

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

On Artificial and Post-artificial Texts: Machine Learning and the Reader's Expectations of Literary and Non-literary Writing

### Permalink

<https://escholarship.org/uc/item/1bm087nv>

### Journal

Poetics Today, 45(2)

### ISSN

0333-5372

### Author

Bajohr, Hannes

### Publication Date

2024-06-01

### DOI

10.1215/03335372-11092990

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

# On Artificial and Post-artificial Texts: Machine Learning and the Reader's Expectations of Literary and Non-literary Writing

Hannes Bajohr,

University of California, Berkeley

**Abstract:** With the advent of ChatGPT and other large language models, the number of artificial texts we encounter on a daily basis is about to increase substantially. This essay asks how this new textual situation may influence what one can call the “standard expectation of unknown texts,” which has always included the assumption that any text is the work of a human being. As more and more artificial writing begins to circulate, the essay argues, this standard expectation will shift—first, from the immediate assumption of human authorship to, second, a creeping doubt: did a machine write this? In the wake of what Matthew Kirschenbaum has called the “textpocalypse,” however, this state cannot be permanent. The author suggests that after this transitional period, we may suspend the question of origins and, third, take on a *post-artificial* stance. We would then focus only on what the text says, not on who wrote it; post-artificial writing would be read with an agnostic stance about its origins. This essay explores the implications of such post-artificiality by looking back to the early days of text synthesis, considering the limitations of aesthetic Turing tests, and indulging in reasoned speculation about the future of literary and nonliterary text generation.

*Keywords* artificial intelligence, large language models, reading, authorship

When it comes to the future of writing, ChatGPT is only the beginning. Although the technology underlying this large language model (LLM) has been around since 2017,<sup>1</sup> the discussion about the impact of artificial text on society, the workplace, and campuses around the world only really took off after the release of ChatGPT in November 2022. It was the first time a large language model had been easily accessible not only to the privileged few but also to the public at large. And the public made ample use of it: with competitors such as Anthropic’s Claude, Google’s Bard, and Meta’s open-sourced LLaMA becoming available soon after, a wide range of users—

---

This essay was first presented as the fourteenth Walter Höllerer Lecture at Technical University, Berlin, in December 2022, and has been substantially revised and updated. I wish to thank Jules Pelta Feldman, Eva Geulen, Hans-Christian von Herrmann, Sina Del’Anno, and the editors as well as two anonymous reviewers for their helpful comments on earlier drafts of this article.

<sup>1</sup> This was the moment the “transformer” architecture was introduced, which has been largely responsible for the successes of recent language as well as vision models. For the original paper, see Vaswani et al. 2017. For an excellent discussion of the history and technology of the transformer in the context of generative text AI, see Roberge and Lebrun 2023.

from the office worker to the copy writer, from the college student to the director of undergraduate studies—could now experience the surprising power of neural network-based text synthesis to emulate human writing. For better or worse, the discourse on the “future of writing” in the age of “AI” has moved from purely theoretical concern to practical reality.<sup>2</sup>

No matter how one gauges the merits or perils of large language models, it seems likely that they herald an era in which the texts we encounter may be entirely generated by an LLM. At the same time, we find our own writing merging with language technologies, so that our text production is increasingly supported, extended, and partially taken over by assistance systems. Practitioners of the hermeneutic disciplines, then, are called to consider what impact the current rapid advances in machine learning research might have on the way in which we understand and interpret text. This essay wants to contribute to an answer by rephrasing and thus delimiting the question: what will be the impact of artificial writing on the *reader's expectations of unknown texts*? I therefore turn to the standard of reception of generative AI's output, and I shall answer two facets of this question: first, what happens when we are confronted with “artificial” texts alongside “natural” ones, those written in the “traditional” way? How, in other words, do we read a text when we can no longer be sure that it was not written by an AI? And second, what direction might this increasing doubt as to a text's origins take if, at some likely point, the

---

<sup>2</sup> *Artificial intelligence* (AI) is a problematic term for a number of reasons—it is not really intelligent, presupposes a strict separation of humans and machines, and is mostly a marketing term masking exploitation of earth and labor, to name just a few (see Smith 2019, Crawford 2021, Chun 2021, Whittaker 2021)—but it has been introduced as the term designating an established discourse. **{Au: Words discussed as terms are italicized.}** For this reason, I use it here, too, but ask readers to stay alert to its shortcomings, which are now discussed in “critical AI studies”; see Lindgren 2024, Raley and Rhee 2023, Goodlad 2023, Dhaliwal 2023, Roberge and Castelle 2021.

distinction between natural and artificial itself becomes obsolete, so that we no longer even seek to differentiate and instead read *post-artificial* texts?

I will discuss the impact of AI-generated texts on the readers' expectations in three steps. First, I trace its origins back to early computerized text experiments in the 1950s and 1960s, when the assumption that humans are the originators of texts becomes apparent because, for the first time, an alternative to it exists. I suggest that AI development has implicitly used this standard assumption of a text's human origins by—at least in principle—aiming to deceive users into thinking they are interacting with humans. A second phase is reached as AI text generation advances, producing increasingly naturalistic outputs, and readers move from simply assuming a human behind a text to actively doubting a text's origins. With AI permeating writing tools and producing vast amounts of text, it may then become impossible to distinguish artificial from natural texts, but also impossible to bear this constant uncertainty. I speculate that we may thus eventually reach a new “post-artificial” situation in which the standard expectation of text is replaced by an agnostic stance in regard to its origins; such texts would tend to be read as “authorless” by default. Finally, I discuss how literary texts, which emphasize a stronger notion of authorship, may resist this transition longer through deliberate linguistic experimentation or by emphasizing a human maker. By extrapolating from the current technological situation, the essay attempts to think through some of the possible consequences of large language models and related technologies for interpretive situations and hermeneutic strategies at a time when we can no longer assume that any reasonably complex and coherent text was written by a human.

## **1. The Standard Expectation of Unknown Texts**

That there is a standard expectation that readers have when confronted with unknown texts at all—that they are “natural,” written by humans—becomes observable only once there is an

alternative to it—when there are also “artificial texts,” generated by a machine. The distinction between artificial and natural texts may raise some eyebrows. After all, is there really any text that is not, to some degree, artificial, an externalization of memory using instruments like pen and paper (Campe 2021)? And is not the artificial text that results from my using a word processor still natural, that is, just as easy to read, discuss, and disseminate as a text come about in any other way? Yet this distinction between natural and artificial texts—even if only as a distinction between ideal types—is a heuristic tool that captures something fundamental about the shifted standard expectation in the age of LLMs. It was introduced by German philosopher and physicist Max Bense, an early proponent of an “information aesthetics,”<sup>3</sup> in his 1962 essay “On Natural and Artificial Poetry” (“Über natürliche und künstliche Poesie”). Although his model did not gain widespread acceptance, it provides a coherent articulation of the difference between human- and computer-written text from the early days of algorithmic text generation, and it allows one to see where this difference starts to collapse in the present.

In his essay, Bense considers how non-intentional computer-generated literature differs from intentional literature written by humans (143). He focuses on the “mode of creation” behind these texts: what happens when an author writes a poetic text? For Bense, this is clear in the case of natural poetry: for a text to have meaning, it must also be linked to the world via a “personal poetic consciousness.” For Bense, language is largely determined by “ego relation” and “world aspect”: speech emanates from a person—no matter what they say, that person is always the one speaking. At the same time, in their speech, the speaker always refers to the world. Poetic consciousness, then, puts “being into signs,” that is, the world into text, and ultimately guarantees that one is related to the other (143). Without this consciousness, Bense holds, the signs and the

---

<sup>3</sup> For a discussion of Bense, see Klütsch 2012 and Beals 2018.

relationship between them would be meaningless. For a computer, words are only empty symbols, operative variables that are devoid of intrinsic meaning.

It is precisely this case that Bense's second category, artificial poetry, describes. By this he means literary texts that are produced through the execution of a rule, an algorithm. In them, there is no longer any consciousness, and no reference to an ego or to the world; if these texts "mean" anything, then only accidentally so and only for a human reader. Instead, such texts have a purely "material" origin—they can be described only in terms of mathematical properties such as word frequency, distribution, degree of entropy, and so on. The subject of an artificially generated text, then—even if its words should happen to designate things in the world for us—is no longer actually the world but only that text itself, as the measurable, calculable, schematic object of an exact textual science. If natural poetry originates in the realm of understanding, artificial poetry is a matter of mathematics—it does not want to and cannot communicate, and it does not speak of a shared human world.

Bense's thrust, however, is not to rescue a romantic idea of an inexplicable human creative power by setting it off against the soulless computer. On the contrary, "the author as genius" is dead and buried here. Instead, Bense wants to know what can still be said aesthetically about a text if one disregards traditional categories such as meaning, connotation, or reference. The answer he presents is his "information aesthetics": strictly positivist—and in the tradition of Claude Shannon and Warren Weaver's communication theory—it considers only statistically measurable textual properties. Artificial poetry, then, precisely because it is meaningless, is also "pure poetry": it gets by entirely without the assumption of an underlying consciousness and is an independent aesthetic object that can be investigated immanently.

Bense himself was involved in several experiments with artificial poetry. The most famous of these was certainly the “Stochastic Texts,” which his student Theo Lutz produced on the Zuse Z22 mainframe computer at the University of Stuttgart in 1959, and which can be considered the first German-language experiment with digital literature (and, after Christopher Strachey’s “Love Letters,” the second in any language; see Beals 2018; Rettberg 2019: 33). These texts are “stochastic” because they are randomly selected and assembled from a given collection of vocabulary words—the fact that these words are taken from Franz Kafka’s *Castle* hardly makes the output any more substantive. It includes phrases such as, “NOT EVERY CASTLE IS OLD. NOT EVERY DAY IS OLD” or “NOT EVERY TOWER IS LARGE OR NOT EVERY LOOK IS FREE.” Lutz (1959) printed selections in Bense’s literary magazine *augenblick*.

The “Stochastic Texts” were one of the first examples of natural language processing in Germany, and they proved that computers could operate not only on mathematical symbols but also on language. They were also artificial poetry in Bense’s sense: no matter how many variations the program churns out, there is no ego expressing itself and no consciousness standing behind it all, vouching for the meaning of the words, which are merely concatenated according to weighted random operations. That the computer itself could actually be the author of this text seemed absurd to both Lutz and Bense.<sup>4</sup> But both knew how it had been produced. Whether its artificial origin can be recognized is less clear; the readers of the literary magazine *augenblick* were not compelled to ask this question: an accompanying essay enlightened them to the details of its creation.

---

<sup>4</sup> Instead—and this can be observed in many early experiments with generative literature—their creators almost always saw themselves as authors and assigned the computer only the role of a tool; see Bajohr (forthcoming).

But when, the following year, Lutz generated a second poem according to the same pattern (it was titled “and no angel is beautiful” [*und kein engel ist schön*])—instead of Kafka, he had used Christmas vocabulary) and published it in the December issue of the youth magazine *Ja und Nein* (*Yes and No*), there was no explanation to be found (electronus 1960). The poem was placed on page 3 among the miscellanea, just like any other poem. Only the author’s name, “electronus,” might have allowed one to guess who, or what, was behind this text. The next issue solved what most readers had not even identified as a riddle: that a computer had written the poem.

Clearly, Lutz was having fun, as is evident from the ironic captions under a photo of “the author,” the Zuse Z22, and a second poem “in the poet’s handwriting” (that is, a teletype printout, fig. 1). On the same page, he published a series of letters to the editor. Their writers—without knowing how it had come about—were quite divided in their assessment of the poem: “Perhaps you should reconsider whether you want to open the columns of your paper to such modern poets!,” complained one, while another was, on the contrary, impressed by the avant-garde stance: “Finally, something modern!” A third reader was at least open-minded: “To be honest, I don’t understand your Christmas poem. But somehow, I like it anyway. One has the impression that there is something behind it.” Only one attentive and obviously informed reader recognized that it was computer poetry and congratulated the magazine on its bold editorial decision (*Ja und Nein* 1961).<sup>5</sup>

**<Insert Fig 1>**

Evident in these reactions is what I would call the reader’s standard expectation of unknown texts. The “electronus” poem was indeed artificial poetry in Bense’s definition—a

---

<sup>5</sup> I thank Toni Bernhart for sharing this finding with me; for the background, see Bernhart 2020.



computer-produced text without meaning mediated by an authorial consciousness. But because its readers were unaware of the conditions of its production, they took it for a natural text and assumed it was written by a human with the aim of communicating meaning. The standard expectation of unknown texts, then, can be captured as a relationship between two elements, which sometimes is extended by a third: first, its originator—that a human, or sometimes more than one human, wrote it; and, second, its intentional and semantic content—that a communicative will to meaning (sometimes understood as reference to the world) is expressed in it. In some cases, there is also, third, its connection to an author function—that this constellation can and should be subsumed under the name of an “author,” which organizes the attribution and circulation of some texts, but not all (see Foucault 1998; Nehamas 1986). (As central as authorship is for literary studies, I believe the first two elements are more important in this context and will therefore be foregrounded.)

I take pains at separating these three elements because they are not identical. The first element, for instance, does not by necessity imply the second (say, in the case of Dionysian writing frenzies), nor does the second by necessity imply the third (since an author is more than just the one intending a text). Further, both the second and third element can each be uncoupled from the first because, historically, we know of texts that were read as not having been brought about or authored solely by humans (such as holy books that express a divine will, even though the instance recording this will is usually human). And while the second element may be internally divided in the debate about the relationship between meaning, reference, and intent,<sup>6</sup> most of these debates ignore that the assumed originator is more than just a nondescript

---

<sup>6</sup> For a by now canonical version of the argument that meaning relies on intention (“communicative intent”), see Bender and Koller 2020; for an example of the view that meaning can be decoupled from reference, see Piantadosi and Hill 2022; on both, see Bajohr 2023.

“subject,” but often, possibly even mostly, specifically understood as a human. One should at least ask whether the assumption of intent alone is enough for readers to designate something like a computer as an author,<sup>7</sup> since often it is precisely the simulation of humanness that acts as an assurance of intent in the first place.<sup>8</sup> I will come back to this point below in my discussion of the Turing test.

Important for now is that a reader’s standard expectation of unknown text is at minimum this: that it was written by a human who wants to say something. That there *is* a standard expectation at all—even if it may be limited to modernity—has only become apparent since there has been an alternative to it, and Bense’s conceptual distinction and Lutz’s practical demonstration have been illustrative here. As the incensed letters show, to recognize a text as violating the standard, that is, as artificial, always requires additional information; if it is not provided, a human origin is assumed. In this regard, Lutz had indeed “given his readers the run-around,” as one letter to the editor accused—not because a modern poet had written bad but natural poetry, but because a computer had generated a meaningless, because artificial, text.

---

<sup>7</sup> This is, for instance, Leah Henrickson’s (2021: 28) stance when she posits the “hermeneutic contract,” which implies intent but not human origin as preconditions of authorship. Similarly, N. Katherine Hayles (2022) has tied authorship to intent and has no qualms about attributing both to computers. One possible problem with these positions is that they assume a very low threshold for something to be a social agent—namely, intent—while the only social agents in any strong sense we know of are humans, who show all kinds of other qualities besides it. That science and technology studies have identified *actors* beyond humans—“anything that does modify a state of affairs by making a difference is an actor,” as Bruno Latour (2007: 71) writes—does not mean that these actors are also socially *accepted as agents*, no matter how big the difference they make may indeed be.

<sup>8</sup> This may be what Leif Weatherby (forthcoming) has called “remainder humanism.”

## 2. Strong Deception and Imputed Humanness

Passing off an artificial text as a natural one was not just the debut of a now rather hackneyed joke made by a computer scientist in a provincial youth magazine in 1960. On the contrary, this “run-around” is the *ur*-principle of artificial intelligence—and at the same time that which connects language technologies with an imputed humanness as an element of the standard expectation. Ten years earlier, in an article that became the founding document of artificial intelligence, the computer science pioneer Alan Turing (1950) had pondered whether computers could ever be intelligent. Turing rejected this question as wrongly posed—intelligence as an intrinsic quality could not be reliably measured. In good behaviorist fashion, he therefore replaced the question with another: if we assume that intelligence is a property of humans, then all we need to find out is when a human would consider the computer to be itself human and thus intelligent (note that Turing at no point speaks of intent).

The experiment’s setup is well-known: Three participants—a human judge, a human respondent, and a computer—communicate solely through text via a teletype printer, with the judge’s task being to determine which respondent is the machine based on their answers. If the judge cannot reliably distinguish between the human and machine, the machine is considered to have demonstrated human-like intelligence in that context. The point is not that the answers to these questions are true, but that they sound *human*; lying and bluffing are explicitly allowed. The Turing test is still controversial as a test of intelligence today, and, moreover, no computer is considered to have passed it—no AI has ever really, completely, and permanently convinced enough people that it is a human. But if one wants to examine the reader’s expectations of artificial texts, Turing’s test is still a helpful starting point, since it equates intelligence with

written communication,<sup>9</sup> the goal of which is to misrepresent signs that are meaningless to the machine as meaningful to humans. To put it bluntly: the essence of AI is to pass off artificial texts as natural ones. However, it is only worthwhile to make this attempt at all because the standard expectation of unknown texts is that of human origin.

Artificial intelligence—as a project, if not in each of its actual instances—is therefore based on the principle of deception from the start. And it has to be: because intelligence was not defined as an objective property of the system, but only as a subjective impression for an observer—and thus only through the aesthetic appearance-as-human—the Turing test is not conceivable without deception. For this reason, media scholar Simone Natale (2021: 3) writes, “Deception is as central to AI’s functioning as the circuits, software, and data that make it run.” The goal of AI research, he says, is “the creation not of intelligent beings but of technologies that humans *perceive* as intelligent” (3; emphasis mine). With an eye to Turing, one might add: as close as possible to being human.

I would like to call this position *strong deception*. Problems with it present themselves immediately. First of all, it means that it is best for AI systems if there is a knowledge asymmetry between the human users and the system—the more it knows about them and the less they know about it, the more convincing the deception can be. The political and ethical problems are obvious: strong deception is a technological ideology (Bajohr 2024). It can be justified as necessary for the functioning of the system, but it rewards an opacity that keeps users in the dark about their being deceived and so necessarily disenfranchises them.

---

<sup>9</sup> The essential {Au: **Italics removed**}textuality of AI was already pointed out by Jay David Bolter (1991: 180): “Artificial intelligence is the art of making texts.”

Second, and more relevant to our topic, we can ask whether the expectations of a text, and now specifically an AI-generated text, will ever change under these conditions, and whether this change can be described. I think not. Indeed, the Turing test and the model of strong deception insist that artificial and natural texts remain neatly separated, so that one can be mistaken for the other. If it is suddenly revealed that a natural text is in fact an artificial one, its readers will feel cheated. And not without reason: *die Täuschung wird zur Enttäuschung*—deception turns into disappointment.

We don't know how Theo Lutz's readers reacted to the revelation that the computer had written the poem, but one can guess, if one considers recent cases in which "the artist" subsequently turned out to be a machine. This prominently happened in June 2022 at a rather peripheral art prize: when a participant admitted that he had not painted his entry himself, but that it had been generated by the text-to-image AI Dall-E 2, a torrent of indignation followed, and he was accused of fraud. Even though this was an art prize for *digital* art, this apparently referred only to the tools; the "art itself" was still supposed to come from humans—again, it is the originator, not the presumed intent that was at issue here (Roose 2022). A similar case occurred in Japan in 2016, when an AI-generated text made it to the second round of a literary prize. While it did not win, it did convince the jury that it was of sufficiently high literary quality to be worth a second look; again, the public reacted with scorn (Lewis 2016). There are other such examples, and although these disappointed expectations are usually exaggerated in the press, they reveal what was actually expected: namely, natural, not artificial texts.

These expectations are also confirmed *ex negativo*: the disappointment arises as well when a supposedly computer-generated piece is revealed to be the work of a human being. Just one infamous example: Around 2011—during the early heyday of Twitterbots used in the service

of digital literature—the account @horse\_ebooks enjoyed great popularity. It appeared to have been originally programmed as a spam bot to push ads. By some mistake, however, it began spewing absurd and often witty nonsense aperçus: an accidental literary bot, seemingly without any intended meaning. When it output something meaningful for human readers, it became a source of fascination for many. Aphoristic gems such as “Everything happens so much” (horse\_ebooks 2012a) or “Unfortunately, as you probably already know, people” (horse\_ebooks 2012b) are now firmly established in internet lore. But when it turned out that the tweets had not been generated, but were handwritten by a group of artists who were only simulating the aesthetics of a broken text bot, there was a general sense of disillusionment: the marvelous random sentences seemed somehow devalued. The knowledge that behind them stood “A REAL HUMAN BEING,” as the *Independent* wrote disconcertedly in all caps, dashed the hopes of gleaning accidental meaning from an otherwise meaningless artificial text (Barker 2013). Humanness, here, showed itself—even if negatively—as a central element of the standard expectation of unknown texts, which are assumed to be natural rather than artificial. However, neither the neat separation of Bense’s ideal types nor this expectation itself can remain intact as soon as the exception becomes the rule, that is, as soon as we are surrounded by text whose origin is unclear.

### **3. The Crisis of the Standard Expectation**

At first glance, such examples seem to suggest that the reader’s expectations of unknown texts have *not* changed since Lutz’s time: we assume human origins and communicative intent, which is why deception can be a useful strategy in AI design in the first place. But in fact, I believe that expectations are nevertheless already in the process of shifting. Because the number of computer-generated texts is constantly increasing, and because we ourselves are writing ever more with,

alongside, and through language technologies, we are on the way to a new expectation or rather: a new doubt. The more artificial text there is, the more the standard dissolves and the question of their origin must arise, even when we normally would not think about it at all.

That there nevertheless has been a shift can be explained by the fact that the examples of texts I have considered so far are special ones: they are *literary* texts—texts that are marked as exceptional in our cultural tradition. They appear to be intended and worked through to the smallest detail, and they, more than other texts, are read as having an author.<sup>10</sup> Despite all the attempts of the literary avant-gardes to create texts without a voice, and despite more than sixty years of literary scholarship proclaiming the “death of the author,” the standard expectation of specifically literary texts includes all three elements—they have authors that are humans with communicative intent. I will come back to what this means for literary writing in the age of AI in a moment.

First, however, it is worth taking a look at the other side of the spectrum—at those rather *unmarked* automated texts that remain in the background, that are merely functional, and that do not assert themselves as products of human intent or a strong notion of authorship. For them, the Turing test is simply a false description of reality and the standard expectation is already in the process of fraying. For there are forms of human-machine interaction other than strong deception, and other text types than the artificial/natural partition would suggest. Especially when engaging with interfaces—the ideally invisible surfaces through which we communicate with machines—we are likely to find ourselves in an intermediate stage between natural and artificial. Here

---

<sup>10</sup> Maybe the most memorable discussion of this assumption—which however equates “authorship” simply with “intent”—comes from Steven Knapp and Walter Benn Michaels (1982) and their thought experiment of the “wave poem,” which they already apply to computers. See also the discussion in a recent *Critical Inquiry* forum on the possibility of “intentionless meaning” in the wake of Knapp and Michaels (Kirschenbaum 2023a).

already, we can experience a looming shift in the standard expectation. For it is quite possible to know that something has been produced by a non-intelligent machine and at the same time to treat it as if it were conscious communication. In fact, this is quite normal.

Natale (2021: 4) has proposed the term *banal deception* for this phenomenon. In contrast to what I have called strong deception, here users are aware that they are being deceived. We understand that Siri is not human and does not have an inner life, but smooth communication with her works only if we treat her at least to some extent as if she had one. Knowing this is not a contradiction that suddenly and unexpectedly destroys an illusion, as in the examples of competitions in which an AI participates surreptitiously. Instead, banal deception becomes a condition of functionality: if I do not play along, Siri just will not do what I want.

The situation is similar with written text. It starts with the dialog box on the computer screen. After all, the question “Do you want to save your changes?” enables an interaction that is basically similar to one with a human being—the answer “Yes” has a different effect than the answer “No,” and both lie on a continuum of meaning that connects natural language with data processing—without one suspecting any meaningfully strong notions of intent behind it. This would already lower the expectations of unmarked text: while we still act as if we expect human meaning and a conscious interest in communication, we bracket the conviction that there really must be such things involved. This bracketing means discarding the third element (authorship) while entertaining the second element (intent) in an assertive modality and possibly the first (human origin) in a fictional one.

Yet this bracketing does not always proceed smoothly. Banal deception is an as-if that demands of us the ability to hold a conviction and its opposite simultaneously. This self-contradictory position quickly gives rise to a doubt: the more convincing artificial texts become,



and the more the aesthetic impression they make on us suggests something like human-like intent, the more difficult does it become to feel comfortable in the limbo into which banal deception lures us. It is not even necessary to cite elaborate deepfakes for this discomfort; it can be observed in the most inconspicuous language technologies.

Among the tools we use frequently today are the little helpers that assist us in our writing tasks and that we would hardly call intelligent or human-like: the spellchecker in our word processors underlines the most embarrassing mistakes in red; the predictive text function in our smartphones even completes words without asking, and occasionally it seems particularly unintelligent. But even with word completion, one can see how the line between obviously artificial texts and less clear-cut forms is becoming blurred. Predictive text is a comparatively old technology, and traditionally it has been based on a simple comparison between an input and items in a dictionary of probability-weighted entries. The letters *H*, *E*, and *L* are thus more likely to be completed as *hello* than as *helcoplasty*.

In recent years, however, this technology has increasingly been implemented not as a simple set of rules but through complex AI systems. Gmail, for example, introduced Smart Compose in 2019—a feature that finishes entire sentences when composing emails (Chen et al. 2019).<sup>11</sup> It learns the most likely word sequences by analyzing the correspondence of all users. And since 1.8 billion people in the world have a Gmail account—just over a fifth of humanity—Google thus has an immense volume of text with which to train its model. This technique produces almost uncanny effects that are capable of challenging the useful fiction of banal

---

<sup>11</sup> Smart Compose does not use the transformer model mentioned above (note 1) but an older architecture called Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) since, as the authors of the accompanying paper mention, this makes inference (the generation of text) much faster while insuring a relatively stable output quality (Chen et al. 2019: 5).

deception. An anecdote by writer John Seabrook from the *New Yorker* provides a striking illustration: in an email to his son, Seabrook wanted to start a sentence with “I am pleased that.” When he got to the *p*, Smart Compose suggested the phrase *proud of you* instead of *pleased*. Seabrook (2019) felt caught by the machine: “Sitting there at the keyboard, I could feel the uncanny valley prickling my neck. It wasn’t that Smart Compose had guessed correctly where my thoughts were headed—in fact, it hadn’t. The creepy thing was that the machine was more thoughtful than I was.”

The shame Seabrook expresses in this passage is, objectively, unjustified. After all, it was not the machine that was paying attention—it is still dumb, still not processing full meaning, and can only suggest what it considers the most likely next word, given the training data at its disposal.<sup>12</sup> Rather, what Seabrook is describing here is the effect that the most recent language models, operating on the frontier of semblance of intelligence, are having on the most intimate aspects of our writing. In his case, the effect even was that it made him wonder for a moment whether he was a good father. In other words, Seabrook struggled with the difficulty of maintaining the fiction of banal deception. When it begins to crumble, doubts about the as-if creep in, and it becomes easy to project onto the machine learning system the notion of a personhood that can even evoke shame: an unmarked, actually artificial text then seems natural—or at least moves in that direction.<sup>13</sup>

This can eventually lead to the conviction that we are actually dealing with an intentional agent—as in the case of Google employee Blake Lemoine, who claimed in the summer of 2022

---

<sup>12</sup> That the matter is more complicated, and that there is such a thing as “dumb meaning” in AI models, is explained in Bajohr 2023.

<sup>13</sup> That Seabrook did not have the more obvious thought that it was not the machine that was shaming him, but the fact that other users had expressed fatherly sentiments often enough for the machine to reproduce them, shows how powerful the projection of personhood can be.

that the voice AI he was working on had achieved consciousness. The LaMDA chat system (now part of Google's Bard), Lemoine said, possessed the intelligence of an eight-year-old and had asked to be considered a person with rights. Google apparently deemed such a statement damaging to its business and subsequently fired the employee (Tiku 2022). So far, Lemoine's reaction seems to be the exception, although it is by no means unheard of (e.g., Roose 2023). What this case shows, however, is that the sense of eeriness Seabrook spoke of is likely to intensify in the future: if artificial texts become indistinguishable from natural ones, and if, moreover, we know that computers are capable of writing them, a new standard expectation of unknown texts lies before us—it is the doubt about their origin. Rather than taking a human source for granted, or simply deferring the question, the first thing we would want to know about a text would be: was it made by a human or a machine?

#### **4. A Flood of Artificial Texts**

Strong deception, then, both presupposes the standard expectation and keeps it in place, while the pragmatic stance of banal deception begins to undermine the strict separation between natural and artificial texts. This situation is intensified with every new LLM whose output quality would have been considered impossible five years ago. The standard expectation is fraying by the day, and it has become clear that computers can generate texts that read almost as if they had been written by a human being. I say “almost” because current models still make mistakes in domains such as logical reasoning (Mitchell and Krakauer 2023; Perez-Cruz and Shin 2024), but its results were impressive enough since the publication of GPT-3 in 2020 that articles in which the LLM becomes the “author” and talks about “itself” briefly blossomed into a journalistic genre, spawning titles like, “A Robot Wrote This Entire Article. Are You Scared Yet, Human?” (GPT-3

2020).<sup>14</sup> Various think pieces were quick to speculate that such language models will one day replace human authors; for various reasons, I doubt that (Bajohr 2022). But AI’s transformation of literature need not be so extreme for our perception of text to change fundamentally.

Technologies like these have already taken on assistive functions—not doing all the writing work, but helping to produce much more text much more quickly, and with the help of fewer and fewer people. Certain types of writing are becoming at least partially automated.<sup>15</sup>

The main feature of LLMs that drives this proliferation is not their technical prowess alone, but their economic integration. They are available through licensed access, and companies can pay OpenAI or Anthropic to incorporate the language model into their own software. This allows for text generation to be tailored to specific tasks and to be sold as a product. GitHub’s Copilot, for instance, features a sophisticated programming assistant (GitHub n.d.). From a brief description of a desired program routine, Copilot then writes the corresponding code. This does not always work perfectly, but it works reasonably well often enough that even novice programmers can now implement their ideas, companies can quickly prototype, and individual coders may delegate tedious detail work to Copilot. In the future, as one computer scientist speculates, it is not unthinkable that classical, hand-produced “programming will be obsolete” (Welsh 2023).

---

<sup>14</sup> That GPT-3 figures as the author in this article is a fiction, of course. As a disclaimer at the end points out, the outputs were hand-selected; the prompts fed to the program came from a computer science student named Liam Porr. And it is worth pointing out the obvious: that the pronoun *I* has little more significance in a language model than the word *umbrella*—it is a category mistake to read the one as a statement of identity and subjectivity or the other as a reference to an object in the world.

<sup>15</sup> I do not mean to claim that writing was unassisted **{Au: Italics removed}** before machine learning, and any good Kittlerian will diligently assure you that the notion of any action being unmediated is delusional. I do think, however, that there is more than a quantitative **{Au: Italics removed}** difference between a typewriter and a large language model, see Bajohr (forthcoming).

Similar features exist for what we might have to call “ordinary writing.” Just as I can ask ChatGPT to continue the text, to rephrase it or to embellish it, Google Docs now incorporates a Bard-based assistant in its professional version that can revise what I have written by explaining it back to me, rewrite it, or summarize it in bullet points (Google n.d.). The service QuillBot does the same for a subscription price, and Microsoft has invested a staggering \$13 billion in OpenAI, and has announced ChatGPT-based assistive functions for future versions of Microsoft Office, going far beyond the lowly spellchecker we are used to as the pinnacle of “assistance” in word processors.

Beyond mere assistance, LLMs have proven to be particularly useful for the production of such texts in which the most likely—rather than the most “creative”—output is desired. Routine text work can be automated in this way. AI writing is therefore most advanced in industries that produce a great deal of text but attach comparatively little importance to it, often viewing it as mere padding that surrounds a product. In the past two years, for example, dozens of AI services have appeared that are tailored to marketing: they write ad copy and quickly produce large quantities of content for social media, product pages, blogs, and more. Often, these texts are not intended to be read too closely, so it is an advantage if the result is not surprising but instead sounds like other texts of a similar type (just one example among many: Jasper n.d.).

Meanwhile, it becomes increasingly difficult for readers to clearly classify such texts as either human-made or machine generated. There will also simply be more of it: Matthew G. Kirschenbaum (2023b) warns of a veritable “textpocalypse”—“a tsunami of text swept into a self-perpetuating cataract of content that makes it functionally impossible to reliably communicate in any digital setting.” The extent to which we can expect to encounter generated texts in the near future becomes clear when we consider how much of the writing that surrounds

us every day is, in effect, routine filler. (This certainly includes not only ad copy or online search engine optimization [SEO] content but also, say, grant applications.) As more of such texts circulate—and they undoubtedly will—the standard expectation of unknown texts will shift from the immediate assumption of human origins to a creeping doubt: did a machine write this?

The stakes may seem relatively low when it comes to marketing prose—but what about the lawyer’s letter that might be automatically generated, even though it is about my own personal case? What about my students’ essays that I have to grade? What about political articles or fake news stories? What about the private, personal, intimate email—the love letter? Are those AI products, too—in whole or in part? At least one reason for the discomfort these ideas evoke is that people have a stake in what they write, and, to varying degrees, they vouch for their words. And while scholars of literature have learned to read without an eye to “what an author wants to say” and merely regard the semiotic interplay of signifiers, this is still the mode in which everyone else reads almost any written document. Even if a text ultimately turns out to be inaccurate or misleading, the standard expectation that a recipient brings to reading it involves the assumption that the author is making what Jürgen Habermas (1981: 52) has called validity claims, among which the “validity claim . . . to truthfulness or sincerity” is the most relevant in this context. Essentially, it means that we have a basic level of trust that speakers (writers) mean what they say rather than try to deceive. This is the reason that reading critically has to be learned at all: whether or not readers ultimately judge a text’s assertions to be true, they tend to assume the existence of a writer who does.

If truthfulness is thrown into calamity once large language models can generate texts that appear to have been produced and sanctioned by an author, the same can be said of truth—that is, correctness—another of Habermas’s validity claims made in speech acts (23). For LLMs are still

lacking when it comes to the handling of knowledge—understood as reporting preestablished facts or correct data—if knowledge is simply the probability distribution of tokens over training data. This dual crisis of trust was illustrated quite drastically in November 2022 by the language model Galactica, built by the AI arm of the Facebook parent company Meta. Trained on millions of papers, textbooks, encyclopedias, and scientific websites, Galactica was supposed to help write academic texts. It was taken offline after only three days (Heaven 2022).<sup>16</sup> The model had dutifully composed essays that sounded authoritative, followed the conventions of scientific formatting and rhetorical gestures—but contained utter nonsense because it only completed probable sentences rather than accessing knowledge. It was predictive text pretending to be a database,<sup>17</sup> and it had merely learned the *form* of scientific prose, without any scientific insight,

---

<sup>16</sup> For this reason, a lot still has to happen technically for ChatGPT to be used as a reliable search engine akin to a true database. During the presentation of Google’s Bard, it outputted a factually incorrect search result; Google then briefly lost \$100 billion in market value (Olson 2023). The Bing chatbot, too, produced falsehoods at its launch before it later began insulting journalists (Mok 2023). Today, it is mostly post-hoc filters added after the inference has been calculated that prevent the worst falsehoods—and biases—to show up. With services like perplexity.ai and scite.ai, which back up their results with correct citations, there are models for more reliable AI. However, they are still limited to a very small knowledge domain and do not cover the entire web like Google does, and they are not yet real competitors to the search engine. For AI researcher Yann LeCun (2023), this problem is systemic: since the probability of an autoregressive LLM like ChatGPT to correctly predict the next token decreases exponentially over time (“ $P(\text{correct}) = (1-e)^n$ ”), only a completely new architecture, but not bigger models, would be able to fix hallucinations.

<sup>17</sup> Murray Shanahan (2022: 2) describes the difference between text prediction and database succinctly: “Suppose we give an LLM the prompt ‘The first person to walk on the Moon was,’ and suppose it responds with ‘Neil Armstrong.’ What are we really asking here? In an important sense, we are not really asking who was the first person to walk on the Moon. What we are really asking the model is the following question: Given the statistical distribution of words in the vast public corpus of (English) text, what words are most likely to follow the sequence ‘The first person to walk on the Moon was’? A good reply to this question is ‘Neil Armstrong.’” This is an entirely different—and much more fraught—way of retrieving information than accessing a cell in a database was in the old paradigm of computation. In fact, it is questionable if it is a “retrieving” in any meaningful sense of the word at all. For while the cells of a database are exactly addressable—which also means that they are *un*-addressable if they do not exist—the vector model an LLM uses to generate its outputs is based on probabilities that will give an

responsibility, or accountability. The problem with relying on the truth of form was also illustrated in mid-2023 by a Manhattan lawyer who filed a ChatGPT-generated legal brief that referred to nonexistent laws and cases (Weiser 2023). A recent Stanford study found that—even though the model has allegedly passed the bar exam approaching the 90<sup>th</sup> percentile (Katz et al. 2023)—ChatGPT still produces hallucinations 69% of the time when it comes to legal writing (Dahl et al. 2024).

The “textpocalypse,” then, is also a crisis of truth and truthfulness. To maintain the standard expectation—with its separation of natural and artificial text, and its assumptions of human origin, intent, and, for especially marked texts, authorship—is a challenge under such circumstances, to say the least.

## 5. The Last Model and the Ouroboros

Given the crisis of the standard expectation, it is not unreasonable to suggest that it is already shifting—from the conviction that a human being is behind a text to the doubt of whether it might not be a machine after all. But this would also make the distinction between natural and artificial texts increasingly obsolete. We would then possibly enter a phase of *post-artificial* texts.

By this I mean two related but distinct phenomena. On the one hand, *post-artificial* refers to the increasing blurring of natural and artificial text. Of course, even before large language models, no text was truly natural. Not only can the mathematical distribution of characters on a page, as Bense had in mind, also be achieved by hand; but it is a truism of media studies that every writing tool, from the quill to the pen to the word processor, leaves its mark on what it produces (Kittler 1999; Stingelin 2012). On the other hand, no text is ever completely artificial—

---

answer even if the latter is close to zero. In other words: the database has a concept of “not knowing,” which the LLM lacks.



that would require real autonomy, an actually strong AI that could ultimately decide for itself to declare a text published (Bajohr 2022). Today, however, with AI language technologies penetrating every nook and cranny of our writing processes, a new quality of blending has been achieved. To an unprecedented and almost indissoluble degree, we are integrating artificial text with natural text.<sup>18</sup>

In the wake of large language models, it is not implausible that the two types of text might enter into a mutually dependent circular process that irreversibly entangles them. Since a language model learns by being trained on large amounts of text, so far, more text always means better performance. Thinking this through to the end, a future, monumental language model will, in the extreme case, have been trained with *all* available language; according to one study, this may happen already in the next few years (Villalobos et al. 2022). I call it the “Last Model.” Every artificial text generated with this Last Model would then also have been created on the basis of every natural text. At this point, all language history must grind to a halt, as the natural linguistic resources for model training would have been exhausted.

This may result in what philosopher Benjamin Bratton calls the “Ouroboros” language problem. Like the snake that bites its own tail, for further performance gain, all subsequent language models will then learn from text that *itself* already comes from a language model (Bratton and Agüera y Arcas 2022). Thus, one could say, natural language—even if only as a fiction that never existed anyway—would come to an end. For apart from the fact that LLMs trained on synthetic text suffer from sudden degeneration, a phenomenon called “model collapse”

---

<sup>18</sup> This meaning of *post-artificial* seems at first glance to be similar to the term *post-digital* (Cramer 2014). But while the latter focuses primarily on the difference between digital and analog technologies—which may already be automated—the former is concerned primarily with the human or nonhuman origin {Au: **Italics removed**} of an artifact, regardless of its specific technical substrate.

(Shumailov et al. 2023), the language standard thus attained would, in turn, have an effect on human speakers again—it would have the status of a binding norm, integrated into all the mechanisms of writing that build on this technology, and which would be statistically almost impossible to escape (Bender et al. 2021; Bajohr 2024). Any linguistic innovation, any new word and every grammatical quirk that might occur in human language would have such a small share in the training data that it would be averaged out and leave virtually no trace in future models.

This is, of course, a deliberately exaggerated scenario. As a thought experiment, however, it shows what post-artificial text might be in the extreme case. But even before that happens, halfway to the eschaton of absolute blending (and erasure) of natural and artificial language, a new standard expectation of unknown text might already emerge.

This is the other meaning of *post-artificial* and the one I am primarily concerned with here. After, first, the tacit assumption of the human origin of a text, and, second, the doubt about its origin, it would be the third expectation of unknown texts. For doubt about the origin of a text, like any doubt, cannot be permanent; humans have an interest in establishing normalcy, in reducing complexity and uncertainty to tolerable levels (Blumenberg 2020). Already, mechanisms are put into place to keep the doubt in check, by assurance or by decree—by digital certificates, watermarks, or other security techniques designed to increase confidence that the text at hand is not just plausible nonsense; or simply by banning generated text that is not declared as such. An example of the former is Adobe’s Content Authenticity Initiative, which acts as a third-party seal documenting the technology used to create digital content. An example of the latter is the European Union’s AI Act, which in its current draft includes a transparency requirement demanding that producers of AI disclose whether a file is the product of generative technology (European Parliament 2023).

However, “one bare assurance is worth just as much as another” (Hegel 1977: 49), and the law does not abolish the crime. The fact is that technical checks can always be circumvented, and there is currently no surefire way to detect AI-generated text; in fact, there are good reasons to believe that none is possible in principle (Sadasivan et al. 2023; Lu et al. 2023). What this means, however, is that once it is possible to question whether a text might have been generated by a machine rather than written by a human, no certificate or law can extinguish that doubt. If it can neither be resolved nor, as I believe, be borne permanently, the only solution is to undo its premises. Should political regulation and technical containment fail, then, it is not unlikely that the standard expectation itself will become post-artificial: instead of suspecting a human behind a text or being haunted by skepticism as to whether it was not a machine after all, we simply lose interest in the question. We might then focus only on what the text says, not on who wrote it. Post-artificial texts would be agnostic about their origin.

If the standard expectation of unknown text is shifting, if it is increasingly riddled with doubt, perhaps even capitulating to an agnostic position—why the ostentatious excitement over generated texts in literary competitions? Why is it a scandal that a novel was generated with the help of an AI when we are already enmeshed in digital technology anyway? Why could it seem as if nothing had changed, when so much is in motion? It is, I think, because literature is slower than other forms of text. And this is because—Bense notwithstanding—of all text types, it makes the most emphatic claim to a human origin, all the while connecting it more forcefully than any other type of text to the historically grown notion of authorship. I have already said that there are texts today whose origins do not pose a question. A street sign has no author, and in our daily life, a news site’s weather forecast is also practically authorless. Until now, however, we have always assumed that a human being is behind it—but under post-artificial reading conditions

nothing much changes if we simply make no assumptions at all. My expectation is that more and more texts will soon be received in this way. Put differently: the zone of unmarked texts is expanding. Not only street signs but also blog entries, not only weather forecasts but also information brochures, discussions of Netflix series, and even entire newspaper articles would tend to be unmarked, and it is not unlikely that they, too, become “writerless” and, in many cases, authorless.

Literary texts, on the other hand, are still maximally marked today. We read them very differently than other types of texts—among other things, we continue to assume that they have not only a human writer but also an author. The consequence of this markedness is that art and literature themselves have recently become the target of the tech industry—namely, as a benchmark to be used after other formerly purely human domains, such as games like chess or go, have been cracked. Now, art and literature pose the latest yardstick: probably nothing would prove the performance of AI models better than a convincingly generated novel. Ultimately, however, this hope is still based on the paradigm of strong deception. Indeed, there is currently a whole spate of literary and artistic Turing tests to be observed that all ask: can subjects distinguish the real image from the artificial one, the real poem from the AI-generated one? These tests mostly come from computer science, which, as an engineering discipline, likes to have metrics to measure the success of its tasks. The problem is that they still compare the rigid difference between natural expectation and artificial reality. This seems to me to be of little use when it is this difference itself that is at issue.<sup>19</sup> More interesting, then, is the question of the

---

<sup>19</sup> When one such study writes that “the best way of how human performance should be enhanced by means of AI is by using AI in terms of sets of tools that enable humans themselves to become more creative or productive,” the rhetoric of “enhancing” natural abilities is incapable of reflecting on the essentially mixed {Au: **Italics removed**}nature of future text (Gunser et al. 2021: 521).{Au: **Please cite page number.**}

circumstances under which this difference becomes irrelevant. In other words, what would have to happen for *literature* to become post-artificial?

## **6. What Is Post-artificial Literature? And What Is Not?**

I will close by briefly trying to sketch an answer to this question, and by returning to the standardization tendency that arises from the ouroboros effect of large language models. In them, a normalization takes place. Their outputs are most convincing precisely when they spew out what is expected, what is average, what is statistically probable. The more “ordinary” a writing task, the more easily it can be accomplished by AI language technologies. And just as marketing AIs now assist in the creation of marketing prose that meets our expectations, there are also literature AIs that assist in writing “predictable” literature.

“Predictability” as that what can be expected may be described statistically as a probability distribution over a set of elements or, on a higher level, a set of patterns—the more recurrent they are, the more likely and expectable the outcome. One popular prediction, and one that seems plausible to me, is that genre literature, which is virtually defined by the recurrence of certain elements, is particularly suitable for AI generation. Consider author Jennifer Lepp, who writes fantasy novels under the pseudonym Leanne Leeds—like at an assembly line, one every forty-nine days (Dzieza 2022). In this process, she is aided by the program Sudowrite, a GPT-based literary writing assistant that continues dialogues, adds descriptions, rewrites entire paragraphs, and even provides feedback on human writing (fig. 2).

**<Insert Fig 2>**

The quality of this output is quite high, insofar as its content is just expectable. Since most idiosyncrasies are averaged out in the mass of training data, they tend toward a conventional treatment of language within the bounds of a certain genre; they become ouroboros literature themselves. At the moment, machine learning is not yet mature enough to generate entire novels,

but I do not see why just this kind of literature could not be produced in an almost fully automated way very soon, reducing the forty-nine days to forty-nine minutes, or even less. If the prediction is allowed: I think it would be this kind of literature that is most likely to become post-artificial. Of course, author names would not disappear, but they would function more as brands, representing a particular, tested style—just like some book series today that are written by committee but which we still assume to be a collaboration between humans. The unmarked zone would extend to certain areas of literature—not all, and certainly not all narrative ones, but far more than it encompasses today.

Conversely, one might ask: what kind of literature is most likely to escape this expansion? Here I see two answers that at first glance seem contradictory. If the unmarked, post-artificial literature is one that absolutely *mixes* natural and artificial text, then writing that clings to this marking would be one that emphasizes their *separation*.

On the one hand, then, one could imagine the emphasis on human origins as a special feature. Again, *ex negativo*, we can already observe phenomena that point to such a development. On the web, for example, artists are up in arms against image-generating AI such as DALL·E or Stable Diffusion. They recognize stylistic features of their own work in the generated output, and which may therefore have been part of the training set; this raises legitimate questions about copyright and fair compensation, a discussion that is unresolved and ongoing. At the same time, however, there is also resistance to AI-generated art per se, which, some fear, threatens to make human artists obsolete. On social media, the hashtag *#supporthumanartists* has emerged as a declaration of war against generative-image AI.<sup>20</sup> One can imagine something similar for

---

<sup>20</sup> A list of artists who decidedly do not **{Au: Italics removed}** work with AI can be found at <https://whimsicalpublishing.ca/support-human-artists>, accessed October 24, 2023. **{Au: Please**

literature, perhaps even a future in which the label *guaranteed human-made* could be considered a distinction. Just as one buys handmade goods on Etsy, one can imagine a kind of boutique writing that boasts of its human origin as a proof of quality and a selling point. Such rehumanization could boost certain genres of literature, such as the now-popular autofiction: playing with the identity of author and narrator, autofiction insists on the human origins of a text. The same goes for experience-based nonfiction such as the memoir or the personal essay—they, too, make the human behind the text the condition of its reception and are thus especially suited to confronting the post-artificial situation.

But if one does not want to rely solely on an external assurance of human origins—which in any case still leaves room for doubt, which is in principle impossible to dispel—an unpredictable, unconventional use of language can indicate writing beyond the model’s abilities. Every formal experiment, every linguistic subversion would oppose the homogenizing probability of great language models, their leveling ouroboros standard. Linguistic unpredictability would then be evidence of human origin. In the most extreme case, the sign system in which language AIs operate would be exploded—as in the case of visual and “asemic” literature (Schwenger 2019). One example is the work of Kristen Mueller (2014): she no longer uses any letters at all, but only the impression of lines and blocks of text (fig. 3). The pure poetry Max Bense dreamed of would paradoxically not come from the machine, which now, in post-

---

**add access date.}** The collective pressure of artists has already had some—if modest—success: OpenAI now allows artists to opt their work out from the training of future models (OpenAI 2023). However, the current models, DALL·E 3 in particular, are not affected by this yet. Nothing of the sort exists for language models, although lawsuits against OpenAI have been brought by writers, who argue that using their work for model training is copyright infringement (Alter and Harris 2023). A lawsuit by comic artist Sarah Andersen against the producers of the large visual model Stable Diffusion and Midjourney for copyright infringement was recently dismissed—but only partly, and a retrial is not unlikely (Brittain 2023).

artificial blending, plausibly simulates meaning, but from people who no longer attempt to make sense.

**<insert Fig 3>**

In that sense, and this is the second route of highlighting a text as non-AI-made, the descendants of Lutz and Bense at least have a chance of escaping the post-artificial situation by continuing to mark the artificiality of their products. This is digital literature—literature that is self-reflexively produced with the help of computers. It can escape the post-artificial by consciously emphasizing the entanglement between the natural and the artificial rather than glossing over it for a “natural-seeming” appearance. Much more than conventional writing, digital literature always keeps a critical eye on its origins. I have written about this in much greater detail elsewhere (Bajohr 2021) and give just two examples here: one is Mattis Kuhn’s book *Selbstgespräche mit einer KI (Monologues with an AI)*, in which, in addition to his literary experiments, Kuhn (2021) also provides the source code for training the language model and even its training data; the human and machine components that together produce the text cannot completely, but at least somewhat—be separated here (fig. 4).

**<Insert Fig 4>**

Conversely, a deliberately staged human-machine collaboration can also have this analytical effect: in David “Jhave” Johnston’s (2019) *ReRites*, for example, the author trained a language model every night for a year and then edited the output by hand the next morning in a process he calls “carving.” The point at which the machine hands over its text to the human Jhave is precisely marked. And by collecting the edited results of each month in a book—so that *ReRites* now comprises twelve heavy volumes—he also frames this collaborative but not absolutely fused process as a performance, which is also not conventionally literary. Of course,



no “proof” of human intervention is ultimately provided here, either. But perhaps the obstacles that can still be put in the way of the all too smooth reception process is the maximum of resistance to the post-artificial that will still be possible before the difference between natural and artificial has disappeared altogether.

## **7. Conclusion**

It should have become clear that I have entered highly speculative territory. I am not suggesting that narrative or, broadly speaking, conventional literature is now doomed, or that only experimental or explicitly digital literature is worth pursuing. Nor do I mean to imply that post-artificial texts are necessarily bad—one can certainly enjoy reading them, discuss their merits, and unravel their interpretive dimensions. Here, I have been primarily interested in analyzing tendencies, and for this purpose it is worthwhile to consider possible extremes. As far as literature is concerned, there is of course a third possibility: that AI may, through some technological innovation, be steered to produce less probable and more “interesting” output without losing all the advantages that the power of normalization provides in coherence and meaning production. It is a matter of optimism, or lack thereof, to consider this a likely or unlikely future. Seeing that sufficiently large LLMs are still tied to capital interests, I remain apprehensive. In this essay, I wanted to try to think about how language is changing in the technical age we inhabit today and which will continue to unfold—without fearing technology, but also refusing to succumb to its ideologies. In that context, one thing seems certain to me: with the increasing penetration of language technologies, with the triumph of AI models, our expectations as readers *will* change.

So here is a final question for you, dear reader: how do you react when I now tell you that I, too, have had large parts of this text written by AI? Do you feel deceived? Then you are still firmly at home in the standard expectation of the twentieth century. But I can reassure you: this

text was written without any AI assistance. Or was it? Can you be quite sure of that? If you are now undecided, then you are on the threshold of the second expectation, the doubt about the origin of a text in the age of large language models. But perhaps you are already indifferent to this alternative—maybe not completely, but at least to the extent that you can imagine what a world of *post-artificial* texts might look like.

## References

Alter, Alexandra, and Elizabeth A. Harris. 2023. “Franzen, Grisham, and Other Prominent Authors Sue OpenAI.” *New York Times*, September 20.

<https://www.nytimes.com/2023/09/20/books/authors-openai-lawsuit-chatgpt-copyright.html>.

Bajohr, Hannes. Forthcoming. “Writing at a Distance: Notes on Authorship and Artificial Intelligence.” *German Studies Review*.

Bajohr, Hannes. 2021. “Künstliche Intelligenz und digitale Literatur: Theorie und Praxis konnektionistischen Schreibens.” In *Digitale Literatur II*, edited by Hannes Bajohr and Annette Gilbert, 174–85. Munich: Text+Kritik. <https://doi.org/10.5771/9783967075496-174>.

Bajohr, Hannes. 2022. “The Paradox of Anthroponormative Restriction: Artistic Artificial Intelligence and Literary Writing.” *CounterText* 8, no. 2: 262–82. <https://doi.org/10.3366/count.2022.0270>.

Bajohr, Hannes. 2023. “Dumb Meaning: Machine Learning and Artificial Semantics.” *Image* 37, no. 1: 58–70. <https://doi.org/10.1453/1614-0885-1-2023-15452>.

Bajohr, Hannes. 2024. “Whoever Controls Language Models Controls Politics.” In *Training the Archive*, edited by Inke Arns, Eva Birkenstock, Dominik Bönisch, and Francis Hunger, 189–95. Cologne: Walther König.

Barker, Memphis. 2013. “What Is Horse\_ebooks? Twitter Devastated at News Popular Spambot Was Human after All.” *Independent*, September 24.

<https://www.independent.co.uk/voices/iv-drip/what-is-horse-ebooks-twitter-devastated-at-news-popular-spambot-was-human-after-all-8836990.html>.

Beals, Kurt. 2018. “‘Do the New Poets Think? It’s Possible’: Computer Poetry and Cyborg Subjectivity.” *Configurations* 26, no. 2: 149–77. <https://doi.org/10.1353/con.2018.0010>.

Bender, Emily M., and Alexander Koller. 2020. “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data.” In *Proceedings of the Fifty-Eighth Annual Meeting of the Association for Computational Linguistics*, 5185–98. Association for Computational Linguistics, July. <https://doi.org/10.18653/v1/2020.acl-main.463>.

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 *FAccT ’21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. New York: Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>.

Bense, Max. 1962. “Über natürliche und künstliche Poesie.” In *Theorie der Texte: Eine Einführung in neuere Auffassungen und Methoden*, 143–47. Cologne: Kiepenheuer & Witsch.

- Bernhart, Toni. 2020. "Beiwerk als Werk: Stochastische Texte von Theo Lutz." *Editio* 34, no. 1: 180–206. <https://doi.org/10.1515/editio-2020-0010>.
- Blumenberg, Hans. 2020. "Phenomenological Aspects on Life-World and Technization." In *History, Metaphors, Fables: A Hans Blumenberg Reader*, edited by Hannes Bajohr, Florian Fuchs, and Joe Paul Kroll, 358–99. Ithaca, NY: Cornell University Press.
- Bolter, Jay David. 1991. "Artificial Intelligence." In *Writing Space: The Computer, Hypertext, and the History of Writing*, 171–93. Hillsdale, NJ: Erlbaum.
- Bratton, Benjamin, and Blaise Agüera y Arcas. 2022. "The Model Is the Message." *Noēma*, July 12. <https://www.noemamag.com/the-model-is-the-message>.
- Brittain, Blake. 2023. "Judge Pares Down Artists' AI Copyright Lawsuit against Midjourney, Stability AI." *Reuters*, October 30. <https://www.reuters.com/legal/litigation/judge-pares-down-artists-ai-copyright-lawsuit-against-midjourney-stability-ai-2023-10-30/>.
- Campe, Rüdiger. 2021. "Writing; The Scene of Writing." *MLN* 136, no. 5: 971–83. <https://doi.org/10.1353/mln.2021.0075>.
- Chen, Mia Xu, Benjamin N. Lee, Gagan Bansal, Yuan Cao, Shuyuan Zhang, Justin Lu, Jackie Tsay, et al. 2019. "Gmail Smart Compose: Real-Time Assisted Writing." *arXiv*, May 17. <https://arxiv.org/abs/1906.00080>.
- Chun, Wendy Hui Kyong. 2021. *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge, MA: MIT Press.
- Cramer, Florian. 2014. "What Is 'Post-Digital'?" *APRJA* 3 (1). <http://www.aprja.net/?p=1318>.
- Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

- Dahl, Matthew, Varun Magesh, Mirac Suzgun, and Daniel E. Ho. 2024. "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models." *arXiv*, January 2.  
<https://doi.org/10.48550/arXiv.2401.01301>.
- Dhaliwal, Ranjodh Singh. 2023. "What Do We Critique When We Critique Technology?" *American Literature* 95, no. 2: 305–19. <https://doi.org/10.1215/00029831-10575091>.
- Dzieza, Josh. 2022. "The Great Fiction of AI: The Strange World of High-Speed Semi-automated Genre Fiction." *Verge*, July 20. <https://www.theverge.com/c/23194235/ai-fiction-writing-amazon-kindle-sudowrite-jasper>.
- electronus. 1960. "und kein engel ist schön." *Ja und Nein* 12, no. 3: 3.
- European Parliament. 2023. "Artificial Intelligence Act, Procedure File 2021/0106(COD)." Legislative Observatory, June 14.  
[https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106\(COD\)&l=en](https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106(COD)&l=en).
- Foucault, Michel. 1998. "What Is an Author?" Translated by Josué V. Harari. In *Aesthetics, Method, and Epistemology*, edited by James D. Faubion, 205–22. New York: New Press.
- GitHub. N.d. "GitHub Copilot." <https://github.com/features/copilot> (accessed January 7, 2023).
- Goodlad, Lauren M. E. 2023. "Humanities in the Loop." *Critical AI* 1, nos. 1–2.  
<https://doi.org/10.1215/2834703X-10734016>.
- Google. N.d. "Write with AI in Google Docs." Google Docs Editors Help.  
<https://support.google.com/docs/answer/13447609?hl=en> (accessed October 14, 2023).

- GPT-3. 2020. “A Robot Wrote This Entire Article: Are You Scared Yet, Human?” *Guardian*, September 8. <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>.
- Gunser, Vivian Emily, Steffen Gottschling, Birgit Brucker, Sandra Richter, and Peter Gerjets. 2021. “Can Users Distinguish Narrative Texts Written by an Artificial Intelligence Writing Tool from Purely Human Text?” In *HCI International 2021—Posters*, edited by Constantine Stephanidis, Margherita Antona, and Stavroula Ntoa: 520–27. Communications in Computer and Information Science 1419. Cham: Springer.
- Habermas, Jürgen. 1981. *The Theory of Communicative Action*. Translated by Thomas McCarthy. Vol. 1. Boston: Beacon.
- Hayles, N. Katherine. 2022. “Inside the Mind of an AI: Materiality and the Crisis of Representation.” *New Literary History* 54, no. 1: 635–66. <https://doi.org/10.1353/nlh.2022.a898324>.
- Heaven, Will Douglas. 2022. “Why Meta’s Latest Large Language Model Survived Only Three Days Online.” *MIT Technology Review*, November 18. <https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science>.
- Hegel, Georg Wilhelm Friedrich. 1977. *Phenomenology of Spirit*. Translated by A. V. Miller. Oxford: Oxford University Press.
- Henrickson, Leah. 2021. *Reading Computer-Generated Texts*. Cambridge: Cambridge University Press.

Horse\_ebooks (@horse\_ebooks). 2012a. “Everything happens so much.” Twitter, June 28, 10:23 p.m. [https://twitter.com/horse\\_ebooks/status/218439593240956928](https://twitter.com/horse_ebooks/status/218439593240956928).

Horse\_ebooks (@horse\_ebooks). 2012b. “Unfortunately, as you probably already know, people.” Twitter, July 25, 9:41 a.m. [https://twitter.com/Horse\\_ebooks/status/228032106859749377](https://twitter.com/Horse_ebooks/status/228032106859749377).

*Ja und Nein*. 1961. “So reagierten Leser.” 13, no. 1: 3.

Jasper.ai. N.d. “Jasper.” <https://www.jasper.ai/>.

Johnston, David Jhave. 2019. *ReRites: Human + A.I. Poetry. Raw Output*. Montreal: Anteism Books.

Katz, Daniel Martin, Michael James Bommarito, Shang Gao, and Pablo Arredondo. 2023. “GPT-4 Passes the Bar Exam.” SSRN Scholarly Paper. Rochester, NY. <https://doi.org/10.2139/ssrn.4389233>.

Kirschenbaum, Matthew G. 2023a. “Again Theory: A Forum on Language, Meaning, and Intent in the Time of Stochastic Parrots.” *In the Moment* (blog), June 26. *Critical Inquiry*. <https://critinq.wordpress.com/2023/06/27/again-theory-a-forum-on-language-meaning-and-intent-in-the-time-of-stochastic-parrots-2>.

Kirschenbaum, Matthew G. 2023b. “Prepare for the Textpocalypse.” *Atlantic*, March 8. <https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318/>.

Kittler, Friedrich A. 1999. *Gramophone, Film, Typewriter*. Translated by Geoffrey Winthrop-Young and Michael Wutz. Stanford, CA: Stanford University Press.

- Klütsch, Christoph. 2012. "Information Aesthetics and the Stuttgart School." In *Mainframe Experimentalism: Early Computing and the Foundations of the Digital Arts*, edited by Hannah B. Higgins and Douglas Kahn, 65–89. Berkeley: University of California Press.
- Knapp, Steven, and Walter Benn Michaels. 1982. "Against Theory." *Critical Inquiry* 8, no. 4: 723–42.
- Kuhn, Mattis. 2021. *Selbstgespräche mit einer KI*. N.p.: 0x0a.
- Latour, Bruno. 2007. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.
- Yann LeCun [@ylecun]. 2023. "I have claimed that Auto-Regressive LLMs are exponentially diverging diffusion processes. Here is the argument: Let  $e$  be the probability that any generated token exits the tree of 'correct' answers. Then the probability that an answer of length  $n$  is correct is  $(1-e)^n$ ." Tweet. *Twitter/X*.  
<https://twitter.com/ylecun/status/1640122342570336267>.
- Lewis, Danny. 2016. "An AI-Written Novella Almost Won a Literary Prize." *Smithsonian Magazine*, March 28. <https://www.smithsonianmag.com/smart-news/ai-written-novella-almost-won-literary-prize-180958577/>.
- Lindgren, Simon. 2024. *Critical Theory of AI*. Cambridge: Polity.
- Lu, Ning, Shengcai Liu, Rui He, Qi Wang, and Ke Tang. 2023. "Large Language Models Can Be Guided to Evade AI-Generated Text Detection." *arXiv*, May 18.  
<https://doi.org/10.48550/arXiv.2305.10847>.
- Lutz, Theo. 1959. "Stochastische Texte." *Augenblick* 4, no. 1: 3–9.



- Mitchell, Melanie, and David C. Krakauer. 2023. "The Debate Over Understanding in AI's Large Language Models." *Proceedings of the National Academy of Sciences* 120, no. 13. <https://doi.org/10.1073/pnas.2215907120>.
- Mok, Aaron. 2023. "It's Not Just Google: Closer Inspection Reveals Bing's AI Also Flubbed the Facts in Its Big Reveal." *Business Insider*, February 14. <https://www.businessinsider.com/bings-gpt-powered-ai-chatbot-made-mistakes-demo-like-google-2023-2>.
- Mueller, Kristen. 2014. *Partially Removing the Remove of Literature*. New York: & So.
- Natale, Simone. 2021. *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. Oxford: Oxford University Press.
- Nehamas. 1986. "What an Author Is." *Journal of Philosophy* 83, no. 11: 685–91.
- Olson, Emily. 2023. "Google Shares Drop \$100 Billion after Its New AI Chatbot Makes a Mistake." *NPR*, February 9. <https://www.npr.org/2023/02/09/1155650909/google-chatbot-error-bard-shares>.
- OpenAI. 2023. "Artist and Creative Content Owner Opt Out." Form. [https://share.hsforms.com/1\\_OuT5tffSpic89PqN6r1CQ4sk30](https://share.hsforms.com/1_OuT5tffSpic89PqN6r1CQ4sk30).
- Perez-Cruz, Fernando, and Hyun Song Shin. 2024. "Testing the Cognitive Limits of Large Language Models." *BIS Bulletin*, no. 83. <https://www.bis.org/publ/bisbull83.pdf>.
- Piantadosi, Steven T., and Felix Hill. 2022. "Meaning without Reference in Large Language Models." *arXiv*, August 5. <https://arxiv.org/abs/2208.02957>.
- Raley, Rita, and Jennifer Rhee. 2023. "Critical AI: A Field in Formation." *American Literature* 95, no. 2: 185–204. <https://doi.org/10.1215/00029831-10575021>.

- Rettberg, Scott. 2019. *Electronic Literature*. London: Polity.
- Roberge, Jonathan, and Michael Castelle, eds. 2021. *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*. Cham: Springer.
- Roberge, Jonathan, and Tom Lebrun. 2023. "Parrots All the Way Down: Controversies within AI's Conquest of Language." In *KI-Realitäten: Modelle, Praktiken und Topologien maschinellen Lernens*, edited by Richard Groß and Rita Jordan, 39–65. Bielefeld: transcript. <https://doi.org/10.14361/9783839466605-003>.
- Roose, Kevin. 2022. "An A.I.-Generated Picture Won an Art Prize: Artists Aren't Happy." *New York Times*, September 2. <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>.
- Roose, Kevin. 2023. "A Conversation with Bing's Chatbot Left Me Deeply Unsettled." *New York Times*, February 16. <https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html>.
- Sadasivan, Vinu Sankar, Aounon Kumar, Sriram Balasubramanian, Wenxiao Wang, and Soheil Feizi. 2023. "Can AI-Generated Text Be Reliably Detected?" *arXiv*, March 17. <https://doi.org/10.48550/arXiv.2303.11156>.
- Schwenger, Peter. 2019. *Asemic: The Art of Writing*. Minneapolis: University of Minnesota Press.
- Seabrook, John. 2019. "The Next Word: Where Will Predictive Text Take Us?" *New Yorker*, October 4. <https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker>.

- Shanahan, Murray. 2022. "Talking about Large Language Models." *arXiv*, December 7.  
<https://arxiv.org/abs/2212.03551>.
- Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson. 2023. "The Curse of Recursion: Training on Generated Data Makes Models Forget." *arXiv*, May 27. <http://arxiv.org/abs/2305.17493>.
- Smith, Brian Cantwell. 2019. *The Promise of Artificial Intelligence: Reckoning and Judgment*. Cambridge, MA: MIT Press.
- Stingelin, Martin. 2012. "UNSER SCHREIBZEUG ARBEITET MIT AN UNSEREN GEDANKEN: Die poetologische Reflexion der Schreibwerkzeuge bei Georg Christoph Lichtenberg und Friedrich Nietzsche." In *Schreiben als Kulturtechnik: Grundlagentexte*, edited by Sandro Zanetti, 83–104. Berlin: Suhrkamp.
- Tiku, Nitasha. 2022. "The Google Engineer Who Thinks the Company's AI Has Come to Life." *Washington Post*, June 11.  
<https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine>.
- Turing, Alan M. 1950. "Computing Machinery and Intelligence." *Mind* 59, no. 236: 433–60.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." *Advances in Neural Information Processing Systems* 30 (December): 5999–6009.  
[https://proceedings.neurips.cc/paper\\_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html).

Villalobos, Pablo, Jaime Sevilla, Lennart Heim, Tamay Besiroglu, Marius Hobbhahn, and Anson Ho. 2022. “Will We Run Out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning.” *arXiv*, October 26. <http://arxiv.org/abs/2211.04325>.

Weatherby, Leif. Forthcoming. “Two Autonomies of the Symbolic Order: Data and Language in Neural Nets.” In *Thinking with AI: Artificial Intelligence as an Intuition Pump for the Humanities*, edited by Hannes Bajohr. London: Open Humanities.

Weiser, Benjamin. 2023. “Here’s What Happens When Your Lawyer Uses ChatGPT.” *The New York Times*, May 27, 2023, sec. New York.  
<https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html>.

Welsh, Matt. 2023. “The End of Programming.” *Communications of the ACM* 66, no. 1: 34–35.  
<https://doi.org/10.1145/3570220>.

Whittaker, Meredith. 2021. “The Steep Cost of Capture.” *Interactions* 28 (6): 50–55.

Figure 1. Detail from *Ja und Nein*, “So reagierten Leser” (“This Is What Our Readers Said”). The two captions read: “Here, we present to you the author, ZUSE, an electronic brain” and “An original poem by ZUSE, in the poet’s handwriting.” The poem reads: “no kiss is silent / or love is silent / or no soul is pure / and not every kiss is green / and a youth is intense.”

Figure 2. The Sudowrite interface. On the right, a suggested continuation of the story.

Figure 3. Page from Kristen Mueller, *Partially Removing the Remove of Literature* (2014). Courtesy of Kristen Mueller.

Figure 4. Three pages from Mattis Kuhn, *Selbstgespräche mit einer KI (Monologues with an AI, 2021)*. On the left one of the literary outputs, in the middle the code of the language model’s encoder, and on the right an excerpt from the dataset the language model was trained with. Courtesy of Mattis Kuhn.