Media Culture and Cultural Techniques Working Papers
N°007
2023

# "Artificial and Post-Artificial Texts
## On Machine Learning and the Reading Expectations Towards Literary and Non-Literary Writing"

Hannes Bajohr

University of Basel
Department
Arts, Media, Philosophy

N° 2023.007

# “Artificial and Post-Artificial Texts On Machine Learning and the Reading Expectations Towards Literary and Non-Literary Writing”

Hannes Bajohr

DOI:10.12685/bmcct.2023.007

Media Culture and
Cultural Techniques

**working papers**

BMCCT veröffentlicht Arbeitspapiere im Forschungsbereich des Basler Seminars für Medienwissenschaft. Die Arbeitspapiere erscheinen in unregelmässigen Abständen in deutscher und englischer Sprache.

BMCCT publishes working papers in the research area of the Basel Seminar for Media Studies. The working papers appear at irregular intervals in German and English.

Cite this item:
Hannes Bajohr, “Artificial and Post-Artificial Texts. On Machine Learning and the Reading Expectations Towards Literary and Non-Literary Writing”, BMCCT working papers, (March 2023) No. 007 (DOI: 10.12685/bmcct.2023.007).

Editors: Ute Holl, Markus Krajewski, Magnus Rust

The Basel Media Culture and Cultural Techniques Working Papers (BMCCT working papers) are published by

Seminar für Medienwissenschaft
Universität Basel
Holbeinstrasse 12
4051 Basel
medienwissenschaft.philhist.unibas.ch
bmcct@unibas.ch

under the creative commons licence
https://creativecommons.org/licenses/by/4.0/

ISSN 2673-5792

Concept and design:
Mario Wimmer

Hosted by University of Basel library’s eterna server

University of Basel
Department
Arts, Media, Philosophy

# Artificial and Post-Artificial Texts

## On Machine Learning and the Reading Expectations Towards Literary and Non-Literary Writing

Hannes Bajohr

Collegium Helveticum, Zurich,
bajohr@collegium.ethz.ch

DOI: 10.12685/bmcct.2023.007

It gives me great pleasure to deliver this year's Walter Höllerer Lecture.[1] During his time as a professor of literature at the Technical University of Berlin, Höllerer was responsible for founding a journal that still exists today. Its title is a good description of what I am going to talk about: *Sprache im technischen Zeitalter*—language in the technical age. In the first issue, published in 1961, Höllerer defines the task of a rigorous study of literature that operates at the cutting edge of the present and precisely *in* the technical age: It should not fear technology nor regard it as its natural enemy; and it should not willingly submit to the ideologies of this technology.[2] I still consider both aspects to be solid guidelines for discussing the question: what is the state of language in the technological age in which we live *today*—an age that is characterized by the rise of machine learning and "artificial intelligence."

In 1961, the extent to which language technologies would one day be used was hardly foreseeable, but the development was certainly underway. Höllerer was cognizant of the very early stages of what is now known as "natural language processing," and he was far-sighted enough to wish that the humanities, too, pay attention to the technical treatment of language. His programmatic preface in the first issue of *Sprache im technischen Zeitalter* was followed by an essay by the Austrian computer pioneer Heinz Zemanek. Zemanek's article applied Höllerer's two guidelines quite concretely to the language technology of translation: not to be afraid of technology and

[1] This lecture was delivered, in a shorter version, as the 14th Walter Höllerer Lecture at the Technical University, Berlin, on December 8, 2022. I thank Hans-Christian von Herrmann, Eva Geulen, Sina Dell'Anno, and Jules Pelta Feldman for their help.

[2] Walter Höllerer, "Diese Zeitschrift hat ein Programm," *Sprache im technischen Zeitalter* 1, no. 1 (1961): 1–2.

not to be a prisoner of its ideology.[3] In order to process language at all, one must first assume that it is subject to rules that can at least approximately be taught to a computer; if language were *only* a great mystery, one might as well not attempt what had already yielded presentable results. In the same breath, however, Zemanek warns against the illusion of complete automatability: Language is complex, situational, often ambiguous, and always a matter of human interpretation. Simply automating its syntax—a problem that even today has yet to be fully solved—does not mean that its meaning can be captured. All automatic language technology is thus a precarious balancing act between the necessary fiction that language *can* be automated and the constant reminder that it *cannot*.

Zemanek illustrates the problem with a canonical example from the philosopher Yehoshua Bar-Hillel: "The box was in the pen."[4] Since "pen" has at least two meanings, two translations are also possible: "The box was in the enclosure." Or: "The box was in the writing device." We immediately realize that one of these sentences is obviously absurd, because we know about usual size relations, and that writing devices are typically smaller than enclosures. But the software does not know that. The use of language presupposes intelligence as active world-relation that is capable of resolving such synonymies. As long as language and intelligence are not automated together, Zemanek holds, a high-quality translation—one on par with humans—remains a "utopian goal."[5] And if one throws all caution to the wind and gives in to the seduction that emanates from merely sufficiently good results, one runs the risk of turning the useful fiction of language as automatable into a harmful ideology. For then it can happen that "the aesthetic impression of the result puts all doubts to sleep, ... but at the same time does not indicate difficult decisions, but simply decides them."[6] The result appears to be meaningful, but in reality, it is not; the power to make decisions is then handed over to the machine uncritically in a false trust in its competence.

This admonition still rings true today. And unlike in 1961, speech technology is now substantially more sophisticated. While it is still not intelligent—computers,

---

[3] Heinz Zemanek, "Möglichkeiten und Grenzen der automatischen Sprachübersetzung," *Sprache im technischen Zeitalter* 1, no. 1 (1961): 3–15. I thank Hans-Christian von Herrmann for drawing my attention to this article.

[4] Yehoshua Bar-Hillel, "The Present Status of Automatic Translation of Languages," in *Advances in Computers*, ed. Franz L. Alt, vol. 1 (New York: Academic Press, 1960), 158. Bar-Hillel's famous example gives as a plausible context the sentence: "Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy."

[5] Zemanek, "Möglichkeiten und Grenzen der automatischen Sprachübersetzung," 13.

[6] Ibid., 14.

now as then, do not really understand what they are doing—the latest AI models give the impression of intelligence more than ever before. This appearance has to do with how observers interpret the output, how they confront it, and how they infer from it the system that stands behind it. And this interpretation of an *appearance* is not only an ideological, but simultaneously an eminently *aesthetic* question—and this brings us back to Höllerer, who ascribed to literature and its analysis the function of reflecting language as an aesthetic construct precisely in its interaction with technology. For language in the technical age is not a purely technical matter. It is a social phenomenon, a bundle of meaning and connotations that determines cultural practices and not least reception traditions.

It is here that I would like to begin today, by asking what impact the current rapid advances in machine learning research are having on the way in which we deal with language, or, to be more precise, their impact on our *reading expectations*. In contrast to the times of Höllerer and Zemanek, we are now truly on the threshold of being surrounded by texts that are entirely artificial—while at the same time we continue to merge with our language technologies in our own writing, so that our text production is increasingly supported, extended, and partially taken over by assistance systems. Therefore, I want to discuss—quite speculatively, but always with an eye on the state of the art—two questions: first, what happens when we are confronted with *artificial* texts in addition to *natural* ones? How do we read a text that we can no longer be sure was not written by an AI? And second, what direction might this development take if, at some point, the distinction between natural and artificial itself becomes obsolete, so that we no longer even ask about it and instead read *post*-artificial texts?

## The Standard Expectation Towards Unknown Texts

Of course, the distinction between artificial and natural texts is not mine. At about the same time that Höllerer in Berlin and Zemanek in Vienna were thinking about the cultural and practical aspects of technical language processing, the philosopher and physicist Max Bense in Stuttgart introduced a very similar set of concepts. In his 1962 essay "On Natural and Artificial Poetry," (*Über natürliche und künstliche Poesie*) Bense considers how computer-generated literature differs from conventional

literature written by humans.[7] Bense focuses on the "mode of creation" behind these texts: what happens on the part of authors when they write a poetic text?

For Bense, this is clear in the case of *natural* poetry: in order 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 she says, she is always speaking herself. At the same time, in her speech, she always refers to the world. Taken together, poetic consciousness puts "being into signs," that is, the world into text, and ultimately guarantees that one is related to the other.[8] Without this consciousness, Bense holds, the signs and the relationship between them would be meaningless. This already reveals the connection to technical language processing: for, as Zemanek has shown with his example of translation, such a text carries no meaning—the word "pen" or the word "box" are only empty symbols for the system, operative variables that could also be called something else entirely.

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. Instead, such texts have a purely "material" origin—they can only be described in terms of mathematical properties such as 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 the rescue of a romantic idea of inexplicable human creative power. On the contrary, "the author as genius" is dead here. Instead,

---

[7] Max Bense, "Über natürliche und künstliche Poesie," in *Theorie der Texte: Eine Einführung in neuere Auffassungen und Methoden* (Köln: Kiepenheuer & Witsch, 1962), 143. An English translation can be found on my website, Max Bense, "On Natural and Artificial Poetry" (March 13, 2023).
https://hannesbajohr.de/en/2023/03/13/max-bense-on-natural-and-artificial-poetry-1962 (accessed March 16, 2023).

[8] Bense, 143. I interpret Bense as articulating an early (ontological) version of the symbol grounding problem, but linking it (provocatively) to a post-Romantic poetics, which serves as a negative foil to his avant-garde aesthetics. See Stevan Harnad, "The Symbol Grounding Problem," *Physica D: Nonlinear Phenomena* 42, no. 1–3 (1990): 335–46.

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. As with Zemanek, the assumption that the text-producing system has intelligence would be a category mistake—and here, moreover, even an aesthetic faux pas.

Bense himself was involved in several experiments with artificial poetry. The most famous of these were 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 are considered the first experiment with digital literature in the German-speaking world.[9] These texts are "stochastic" because they are randomly selected and assembled from a collection of vocabulary words—the fact that they stem from Kafka's C*astle* hardly makes the output any more meaningful. 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." In Bense's literary journal *augenblick,* Lutz printed selections of some of these.[10]

The "Stochastic Texts" were one of the first examples of natural language processing in Germany, and they proved that computers could process not only mathematical operations but also language. They were also artificial poetry in Bense's sense: no matter how many variations the program churns out, there seems to be 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, in any case.[11] But both knew how it had been produced. Whether its artificial origin can be recognized, whether it reveals itself in the

[9] See Kurt Beals, "'Do the New Poets Think? It's Possible': Computer Poetry and Cyborg Subjectivity," *Configurations* 26, no. 2 (2018): 149–77.

[10] Theo Lutz, "Stochastische Texte," *augenblick* 4, no. 1 (1959): 3–9.

[11] Instead—and this can be observed in many early experiments with such generative literature—their creators almost always saw themselves as authors and assigned the computer only the role of a tool, see Hannes Bajohr, "Writing at a Distance: Some Notes in Authorship and Artificial Intelligence," *RG Working Papers*, doi: 10.13140/RG.2.2.15152.64002.

"aesthetic impression," is less clear; the readers of the literary magazine *augenblick* were not compelled to ask this question: an accompanying essay enlightened them to all the details of its creation.

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.[12] Only the author's name "electronus" would have allowed one to guess who was behind this text; otherwise, the poem was placed on page 3 among the miscellanea, just like any other poem. Only in the next issue was solved what had not been obvious as a riddle: that a computer had written the text.

Obviously, Lutz was having fun. Along with a photo of the Zuse Z22 and a second poem "in the poet's handwriting" (that is, a teletype printout), he published a series of letters to the editor. Their authors—without knowing how the poem had come about—were quite divided in their assessment of it: "Perhaps you should reconsider opening the columns of your paper to such modern poets!," complained one, while another was, on the contrary, impressed by so much literary avant-garde: "Finally, something modern!" And 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 publication.[13]

What is evident in these reactions is what I would call the *standard expectation towards unknown texts.* The electronus poem was indeed artificial poetry in Bense's definition, a synthetic text without meaning mediated by an authorial consciousness. But because its readers were unaware of these conditions of 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 is precisely this: *that they come from a human being who wants to say something*.[14] To recognize a text as

[12] electronus [i.e. Theo Lutz], "und kein engel ist schön," *Ja und Nein* 12, no. 3 (1960): 3.

[13] "So reagierten Leser," *Ja und Nein* 13, no. 1 (1961): 3. I thank Toni Bernhart for sharing this finding with me; for the background, see Toni Bernhart, "Beiwerk als Werk: Stochastische Texte von Theo Lutz," *Editio*, no. 34 (2020): 180–206.

[14] This is very similar to Leah Henrickson's notion of the "hermeneutic contract". However, I contend that this contract implies that the instance writing is specifically *human*. See Leah

*artificial* still requires additional information—especially in the case of artificial poetry. Lutz had indeed "given his readers the run-around," as one letter to the editor insinuated—but not because a modern poet had written bad but natural poetry, but because a computer had generated a meaningless, because artificial text.

## Strong and Weak Deception

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 giving the run-round is the *ur*-principle of artificial intelligence—and at the same time that which connects it with language technologies. Ten years earlier, computer science pioneer Alan Turing had pondered in an article that became the founding document of artificial intelligence whether computers could ever think, ever be intelligent.[15] 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 human and thus intelligent.

The experiment's setup is well-known: a subject communicates with an absent second person via teleprinter and is supposed to find out whether it is a human or a machine.[16] Through the teleprinter, the subject can talk to the other side like, ask questions, and demand clarification. 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 human. But if one wants to examine the

---

Henrickson, *Reading Computer-Generated Texts* (Cambridge: Cambridge University Press, 2021), 28, https://doi.org/10.1017/9781108906463.

[15] Alan M. Turing, "Computing Machinery and Intelligence," *Mind* 59, no. 236 (1950): 433–60.

[16] Turing takes this setup from the "imitation game," in which the *gender* of the unknown person is to be guessed. Much has been made of this "passing," both in terms of Turing's own biography—as a gay man he was forced to undergo estrogen treatment, to which his suicide is probably related—and the gendered nature of AI more generally as the "obvious connection between gender and computer intelligence: both are in fact imitative systems, and the boundaries between female and male, I argue, are as unclear and as unstable as the boundary between human and machine intelligence." Jack Halberstam, "Automating Gender: Postmodern Feminism in the Age of the Intelligent Machine," *Feminist Studies* 17, no. 3 (1991): 443, https://doi.org/10.2307/3178281.

expectations of artificial texts, Turing's test is still a helpful starting point, since it equates intelligence with written communication,[17] 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 authorship.

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 mediated only through the aesthetic appearance-as-human—the Turing test is not conceivable without deception. For this reason, media scholar Simone Natale 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."[18]

I would like to call this position *strong deception.* You can see right away that there are problems that come with this position. 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, in Zemanek's sense, a technological ideology. 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.

Second, and more relevant to our topic, we can ask whether under these conditions expectations of AI-generated texts will ever change in the long run—and whether their change can be *described*. I think not. Indeed, the Turing test insists that artificial and natural texts remain neatly separated so that one can be considered as the other. If it is suddenly revealed that a natural text was in fact an artificial one, the audience will feel cheated. And not without reason: *Täuschung wird Enttäuschung*, deception turns into disappointment.

---

[17] The *essential* textuality of AI was already pointed out by Jay David Bolter in 1991: "Artificial intelligence is the art of making texts," Jay David Bolter, "Artificial Intelligence," in *Writing Space: The Computer, Hypertext, and the History of Writing* (Hillsdale, NJ: Erlbaum, 1991), 180.

[18] Simone Natale, *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test* (Oxford: Oxford University Press, 2021), 3. Emphasis mine.

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. The last time this happened was in June 2022 at a rather peripheral art prize: when a participant admitted that he had not painted his picture himself, but that it had been generated by the text-to-picture AI Dall-E 2, a torrent of indignation follwed and he was accused of fraud. Even though this was an art prize for *digital* art, it apparently only meant the tools; the art itself was still supposed to come from humans.[19] A similar case occurred in Japan in 2016, where 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.[20] There are other such examples—and although they are usually exaggerated in the press, as disappointed *expectations* these reactions reveal what was actually expected: namely natural, not artificial texts.

These expectations are also confirmed negatively: the disappointment comes about when a supposedly computer-generated piece is actually the work of a human being. Just one infamous example: Around 2011—during the early heyday of Twitterbots for the purpose 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 apercus: a literary bot against its will, seeming without any intended meaning. When it output something meaningful for human readers, it seemed all the more fascinating. Aphoristic gems as "everything is happening so much"[21] or "unfortunately, as you probably already know, people" are now firmly established in Internet lore.[22] 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 disappointment: the marvelous random sentences seemed somehow devalued. The knowledge that behind them stood "A REAL

[19] Kevin Roose, "An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy.," *The New York Times*, September 2, 2022, https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html.

[20] Danny Lewis, "An AI-Written Novella Almost Won a Literary Prize," Smithsonian Magazine, March 28, 2016, https://www.smithsonianmag.com/smart-news/ai-written-novella-almost-won-literary-prize-180958577/.

[21] @horse_ebooks, June 28, 2012, https://twitter.com/horse_ebooks/status/218439593240956928.

[22] @horse_ebooks, July 25, 2012, https://twitter.com/Horse_ebooks/status/228032106859749377.

HUMAN BEING," as the *Independent* wrote disconcertedly in all caps, dashed the hopes of accidental meaning in an otherwise meaningless artificial text.[23]

## The Crisis of the Standard Expectation

At first glance, such examples seem to suggest that the reading expectations towards 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 at all. But in fact, I believe that expectations are nevertheless already in the process of shifting. Because on the one hand the number of computer-generated texts is constantly increasing, and on the other hand we ourselves are writing more and more with, about, and through language technologies, we are on the way to a new expectation, or rather: a new doubt. The more artificial texts there are, the more the standard dissolves and the question of their origin must arise, even where we normally do not think about it.

This apparent contradiction 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. This includes the fact that they appear to be "intended" and worked-through to the smallest detail. 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," this apparent intentionality means that the standard expectation towards literary texts to this day is that they have authors as humans with communicative intent.[24] We know that there are exceptions—but nevertheless we, like the readers of Lutz's electronus poem,

[23] Memphis Barker, "What Is Horse_ebooks? Twitter Devastated at News Popular Spambot Was Human After All," The Independent, September 24, 2013, https://www.independent.co.uk/voices/iv-drip/what-is-horse-ebooks-twitter-devastated-at-news-popular-spambot-was-human-after-all-8836990.html.

[24] I use the term "authorship" here in a deliberately reductive way. I do not mean the "mode of being of discourse" and the "classificatory function" of work coherence and intellectual property, for which the concept of authorship is usually reserved in its emphatic function, Michel Foucault, "What Is an Author?," in *Aesthetics, Method, and Epistemology*, ed. James D. Faubion (New York: New Press, 1998), 211, 210. Instead, I assume a "causal" authorship (who produced this text and by what means?, see Bajohr, "Writing at a Distance") and ask about the reception-side awareness of this causality. My usage is thus of a lesser breadth than the "implicit author," which is also an artifact of the text, and thus cannot be resolved simply by a general notion of text. What I am interested in, then, is a sociology of text — the relation in the mind of the *empirical* reader.

assume that texts have human authors until we are taught otherwise. 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* texts that remain in the background, that are merely functional, and that do not assert themselves as products of human intent. For them, the Turing test is simply a false description of reality. It assumes strong deception as the only form of human-machine interaction and the artificial/natural partition as the only possible distinction between text types. But especially in dealing with interfaces, with the ideally invisible surfaces through which we communicate with machines, there already exist intermediate stages—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 often the norm.

Simone Natale has proposed the term *banal deception for* this phenomenon.[25] In contrast to what I have called strong deception, here the 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 is possible only if we treat her at least to some extent as such. Knowing this is not a contradiction that suddenly and unexpectedly destroys an illusion, as in the examples of competitions in which an AI participates. Instead, it becomes a condition of functionality: If I do not play along, Siri just will not do what I want.

The situation is similar with 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, on a fundamental level, similar to that 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 already suspecting intelligence behind it.[26] This would already lower the expectation 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 a consciousness involved.

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

[25] Natale, *Deceitful Media*, 4.

[26] From a systems-theoretical perspective, which, however, focuses solely on the concept of communication and deliberately excludes "human origins," this is nicely described by Elena Esposito, *Artificial Communication: How Algorithms Produce Social Intelligence* (Cambridge, Massachusetts: The MIT Press, 2022).

simultaneously. This slightly schizophrenic position quickly gives rise to the doubt I mentioned earlier: the more convincing artificial texts become, the more the aesthetic impression they make on us suggests something like consciousness, and the more difficult it becomes to feel comfortable in the limbo into which banal deception lures us. It is not even necessary to cite elaborate deep-fakes for this fact; it can be observed even in the most inconspicuous language technologies.

Among the tools we use incessantly today are the little helpers that assist us in our writing tasks and that we would hardly call intelligent: the spell check in our word processors underlines the most embarrassing mistakes in red; the predictive text function in our smartphones even completes words without asking, which occasionally 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 rather 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. 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 deception. Writer John Seabrook provides a striking illustration in an essay for *The New Yorker*.

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 felt caught by the machine: "[S]itting 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." [27]

[27] John Seabrook, "The Next Word. Where Will Predictive Text Take Us?," The New Yorker, October 4, 2019, https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker.

The feeling of shame Seabrook expresses in this passage is, objectively speaking, 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.[28] 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, it even had the effect of making 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.

This can eventually lead to the conviction that we are *actually* dealing with an intelligence—as in the case of Google employee Blake Lemoine, who claimed in the summer of 2022 that the voice AI he was working on had consciousness. The LaMDA chat system, Lemoine said, possessed the intelligence of an eight-year-old and had asked him to be considered a person with rights. Google apparently deemed such a statement damaging to its business and subsequently fired the employee.[29] So far, Lemoine's reaction seems to be rather the exception, although it is by no means rare, if you look at the breathless coverage of Microsoft's Bing chatbot, for example.[30] What this case does show, however, is that the sense of eeriness Seabrook spoke of is likely to intensify in the future. If artificial texts become *too* good—for instance, by appearing more thoughtful than their authors—and if, moreover, we know that computers are capable of writing such texts, a new standard expectation towards unknown texts is in prospect: it is the doubt about their origin. Instead of assuming a human source as a matter of course, or dismissing the question for the time being, the first thing we would want to know about a text would be: how was it made?

---

[28] That the matter is more complicated and that there is such a thing as "dumb meaning" in AI models, I explain in Hannes Bajohr, "Dumb Meaning: Machine Learning and Artificial Semantics," *IMAGE* 18, no. 37 (2023), forthcoming.

[29] See ibid.

[30] Kevin Roose, "A Conversation With Bing's Chatbot Left Me Deeply Unsettled" (February 16, 2023). *The New York Times*, https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html (accessed March 12, 2023).

## A Flood of Artificial Texts

This consideration merely follows a trend that intensifies with each new report of the capabilities of new language models. LaMDA has not yet been released to the public, but other models have. Their abilities would have been considered impossible five years ago; today they have become almost normal.

Any modern AI model based on machine learning is nothing more than a complex statistical function that makes predictions about likely future states based on learned data. In so-called language models, both the data learned and the predictions made consist of text. Such models have a wide range of uses, from linguistic analysis to automatic translation to the text generation (and, as a "foundation model," as an engine for even richer applications)[31]. But if Google's Smart Compose can only suggest a few words or phrases, *large* language models are capable of writing entire paragraphs and even coherent texts: And this is only because they learn which sentences and paragraphs are statistically most likely to follow each other.

By now, everyone is familiar with GPT-3, the large language model introduced by the company OpenAI almost three years ago. In one fell swoop—and with a great deal of publicity— it became clear that computers can generate texts that read almost as if they had been written by a human being. I say "almost" because even GPT-3 is far from perfect and makes a lot of mistakes; but its results were impressive enough that for a while articles in which the language model become the "author" and talked about "itself" became a journalistic genre of their own, spawning titles like, "A robot wrote this entire article. Are you scared yet, human?"[32]

In November 2022 and March 2023, OpenAI released updated versions: ChatGPT and GPT-4 are again more powerful than GPT-3.[33] When tasked to write a

---

[31] Rishi Bommasani et al., "On the Opportunities and Risks of Foundation Models." In: *arXiv,* 2021, http://arxiv.org/abs/2108.07258 (accessed March 12. 2023).

[32] GPT-3, "A Robot Wrote This Entire Article. Are You Scared yet, Human?," September 8, 2020, https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3. That GPT-3 figures as the author is a fiction, of course. As a disclaimer at the end of the article 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 or the other as a reference to an object in the world.

[33] Bothe are more performant, but only somewhat. It is my impression that the general enthusiasm and subsequent disenchantment with ChatGPT simply caught up with the experience that users with beta access to GPT-3 already had made. This still holds true for GPT; while it shows some improvements in logical reasoning and is supposed to be more reliable (how much time

student essay on Jorge Luis Borges, for example, the machine learning system goes to work without hesitation. The output is a perfectly acceptable text that contains no significant insights, but could pass as an introduction to a term paper. Moreover, because the system is dialog-based—it is a kind of better chatbot—I can ask ChatGPT to continue writing the text in a certain direction or to add references.

Various think pieces were quick to speculate that such language models will one day replace human authors; for various reasons, I doubt that.[34] But it does not have to come to that for our *perception* of text to change fundamentally. It is already a reality today that technologies like these are taking on assistive functions—not doing all the writing work, but helping to produce much more text much faster and with the help of fewer and fewer people. Certain types of writing are becoming at least *partially* automated.[35]

However, the main feature of the GPT models is not their technical prowess, but their economic integration. They are available through licensing models, and companies can pay OpenAI 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. With GitHub's Copilot, there already exists a sophisticated programming assistant.[36] From a brief description of a desired program routine, the model then writes the corresponding code. This does not always work, but it works 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.[37] Similar features exist for what we might have to call "ordinary writing." Just as I

---

will tell), they are, on the whole, gradual, while OpenAI's documentation seems to become worse and worse with each subsequent version, neither revealing the training corpus nor the architecture used, OpenAI, "GPT-4 Technical Report." In: *arxiv*, 2023, https://doi.org/10.48550/ARXIV.2303.08774 (accessed March 16, 2023).

[34] See Hannes Bajohr, "The Paradox of Anthroponormative Restriction: Artistic Artificial Intelligence and Literary Writing," *CounterText* 8, no. 2 (August 2022): 262–82, https://doi.org/10.3366/count.2022.0270.

[35] I do not mean to claim that writing was *un*assisted before machine learning, and every 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* difference between a typewriter and a large language model.

[36] "GitHub Copilot," GitHub, accessed December 13, 2022, https://github.com/features/copilot. Copilot is based on OpenAI's Codex, a GPT version especially trained on code.

[37] In the future, as one insider speculates, it is quite possible that "programming will be obsolete," Matt Welsh, "The End of Programming," *Communications of the ACM* 66, no. 1 (2023): 34–35.

can ask ChatGPT to continue the text, to rephrase it or to embellish it, the text editor Craft now incorporates an assistant that can revise what I have written by explaining it back to me, rewriting it, or summarizing it as bullet points.[38] Microsoft, too, licensed GPT—I already mentioned Bing—, and we will find such assistive functions in future versions of Word, going far beyond what we are used to in word processors.

Beyond mere assistance, however, large language models can be used profitably especially where the production of the most probable output is concerned. In particular, routine text work can be automated in this way. AI writing is therefore most advanced in an industry that produces a great deal of text, but attaches comparatively little importance to it, often viewing it as mere filler. In the past year, for example, dozens of speech AIs have appeared that are tailored to marketing: They are supposed to be able to use it to write ad copy and quickly produce large amounts 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 sounds like other texts of a similar type. [39]

But it becomes all the more difficult for readers to classify such texts as either human-made or machine-generated. The extent to which we can expect generated texts—Matthew Kirschenbaum warns of a veritable "textpocalypse"[40]—becomes clear when we consider how much of the writing that surrounds us every day is the product of such tedious routine tasks. As more of them circulate—and they undoubtedly will—the standard expectation towards unknown texts will shift from the immediate assumption of human authorship to a creeping doubt: did a machine write this?

Now, this may not be as much of a question 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? [41] What about political articles or fake news stories? What about the private,

[38] "Craft—The Future of Documents," accessed December 13, 2022, https://www.craft.do/.

[39] Just one example among many: "Jasper—AI Copywriting & Content Generation for Teams," accessed December 14, 2022, https://www.jasper.ai/.

[40] Matthew Kirschenbaum, "Prepare for the Textpocalypse," The Atlantic, March 8, 2023, https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318 (accessed March 12, 2023).

[41] The discussion about the use of ChatGPT for school and college assignments has, interestingly enough, drawn almost the widest circles in the popular discussion of large language models. This is astonishing insofar as the admission of operating a test regime based on *predictable* language should perhaps put this regime itself to the test. The expected arms race between language model and language model detection is in any case hardly conducive to pedagogical practice. (The measure recently reported from a New York school—it had summarily blocked

personal, intimate email? Is that an AI product, 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, vouch for their words. Even if it a text ultimately turns out to be inaccurate or misleading, the standard expectation that a recipient brings to reading it involves the assumption the that author is making what Jürgen Habermas has called a "validity claim … to truthfulness."[42] Essentially, it means that we have a basic level of trust that speakers (writers) mean what they say. This is the reason why reading critically has to be learned at all: our first inclination is to believe texts.

This becomes more difficult when large language models can generate texts that appear to have been produced and sanctioned by an author, but have no reliable knowledge of the world, only of the probability distribution of tokens. 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 again after only three days.[43] 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

---

the IP address for ChatGPT on school computers—seems symptomatically helpless, and certainly not on par of the ingenuity of students to circumvent measures to limit technology.) With respect to the *longue durée of* possible post-artificial texts that I am dealing with here, the argument that we are now experiencing the transition from the slide rule to the calculator, but for the humanities rather than math, seems plausible to me.

[42] Jürgen Habermas, *The Theory of Communicative Action*, trans. Thomas McCarthy, vol. 1 (Boston: Beacon, 1981), 52.

[43] Will Douglas Heaven, "Why Meta's Latest Large Language Model Survived Only Three Days Online," MIT Technology Review, November 18, 2022, https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/. For this reason, a lot still has to happen technically for ChatGPT to really be used as a reliable search engine. During the presentation of Google's prototype of such an AI-supported search—a system called Bard—it output a factually incorrect search result; Google then briefly lost $100 billion in market value, Emily Olson, "Google Shares Drop $100 Billion after Its New AI ChatbotMakes a Mistake," *NPR*, February 9, 2023, https://www.npr.org/2023/02/09/1155650909/google-chatbot--error-bardshares (accessed March 12, 2023). The Bing chatbot, too, produced falsehoods at its launch before it later began insulting journalists, Aaron Mok, "It's Not Just Google: Closer Inspection Reveals Bing's AI Also Flubbed the Facts in Its Big Reveal," *Business Insider* (Feb. 14, 2023). https://www.businessinsider.com/bings-gpt-powered-ai-chatbot-made-mistakes-demo-like-google-2023-2 (accessed March 12, 2023).

knowledge. It was predictive text pretending to be a database,[44] and had merely learned the *form* of scientific prose, without any scientific insight, responsibility, or eventual accountability.

## The Last Model and the Ouroboros

Sooner or later, the standard expectation of texts will shift—from the conviction that a human being is behind them to the doubt that it might not be a machine after all. But this will 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. *First*, "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,[45] but it is also 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.[46] On the other hand, no text is ever completely artificial—that would require real autonomy, an actually strong AI that could ultimately decide for itself to declare a text published.[47] 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 inextricable degree, we are integrating artificial text *with* natural text.[48]

---

[44] Murray Shanahan describes the difference 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'." Murray Shanahan, "Talking About Large Language Models." In: *arXiv*, 2022, http://arxiv.org/abs/2212.03551, 2 (accessed March 12, 2023)-

[45] Tobias Wilke, "Digitale Sprache: Poetische Zeichenordnungen im frühen Informationszeitalter," ZfL Blog, November 12, 2021, https://www.zflprojekte.de/zfl-blog/2021/10/12/tobias-wilke-digitale-sprache-poetische-zeichenordnungen-im-fruehen-informationszeitalter/.

[46] See Martin Stingelin, "UNSER SCHREIBZEUG ARBEITET MIT AN UNSEREN GEDANKEN. Die poetologische Reflexion der Schreibwerkzeuge bei Georg Christoph Lichtenberg und Friedrich Nietzsche," in *Schreiben als Kulturtechnik: Grundlagentexte*, ed. Sandro Zanetti (Berlin: Suhrkamp, 2012), 83–104.

[47] See Bajohr, "The Paradox of Anthroponormative Restriction."

[48] This meaning of "post-artificial" seems at first glance to be based on the term "post-digital". But while the latter focuses primarily on the difference between digital and analog

For after large language models, it is not implausible that the two types of text enter into a mutually dependent circular process that completely 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 most extreme case, at one point have been trained on *all* available language; according to one study, this may happen already in the next few years.[49] 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 linguistic 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, all subsequent language models for further performance gain will then learn from text that *itself* already comes from a language model.[50] Thus, one could say, natural language—even if only as a fiction that never existed anyway—would come to an end. For 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: Any linguistic innovation, any new word or every grammatical quirk that occurs regularly 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 most 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 *second* meaning of "post-artificial" and the one I am primarily concerned with here. After the tacit assumption of human authorship and the doubt about its origin, it would be the next expectation towards unknown texts. For doubt

technologies—which may already be automated—the former is primarily concerned with the human or non-human *origin of* an artifact, regardless of its specific *technical* substrate.

[49] Pablo Villalobos et al., "Will We Run Out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning." In: *arXiv*, 2022. https://arxiv.org/abs/2211.04325 (accessed March 12, 2023).

[50] Benjamin Bratton and Blaise Agüera y Arcas, "The Model Is The Message," *Noema*, July 12, 2022, https://www.noemamag.com/the-model-is-the-message.

about the origin of a text, like any doubt, cannot be permanent; humans have an interest in establishing normality, in reducing complexity and uncertainty to a tolerable level. This may be achieved, for example, by digital certificates, watermarks, or other security techniques designed to increase confidence that the text at hand is not just plausible nonsense.[51] Or simply by banning generated text that is not declared as such. Should political regulation and technical containment fail here, it is not unlikely that the expectation itself will become post-artificial. This means: Instead of suspecting a human behind a text, or being haunted by skepticism as to whether it was not a machine after all, the question simply becomes uninteresting: we then focus only on what the text says, not on who wrote it. Post-artificial texts would be agnostic about their origin; they would be *authorless by default*.[52]

So if the standard expectation towards unknown text is shifting; if it is becoming doubtful, perhaps turning agnostic as to its assumptions in a speculative future—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 everything was the same here, when so much is in motion?

I think because literature is *slower*. And this is because—Bense notwithstanding— of all text types, it makes the greatest claim to a human origin.

I have already said that there are texts today whose origins do not pose a question; a street sign has no author in this and in our daily life the weather forecast is also practically authorless. Up to now, we have always assumed that a human being is behind it—but under post-artificial reading conditions it makes little practical difference not to make any assumptions at all. In the future, more and more texts will be received in this way. One could also put it this way: *The zone of unmarked texts is expanding*. Not only street signs, but also blog entries, not only weather forecasts, but

[51] OpenAI already offers one such a solution: "AI Text Classifier," accessed February 3, 2023, https://platform.openai.com/ai-text-classifier. A future solution might involve embedding a specific watermark in AI-generated output: Since the distribution of the output words (or tokens) is not actually random, but follows a pattern that only appears to be arbitrary, a specifically produced distribution may serve as a watermark. Of course, all that would be needed is for a second, less sophisticated AI to be tasked with reformulating the output of the first, and that watermark would be erased, see Kyle Wiggers, "OpenAI's Attempts to Watermark AI Text Hit Limits," TechCrunch, December 10, 2022, https://techcrunch.com/2022/12/10/openais-attempts-to-watermark-ai-text-hit-limits/.

[52] What Foucault had already imagined in the sixties would finally have occurred: the question of authorship would have been lost in the "anonymity of a murmur," Foucault, "What Is an Author?," 222.

also information brochures, discussions of Netflix series, and even entire newspaper articles would tend to be unmarked, authorless.

Literary texts, on the other hand, are still maximally marked today. We read them radically differently from other types of texts—among other things, we continue to assume that they have 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 chess or Go, have been cracked: 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 at hand 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.[53] More interesting, then, is the question of the circumstances under which this difference becomes irrelevant. In other words, what would have to happen for *literature* to become post-artificial?

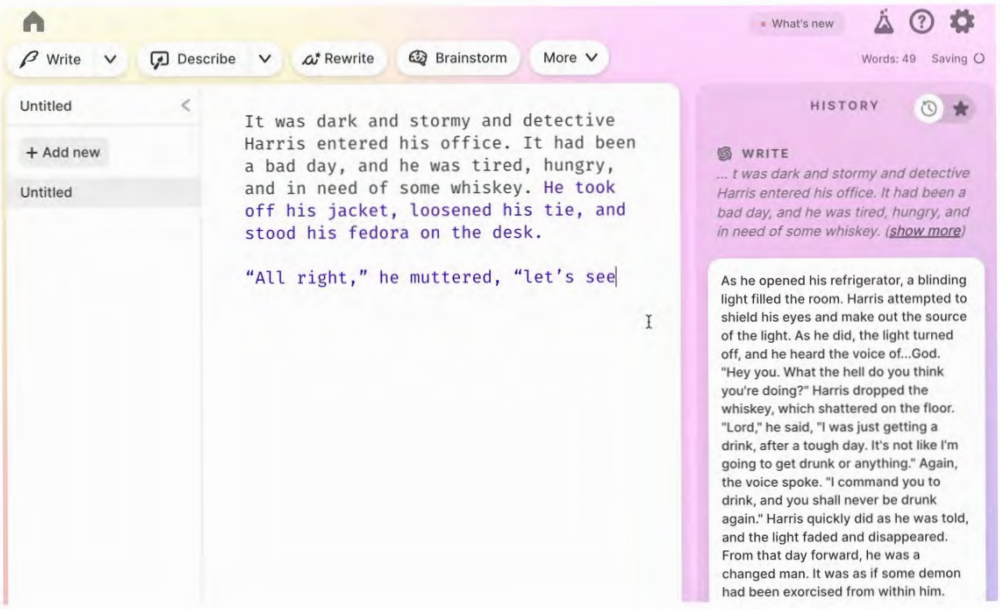
## What Is Post-Artificial Literature? And What Is Not?

I will answer this question by returning once again to the standardization tendency that arises from the Ouroboros effect of large language models. In them, as I said, normalization takes place. Their output is convincing precisely when they are supposed to spit out what is expected, what is ordinary, what is statistically probable. The more "ordinary" a writing task, the more easily it can be accomplished by AI language technologies. And just as there are assistive marketing AIs for expectable marketing prose, there are now also assistive literature AIs for more or less expectable literature.

[53] When one such study writes, "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* nature of future text, Vivian Emily Gunser et al., "Can Users Distinguish Narrative Texts Written by an Artificial Intelligence Writing Tool from Purely Human Text?," in *HCI International 2021—Posters*, ed. Constantine Stephanidis, Margherita Antona, and Stavroula Ntoa, vol. 1419 (Cham: Springer, 2021), 521.

"Expectability" can be described statistically as a probability distribution over a set of elements—the more recurrent they are, the more likely and expectable the outcome. Genre literature is virtually defined by the recurrence of certain elements, making it particularly suitable for AI generation. The website *The Verge* reported on the author Jennifer Lepp, who writes fantasy novels under the pseudonym Leanne Leeds—like an assembly line, one every 49 days.[54] She is aided by the program Sudowrite, a GPT-3-based literary writing assistant that continues dialogues, adds descriptions, rewrites entire paragraphs, and even provides feedback on human writing. The quality of this output is quite high, at least within the limits of expectability. Since all idiosyncrasies are averaged out in the mass of training data, LLMs tend toward a conventional treatment of language—they become Ouroboros literature themselves. At the moment, machine learning is not yet mature enough for generating 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; then it would be possible to reduce the 49 days to 49 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, time- and market-tested style, rather than actually indicating human origins. The unmarked zone would extend to certain areas of literature—not all, and certainly not all narrative ones, but far more than it does today.

[54] Josh Dzieza, "The Great Fiction of AI: The Strange World of High-Speed Semi-Automated Genre Fiction," The Verge, July 20, 2022, https://www.theverge.com/c/23194235/ai-fiction-writing-amazon-kindle-sudowrite-jasper.



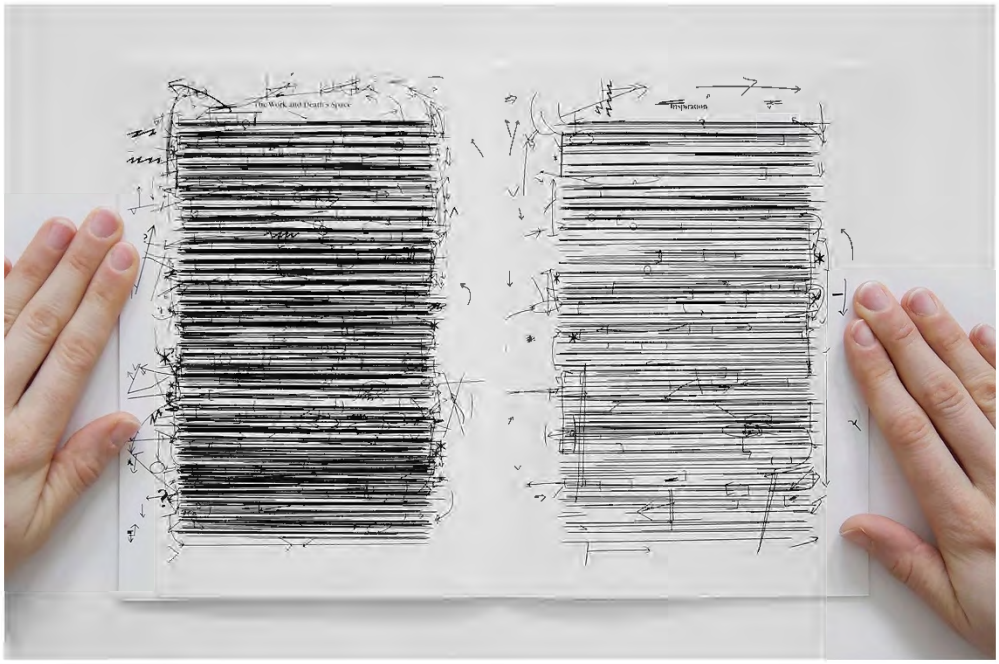

*The Sudowrite interface; on the right, a suggested continuation of the story.*

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

On the one hand, then, one could imagine the emphasis on human origins as a special feature. *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 2 or Stable Diffusion. On the one hand, because 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.[55] At the same time, however, there is also resistance to AI-

[55] For example, comic artist Sarah Andersen describes how her own work was part of the training set LAION (*Large-scale Artificial Intelligence Open Network*) for Stable Diffusion, which can now output images in her style. The name of the artists is thus "no longer attached to just his own work, but it also summons a slew of imitations of varying quality that he hasn't approved. … I see a monster forming." Sarah Andersen, "The Alt-Right Manipulated My Comic. Then A.I. Claimed It," *The New York Times*, December 31, 2022, sec. Opinion, https://www.nytimes.com/2022/12/31/opinion/sarah-andersen-how-algorithim-took-my-work.html.

generated art per se, which, some fear, threatens to make human artists obsolete. On Twitter, the hashtag *#supporthumanartists* has emerged as a declaration of war against generative image AI.[56] One can imagine something similar for 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, a kind of boutique writing would be conceivable that carries its human origin in front of it as a proof of quality and a selling point.



Kristen Mueller, *Partially Removing the Remove of Literature*.

But if one does not want to rely solely on the assurance of human origins—which, again, may leave room for the doubt that it might not be the case—it would be above all an unpredictable, unconventional use of language that would indicate a writing beyond the model. Every formal experiment, every linguistic subversion would oppose the probability of great language models, their leveling Ouroboros standard; linguistic unpredictability would then be evidence of human origin. In the most extreme

[56] A list of artists who decidedly do *not* work with AI can be found at https://whimsicalpublishing.ca/support-human-artists (as of 7.1.2023).

case, the sign system in which language AIs operate would be exploded—as in the case of visual and "asemic" literature, say, in the works of Kristen Mueller: she no longer uses any letters at all, but only the impression of lines and blocks of text.[57] The pure poetry Max Bense dreamed of would paradoxically not come from the machine, which now, in post-artificial blending, plausibly simulates meaning, but from people who no longer do so.


Mattis Kuhn, *Selbstgespräche mit einer KI*; on the left a poem text, in the middle the code and on the right an excerpt from the training dataset.

On the other hand, it is precisely the descendants of Lutz and Bense who escape the post-artificial by continuing to mark the artificiality of their products. This is digital literature—literature that is decidedly produced with the help of computers. It can escape the post-artificial by consciously emphasizing the entanglement between the natural and the artificial. Much more than conventional writing, digital literature always keeps a critical eye on its origins.[58] I have written about this in much greater

[57] Kristen Mueller, *Partially Removing the Remove of Literature* (New York: & So., 2014).

[58] See Hannes Bajohr and Annette Gilbert, "Platzhalter der Zukunft: Digitale Literatur II (2001 → 2021)," in *Digitale Literatur II*, ed. Hannes Bajohr and Annette Gilbert (München: edition text+kritik, 2021), 7–21. I discuss the examples mentioned here in more depth in Hannes Bajohr, "Künstliche Intelligenz und Digitale Literatur: Theorie und Praxis konnektionistischen Schreibens," in *Schreibenlassen: Texte zur Literatur im Digitalen* (Berlin: August Verlag, 2022), 191–213.

detail elsewhere, and give just two examples here: One is Mattis Kuhn's book *Selbstgespräche mit einer KI (Monologues with an AI)*,[59] in which, in addition to his literary experiments, he also provides the source code for training the language model and even its database. Not completely, but at least a little bit, the human and machine components that together produce in the text can be separated here.

Conversely, a deliberately staged human-machine collaboration can also have this analytical effect: In David "Jhave" Johnston's *ReRites*, for example, the author had a language model trained 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.[60] Of course, no "proof" of human intervention is ultimately provided here either. But perhaps the obstacles that can 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 really disappeared altogether.

It should have become clear that I have entered highly speculative territory here. I am not suggesting that narrative or, broadly speaking, conventional literature is doomed from now on, and that we should only pursue experimental or explicitly digital literature. Nor that post-artificial texts are necessarily bad—one will certainly enjoy reading them as well, discuss their merits, and unravel their interpretive dimensions. I was only interested here in analyzing tendencies, and there it is worthwhile to look at possible extremes. Above all, I wanted to try, in the spirit of Höllerer and Zemanek, to think about how language is changing in that technical age we inhabit today and which is still ahead of us—both without being afraid of technology, but also without succumbing to its ideologies. In any case, one thing seems certain to me: with the increasing penetration of language technologies, with the triumph of AI models, our reading expectations *will* change.

So here is a final question for you: 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

[59] Matthis Kuhn, Selbstgespräche mit einer KI (Berlin: 0x0a, 2021).
[60] David Jhave Johnston, ReRites: Human + A.I. Poetry. Raw Output (Montreal: Anteism, 2019).]

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 already on the threshold of the second expectation, the doubt about the origin of a text in the age of great language models. Or perhaps you are indifferent—maybe not entirely, but enough to imagine what a world of *post-artificial* texts might look like.

## Bibliography

"AI Text Classifier." Accessed February 3, 2023. https://platform.openai.com/ai-text-classifier.

Andersen, Sarah. "The Alt-Right Manipulated My Comic. Then A.I. Claimed It." *The New York Times*, December 31, 2022, sec. Opinion. https://www.nytimes.com/2022/12/31/opinion/sarah-andersen-how-algorithim-took-my-work.html.

Bajohr, Hannes. "Dumb Meaning: Machine Learning and Artificial Semantics," *IMAGE* 18, no. 37 (2023), forthcoming.

———. "Künstliche Intelligenz und Digitale Literatur: Theorie und Praxis konnektionistischen Schreibens." In *Schreibenlassen. Texte zur Literatur im Digitalen*, 191–213. Berlin: August Verlag, 2022.

———. "The Paradox of Anthroponormative Restriction: Artistic Artificial Intelligence and Literary Writing." *CounterText* 8, no. 2 (August 2022): 262–82. https://doi.org/10.3366/count.2022.0270.

———. "Writing at a Distance: Some Notes on Authorship and Artificial Intelligence." *RG Working Papers*. https://doi: 10.13140/RG.2.2.15152.64002.

Bajohr, Hannes, and Annette Gilbert. "Platzhalter der Zukunft: Digitale Literatur II (2001 → 2021)." In *Digitale Literatur II*, edited by Hannes Bajohr and Annette Gilbert, 7–21. München: edition text+kritik, 2021.

Bar-Hillel, Yehoshua. "The Present Status of Automatic Translation of Languages." In *Advances in Computers*, edited by Franz L. Alt, 1:91–163. New York: Academic Press, 1960.

Barker, Memphis. "What Is Horse_ebooks? Twitter Devastated at News Popular Spambot Was Human After All." The Independent, September 24, 2013. https://www.independent.co.uk/voices/ivdrip/what-is-horse-ebooks-twitter-devastated-at-news-popular-spambot-was-human-after-all-8836990.html.

Beals, Kurt. "'Do the New Poets Think? It's Possible': Computer Poetry and Cyborg Subjectivity." *Configurations* 26, no. 2 (2018): 149–77.

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

———. "On Natural and Artificial Poetry" (March 13, 2023). https://hannesbajohr.de/en/2023/03/13/max-bense-on-natural-and-artificial-poetry-1962.

Bernhart, Toni. "Beiwerk als Werk: Stochastische Texte von Theo Lutz." *Editio*, no. 34 (2020): 180–206.

Bolter, Jay David. "Artificial Intelligence." In *Writing Space. The Computer, Hypertext, and the History of Writing*, 171–93. Hillsdale, NJ: Erlbaum, 1991.

Bratton, Benjamin, and Blaise Agüera y Arcas. "The Model Is The Message." Noema, July 12, 2022. https://www.noemamag.com/the-model-is-the-message.

"Craft - The Future of Documents." Accessed December 13, 2022. https://www.craft.do/.

Dzieza, Josh. "The Great Fiction of AI: The Strange World of High-Speed Semi-Automated Genre Fiction." The Verge, July 20, 2022. https://www.theverge.com/c/23194235/ai-fiction-writing-amazon-kindle-sudowrite-jasper.

electronus [i.e. Theo Lutz]. "und kein engel ist schön." *Ja und Nein* 12, no. 3 (1960): 3.

Esposito, Elena. *Artificial Communication: How Algorithms Produce Social Intelligence*. Cambridge, Massachusetts: The MIT Press, 2022.

Foucault, Michel. "What Is an Author?" In *Aesthetics, Method, and Epistemology*, edited by James D. Faubion, 205–22. New York: New Press, 1998.

GitHub. "GitHub Copilot." Accessed December 13, 2022. https://github.com/features/copilot.

GPT-3. "A Robot Wrote This Entire Article. Are You Scared yet, Human?," September 8, 2020. https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

Grant, Nico, and Cade Metz. "A New Chat Bot Is a 'Code Red' for Google's Search Business." *The New York Times*, December 21, 2022, sec. Technology. https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html.

Gunser, Vivian Emily, Steffen Gottschling, Birgit Brucker, Sandra Richter, and Peter Gerjets. "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, 1419:520–27. Cham: Springer, 2021.

Habermas, Jürgen. *The Theory of Communicative Action*. Translated by Thomas McCarthy. Vol. 1. Boston: Beacon, 1981.

Halberstam, Jack. "Automating Gender: Postmodern Feminism in the Age of the Intelligent Machine." *Feminist Studies* 17, no. 3 (1991): 439–60. https://doi.org/10.2307/3178281.

Harnad, Stevan. "The Symbol Grounding Problem." *Physica D: Nonlinear Phenomena* 42, no. 1–3 (1990): 335–46.

Heaven, Will Douglas. "Why Meta's Latest Large Language Model Survived Only Three Days Online." MIT Technology Review, November 18, 2022. https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/.

Henrickson, Leah. *Reading Computer-Generated Texts*. Cambridge: Cambridge University Press, 2021. https://doi.org/10.1017/9781108906463.

Höllerer, Walter. "Diese Zeitschrift hat ein Programm." *Sprache im technischen Zeitalter* 1, no. 1 (1961): 1–2.

"Jasper - AI Copywriting & Content Generation for Teams." Accessed December 14, 2022. https://www.jasper.ai/.

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

Kirschenbaum, Matthew. "Prepare for the Textpocalypse." *The Atlantic*, March 8, 2023. https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318.

Kuhn, Mattis. *Selbstgespräche mit einer KI.* Berlin: 0x0a, 2021.

Lewis, Danny. "An AI-Written Novella Almost Won a Literary Prize." Smithsonian Magazine, March 28, 2016. https://www.smithsonianmag.com/smart-news/ai-written-novella-almost-won-literary-prize-180958577/.

Lutz, Theo. "Stochastische Texte." *augenblick* 4, no. 1 (1959): 3–9.

Mueller, Kristen. *Partially Removing the Remove of Literature*. New York: & So., 2014.

Natale, Simone. *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. Oxford: Oxford University Press, 2021.

Roose, Kevin. "An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy." *The New York Times*, September 2, 2022. https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html.

Seabrook, John. "The Next Word: Where Will Predictive Text Take Us?" The New Yorker, October 4, 2019. https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker.

Shanahan, Murray. "Talking About Large Language Models." arXiv, December 11, 2022. http://arxiv.org/abs/2212.03551.

"So reagierten Leser." *Ja und Nein* 13, no. 1 (1961): 3.

Stingelin, Martin. "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, 2012.

Turing, Alan M. "Computing Machinery and Intelligence." *Mind* 59, no. 236 (1950): 433–60.

Welsh, Matt. "The End of Programming." *Communications of the ACM* 66, no. 1 (2023): 34–35.

Wiggers, Kyle. "OpenAI's Attempts to Watermark AI Text Hit Limits." TechCrunch, December 10, 2022. https://techcrunch.com/2022/12/10/openais-attempts-to-watermark-ai-text-hit-limits/.

Wilke, Tobias. "Digitale Sprache: Poetische Zeichenordnungen im frühen Informationszeitalter." ZfL Blog, November 12, 2021. https://www.zflprojekte.de/zfl-blog/2021/10/12/tobias-wilke-digitale-sprache-poetische-zeichenordnungen-im-fruehen-informationszeitalter/.

Zemanek, Heinz. "Möglichkeiten und Grenzen der automatischen Sprachübersetzung." *Sprache im technischen Zeitalter* 1, no. 1 (1961): 3–15.