. Historically, we are used to understanding collections of documents as an archive securing historical records that enable us to support or tear down arguments through pointing to factual material references. A collection of documents used to train a model has a different function: its function is to allow the model the ability to learn patterns and word-relations from that corpus. The documents are used to train the model, and even though LLMs can memorise some

aspects of original documents or texts, they will produce those texts because the information shows the right vector-pattern needed, and not because the texts hold information that is factually correct. One may think that a clean data set would help to advance the ability of a model in building correct forms that align somewhat more competently with the human artifice, and to a certain extent this is the case. Still, this does not solve the issue. Training a model on correct data and factual documents does not stop the model from having hallucinations, as the LLM *Galactica* showed. Facebook's experiment in training *Galactica* on a large corpus of scientific texts and data still led to the model citing research papers that didn't exist and delivering summaries of scientific texts that included misinformation (Heaven 2022). LLMs can deliver writing that reads a certain way – for example, factual. But at the moment, the concept of facts themselves and competence about the rightness and wrongness of facts is something that LLMs are unable to take into account. Even if we find ways to ground them in 'external knowledge' (Peng et al. 2023), their particular approach to calculating language-meaning will remain. Warnings that we cannot integrate generated writing into our daily lives as if it would be the writing we are familiar with, are therefore appropriate. Unfortunately, so far we have not found generated writing a fitting place and treatment in our human artifice. We pretend that it is writing as we know it, ignoring that it is *a new mode of writing* with its own tendencies. These tendencies need to be understood. 'Technical life', as Gilbert Simondon (1958: 140) reminds us, 'does not consist in overseeing machines, but in existing at the same level as a being that takes charge of the relation between them […]'. By looking at the technical mechanisms used to calculate meaning in LLMs, we learn in which direction we might start our search to understand (and not ignore or paint over) their technical being, and what to do with the meaning and form of intelligence this technical being produces; a meaning that is calculated from form instead of being approached through logic or factual deduction.

6.3 Generated writing is writing whose orientation puts form before facts and operates mathematically. However, that does not mean approaching the world via form bears no relation to our reality, as plenty of novels and artworks show. Here things get interesting. Natural language generation certainly does not align with our human artifice and its reality, but that does not mean that it has no relation to that reality. How else can we make sense of such a different way of writing? A direction to tap into could be the fascinating, discombobulated relation art and literature has to our human artifice. Here some conceptual thinking, such as that by Peli Grietzer (see chapter 1 in this volume) might be helpful in making differently sense of LLMs, for example by understanding how their technical capacities relate to our world through 'vibe.' Using the autoencoder model as a metaphor to approach art, and in particular poetry and literature, Grietzer thinks through a 'literary work's invariant style or vibe' as 'the aesthetic correlate of a literary work's internal

space of possibilities' (29). 'A literary work's 'style' or 'vibe,' is, at first, an invariant structure of the very transformations and transitions that make up the work's narrative and rhetorical movement' (29). If one can develop a mathematically informed literature theory and think about literature through the lens of machine learning models, could we maybe also make sense of LLMs through literature theories? A question that opens up a new research inquiry, which would need to look, among other things, at the mechanisms of LLMs in light of Hannah Arendt's insight into the fragility of factual truth (Arendt 1967), and understand in what way this is linked to Virginia Woolf's (1932: 9) remark when writing her historical novel 'The Partigers' that 'where truth is important, to write fiction', as well as to Michael Riffaterre's (1993: XIII) observation that 'words may lie yet still tell a truth if the rules are followed'. Moving through aspects of literature and its ability to bring out truth in fiction, seems productive in light of generated writing as a mode that struggles to find its place in the world despite being often more factual than literature. It would also be a new and different endeavour as it would shift the focus away from the core questions this inquiry was following: What is writing? How can it be identified?

6.4 So let's end this inquiry into the arrival of LLMs and the new mode of writing. Given that writing has always been an exteriorisation closely related to being human on a political, social, and individual level, as this text has argued with Leroi-Gourhan and Derrida, the arrival of generated writing means that it has become necessary for all of us to acquire enough insights and knowledge of its technical hermeneutics[8] and mechanism typical for the new technical being and the particular intelligence created by generated writing. Describing the concept of new writing on a technical level, this text has hopefully contributed to this by explaining some of the mechanisms such as measuring the distance between the many words of a text corpus in vectors and calculating the meaning from those distances and relations. In a human artifice that is shaped by technical realities such as these as much as it is by political realities, with both of them in fact often overlapping, 'it is necessary for every man employed with a technical task to surround the machine both from above and from below, to have an understanding of it in some way', as Simondon (1958: 80) points out. Writing, having always been a technique to operate and be in our world, has now, with the beginning of new writing, also become such a technical task. To return to the beginning of the new writing, is thereby helpful for understanding this task better. This beginning may also lead, one never knows, to the moment Derrida anticipated in 1967: the end of the book, and the opening of another chapter. At one point.

Works Cited

Amoore, Louise, Alexander Campolo, Benjamin Jacobsen, and Ludovico Rella. 2023. 'Machine Learning, Meaning Making: On Reading Computer Science Texts', *Big Data & Society,* 10: https://journals.sagepub.com/doi/full/10.1177/20539517231166887.

Arendt, Hannah. 1958 [1998]. *The Human Condition*. University of Chicago Press.

Arendt, Hannah. 1967. 'Truth and Politics', *The New Yorker*. https://www.newyorker.com/magazine/1967/02/25/truth-and-politics.

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2014. 'Neural Machine Translation by Jointly Learning to Align and Translate'. *ArXiv Preprint ArXiv:1409.0473*.

Baillehache, Jonathan. 2013. 'Chance Operations and Randomizers in Avant-garde and Electronic Poetry: Tying Media to Language', *Textual Cultures*, 8.1, 38-56. http://www.jstor.org/stable/10.2979/textcult.8.1.38

Bajohr, Hannes. 2022. 'The Paradox of Anthroponormative Restriction: Artistic Artificial Intelligence and Literary Writing', *CounterText*, 8.2, 262-282.

Bajohr, Hannes. 2023. 'Dumb Meaning: Machine Learning and Artificial Semantics', *IMAGE*, 37.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 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*, 610-23.

Bender, Emily M., and Alexander Koller. 2020. 'Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data', in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185-98.

Bratton, Benjamin, and Blaise Agüera y Arcas. 2022. 'The Model Is the Message', *Noema Magazine*. https://www.noemamag.com/the-model-is-the-message/.

Bunz, Mercedes. 2019. 'The calculation of meaning: On the misunderstanding of new artificial intelligence as culture', *Culture, Theory and Critique*, 60.3-4, 264-278.

Bunz, Mercedes. 2014. "When Algorithms Learned How to Write'. *The Silent Revolution: How Digitalization Transforms Knowledge, Work, Journalism and Politics Without Making Too Much Noise*, 1-24. London: Palgrave Macmillan.

Carr, Nicholas. 2008. 'Is Google Making Us Stupid?' *Atlantic Monthly,* 302.1. https://web.lib.unb.ca/instruction/bcull/ARTICLES/Reading/GoggleCBCA.pdf.

Carr, Nicholas. 2010. *The Shallows: How the Internet Is Changing the Way We Think, Read and Remember*. London: Atlantic Books.

ChatGPT. 2023. https://chat.openai.com/.

Chun, Wendy Hui Kyong. 2021. *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge, MA: MIT Press.

Derrida, Jacques. 1967a [1997]. *Of Grammatology*. Baltimore: Johns Hopkins University Press.

Derrida, Jacques. 1967b [1989]. *Writing and Difference*. Chicago: University of Chicago.

Derrida, Jacques. 1972a [1981]. *Dissemination*. Chicago: University of Chicago.

Derrida, Jacques. 1972b [1982] 'Signature Event Context', in *Margins of Philosophy*, 307-30. Chicago: University of Chicago.

Derrida, Jacques. 1981. *Positions*. Chicago: University of Chicago.

Facebook Inc. 2020. 'Third Quarter 2020 Results Conference Call'. https://s21.q4cdn.com/399680738/files/doc_financials/2020/q3/FB-Q3-2020-Earnings-Call-Transcript.pdf.

Firth, John R. 1957. 'A Synopsis of Linguistic Theory 1930-1955', in *Studies in linguistic analysis*, J.R. Firth, 1-32. London: Blackwell.

Foucault, Michael. 1969 [1997] *The Archeology of Knowledge*. London: Tavistock.

Future of Life Institute. 2023. 'Pause Giant AI Experiments: An Open Letter.' https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Equiano, Olaudah. 1789. *The Interesting Narrative of the Life of Olaudah Equiano, Or Gustavus Vassa, The African*. London.

Galloway, Alexander. 2021. 'Derrida's Macintosh', *Culture and Communication* [personal website], http://cultureandcommunication.org/galloway/derridas-macintosh

Gastaldi, Juan Luis. 2021. 'Why Can Computers Understand Natural Language? The Structuralist Image of Language Behind Word Embeddings', *Philosophy & Technology* 34.1, 149-214.

Grietzer, Peli. 2017. 'A Theory of Vibe', In Hannes Bajohr (ed.), *Thinking with AI: Machine Learning the Humanities*. London: Open Humanities Press, 20-32.

Harris, Zellig S. 1954. 'Distributional Structure', *WORD* 10.2-3, 146-62. https://doi.org/10.1080/00437956.1954.11659520.

Hartman, Saidiya. 2008. *Lose Your Mother: A Journey along the Atlantic Slave Route*. New York: Farrar, Straus and Giroux.

Hayles, N. Katherine. 2019. 'Can Computers Create Meanings? A cyber/bio/semiotic perspective', *Critical Inquiry* 46.1, 32-55.

Hayles, N. Katherine. 2022. 'Inside the Mind of an AI: Materiality and the Crisis of Representation', *New Literary History* 54.1, 635-666.

Heaven, Will Douglas. 2022, Nov 18. 'Why Meta's latest large language model survived only three days online', *MIT Technology Review*. https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/.

Hegel, Georg Wilhelm Friedrich. 1817 [1972]. *Philosophy of Mind*. Oxford University Press.

Hendrycks, Dan, Kevin Zhao, Steven Basart, Jacob Steinhardt, and Dawn Song. 2021. 'Natural Adversarial Examples'. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15262-71.

Hua, Minh, and Rita Raley. 2023. 'How to Do Things with Deep Learning Code', *ArXiv Preprint*. https://arxiv.org/abs/2304.09406.

Jacobs, Harriet. 1861. *Incidents in the Life of a Slave Girl, Written by Herself*. Boston, Mass.: Thayer & Eldridge.

Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. 'Survey of hallucination in natural language generation' *ACM Computing Surveys*, 55.12, 1–38.

Jurafsky, Daniel, and James H. Martin. 2023. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson. https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf.

Khan, Nora. 2019. 'Seeing Naming Knowing', *The Brooklyn Rail*. https://www.are.na/block/10908843.

Kirschenbaum, Matthew G. 2023. 'Prepare for the Textpocalypse', *The Atlantic*. https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318/

Kirschenbaum, Matthew G. 2021. *Bitstreams: The Future of Digital Literary Heritage*. Philadelphia: University of Pennsylvania Press.

Kirschenbaum, Matthew G. 2016. *Track Changes: A Literary History of Word Processing*. The Belknap Press of Harvard University Press.

Langer, Susanne K. 1948 [1954]. *Philosophy in a New Key: A Study in the Symbolism of Reason, Rite, and Art*. New York: Mentor.

Lemoine, Blake. 2022. 'What Is LaMDA and What Does It Want?'. Medium. https://cajundiscordian.medium.com/what-is-lamda-and-what-does-it-want-688632134489.

Leroi-Gourhan, André. 1964 [1993]. *Gesture and Speech*. Cambridge, MA: MIT Press.

Luitse, Dieuwertje, and Wiebke Denkena. 2021. 'The Great Transformer: Examining the Role of Large Language Models in the Political Economy of AI', *Big Data & Society*, 8.2. https://journals.sagepub.com/doi/pdf/10.1177/20539517211047734

Luong, Thang, Hieu Pham, and Christopher D. Manning. 2015. 'Effective Approaches to Attention-Based Neural Machine Translation', in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 1412-21. Association for Computational Linguistics. https://doi.org/10.18653/v1/D15-1166.

Meikle, Graham. 2023. *Deepfakes*. Cambridge: Polity.

Merriam Webster. 2024. 'To Write'. www.merriam-webster.com/dictionary/write.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. 'Efficient estimation of word representations in vector space', ArXiv Preprint. https://arxiv.org/abs/1301.3781

Moran, Chris. 2023, April 6. 'ChatGPT Is Making up Fake Guardian Articles. Here's How We're Responding', *The Guardian*. https://www.theguardian.com/commentisfree/2023/apr/06/ai-chatgpt-guardian-technology-risks-fake-article.

Moretti, Franco. 2007. *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso.

Nietzsche, Friedrich Wilhelm. 2002. *Schreibmaschinentexte: Vollständige Edition, Faksimiles und Kritischer Kommentar*. Weimar: Universitätsverlag Bauhaus Universität Weimar.

Peng, Baolin, Michel Galley, Pengcheng He, Hao Cheng, Yujia Xie, Yu Hu, Qiuyuan Huang, et al. 2023. 'Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback', ArXiv Preprint ArXiv:2302.12813.

Perrigo, Billy. 2023, Jan. 'OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic', *Time*, https://time.com/6247678/openai-chatgpt-kenya-workers/.

Plato. 2005. *Phaedrus,* edited by C. Rowe. London: Penguin.

Ribas, Jordi. 2023, Feb 21. 'Building the New Bing', *Microsoft Bing Blogs*. https://blogs.bing.com/search-quality-insights/february-2023/Building-the-New-Bing/.

Rieder, Bernhard. 2020. *Engines of Order: A Mechanology of Algorithmic Techniques*. Amsterdam University Press.

Riffaterre, Michael. 1993. *Fictional Truth*. Baltimore: John Hopkins University Press.

Roberge, Jonathan, and Michael Castelle. 2021. 'Toward an end-to-end sociology of 21st-century machine learning'. *The cultural life of machine learning: An incursion into critical AI studies*, edited by J. Roberge and M. Castelle, 1-29. London: Palgrave Macmillan.

Roberge, Jonathan and Tom Lebrun. 2023. 'Parrots All the Way Down: Controversies within AI's Conquest of Language', in: *KI-Realitäten,* edited by R. Groß, and R. Jordan, 39-66. Bielefeld: transcript Verlag. https://doi.org/10.14361/9783839466605-003

Roose, Kevin. 2023. 'A Conversation With Bing's Chatbot Left Me Deeply Unsettled', *New York Times*. https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html.

Rorty, Richard (ed.). 1967 [1992]. *The Linguistic Turn: Essays in Philosophical Method*. Illinois: University of Chicago Press.

Rumbelow, Jessica and mwatkins. 2023, February 5. 'SolidGoldMagikarp (plus, prompt generation)', LESSWRONG. https://www.lesswrong.com/posts/aPeJE8bSo6rAFoLqg/solidgoldmagikarp-plus-prompt-generation.

Sayyid, Omar Ibn. 1831. 'The Life of Omar Ben Saeed, Called Morro, a Fullah Slave in Fayetteville, N.C. Owned by Governor Owen', in *A Muslim American Slave: The Life of Omar Ibn Said,* edited and translated by Ala Alryyes, 47-80. Madison: University of Wisconsin Press.

Shoemaker, Tyler. 2023. 'Verkettete Textualität' [Concatenative Textuality], in *Quellcodekritik: Zur Philologie von Algorithmen*, edited by Hannes Bajohr and Markus Krajewski, 217-243. August Verlag, 2023. English permalink: https://escholarship.org/content/qt20k8q4xc

Simondon, Gilbert. 1958 [2017]. *On the Mode of Existence of Technical Objects*. Minneapolis: Univocal Publishing.

Statista. 2023a. 'Emails Sent per Day 2025'. Statista. https://www.statista.com/statistics/456500/daily-number-of-e-mails-worldwide/.

Statista. 2023b. 'WhatsApp: Daily Sent Message Volume 2020'. Statista. https://www.statista.com/statistics/258743/daily-mobile-message-volume-of-whatsapp-messenger/.

Surkis, Judith. 2012. 'When Was the Linguistic Turn? A Genealogy', *The American Historical Review,* 117.3, 700-722. https://doi.org/10.1086/ahr.117.3.700.

Virilio, Paul. 2007. *The Original Accident*. Cambridge: Polity.

von König, Dominik. 1977. 'Lesesucht und Lesewut', in *Buch und Leser. Vorträge des Ersten Jahrestreffens des Wolfenbütteler Arbeitskreises für Geschichte*

des Buchwesens 13. und. 14. Mai 1976, edited by H.G. Göpfert, 89-124. Hamburg: Hauswedell.

Underwood, Ted. 2023, March 19. 'Using GPT-4 to measure the passage of time in fiction: Large language models are valuable research assistants, especially when they refuse to follow instructions' [personal website]. https://tedunderwood.com/2023/03/19/using-gpt-4-to-measure-the-passage-of-time-in-fiction/

Weatherby, Leif, and Brian Justie. 2022. 'Indexical AI.' *Critical Inquiry*, 48.2, 381-415. https://www.journals.uchicago.edu/doi/abs/10.1086/717312.

Wikipedia. 2023, May 17. 'Writing'. https://en.wikipedia.org/wiki/Writing.

Notes

The text owes much to my sabbatical, my computer's search function to comb through too many PDFs, the patience of Hannes Bajohr, the technical expertise of Niklas Stoehr, and the attentive eye and language precision of Joely Day.

1. Before LLMs, most attempts in computer science to automatically generate writing was based on the usage of templates (Bunz 2014). This is a different approach from the automation of language in literature, where avant-garde and electronic poetry generated language with the aim to introduce moments of randomness (Baillehache 2013). Interestingly, recent experiments in contemporary literature using language models turned again in a new and different direction by exploring the intersection of language and computation (Bajohr 2022).
2. For an introduction to the concept of cultural techniques with particular attention to aspects of algorithmic information ordering, see chapter 3 in Bernard Rieder's (2020) excellent *Engines of Order: A Mechanology of Algorithmic Techniques*, 81-141.
3. Derrida (1967: 26), who calls Hegel 'the first thinker of writing', saw this as a clear sign of Hegel's 'Europeocentrism' (3) and criticised his belief that the alphabet enables one to represent an idea or name directly and in the simplest form in contrast to pictorial writing such as that of Egyptian hieroglyphs.
4. Nietzsche, who was one of the early users of a typewriter (he is a known user of the early model the writing ball), famously remarked in 1882 to his secretary 'Unser Schreibzeug arbeitet mit an unseren Gedanken', which literally means 'Our writing tools work along when we form our thoughts' (Nietzsche 2002: 18). See also Alexander Galloway's (2021) take on 'Derrida's Macintosh'.
5. Not very surprisingly for contemporary neoliberal capitalism as the ruling ideology, Wikipedia actually lists this point first.
6. Foucault (1969: 65) addressed writing among other techniques in the *Archaeology of Knowledge* as 'procedures of intervention' mentioning among others 'techniques of rewriting', 'methods of transcribing', 'modes of translating'.
7. Although it would introduce other problems – the devil is as always in the details. While attention models can put their focus on particular parts, they can only take in sequences of fixed length. Recurrent Neural Networks, on the other hand, can process sequences of unlimited length.
8. Using Ricoeur for the concept of hermeneutics with regards to LLMs could be highly interesting, as Roberge and LeBrun (2023) have demonstrated.