GENERATIVE AI: BETWEEN TOOL AND COMMUNICATION PARTNER

Benedikt Zönnchen
Hochschule München University of Applied Sciences
Munich Center for Digital Sciences and AI

April 5, 2024

1 INTRODUCTION

In November, I had the opportunity to briefly illuminate the field of generative artificial intelligence in two impulse lectures. The first lecture dealt with its application in so-called *extended reality* *XR für den Mittelstand* (2023) and the second was dedicated to the evolving challenges and opportunities for *higher education*, as part of the *Dialogforum Generative KI und Hochschule* (2023). I discussed the various possibilities of creating new media content through generative AI—from text-to-image, image-to-image, to text-to-video generators. I made a forecast on how far these tools will change both domains, including the question of how the quantity and quality of the generated media content may change.

The extraordinary capabilities of this technology both impress and frighten many people. As these opaque systems generate incredibly enticing texts, even some engineers of certain language models suspect a “ghost in the machine”. Talk has turned to an “era of conscious Terminators”. However, such speculations about the purportedly dangerous intelligence of machines cloud our vision of the essential aspects and major challenges.

Against this backdrop, my mini-lectures aimed to introduce a new perspective that I picked up from (Esposito, 2022) and find particularly enlightening in the debate about generative AI. This article is dedicated to that very perspective. It may bridge the gap between current technology, its application, and possible impacts.

2 WHAT IS GENERATIVE AI?

The following informal description is intended to convey an intuition. For an extensive review I refer to (Gozalo-Brizuela & Garrido-Merchan, 2023).

When we talk about generative AI today, we usually mean IT systems based on deep generative models, that is, we talk about *machine learning* or more precisely *deep learning*. The defining feature of these models is their ability to generate new data instances[1] that are plausible according to a "learned" probability distribution. Essentially, this means it is statistically very difficult to distinguish whether the generated data instance comes from the training data[2] or not. Generative models do not simply retrieve a dataset from the training data, but rather generate novel outputs by drawing it from a probability distribution. This distribution is modelled by the training data, the model architecture, and the training process—in short, by the trained model. Therefore, generated instances are neither arbitrary nor exactly determined. The models are *deep* because they are based on a neural network consisting of multiple layers of interconnected units or neurons.

Unlike models that are supposed to classify an input, such as calculating whether an image displays a cat or a dog, these models generate new images of cats and dogs. To discriminate between cat and dog images, it is not necessary to know the entire distribution of pixels of cat or dog images; it suffices if the boundaries between dog and cat images are "learned". For generation, however, an approximation of the probability distribution must be "learned". Consequently generation is significantly more challenging than discrimination, which is reflected in the necessary amount of data, training time, and size of generative models.

Today's large language models like ChatGPT, a class of generative models that particularly interests me in this article, generate text. If we draw many data points from the modelled distribution of one of these models, we obtain texts whose (empirical) distribution appears as if they come from the modelled distribution. In the case of chatbots, the entire distribution is conditioned by the user's input, i. e., our input is the beginning of the drawn text and thus determines what can come next. Moreover, the distribution modelled by the chatbots has been *aligned* so that they are suitable as assistants—usually, a question is followed by an answer, not another question. It's also interesting that about the same amount of computational work is done for each generated word[3], so it might make sense to let the model chatter a bit.

Based on the autoregressive method of successively predicting the next word, combined with a consistently similar computational effort required for each word, it's difficult to see that these language models operate similarly to our "mind". Humans typically consider which concept or idea they want to express and then find a way, i. e., a verbal and thus abstract representation to it. This representation is discrete to be able to communicate over a noisy medium. If we, instead, respond reflexively like a language model, we are in a different mode of operation—we are in the fast thinking mode as described by Kahneman (2011). The same system is active when we catch a ball—we just do it (subconsciously), without performing or solving a differential equation. Critics might claim that large language models seem so intelligent because we

[1] The model's output.

[2] The data by which the model is being optimized.

[3] Actually, the model generates tokens and a token may only represent a part of a word.

rarely engage in slow thinking. They are just better at providing reflexive answers.

It was ultimately very surprising that by processing or modelling language using gigantic models, researchers were able to create systems with remarkable abilities that we intuitively classify as "intelligent". The spark was ignited with GPT-2 (Radford et al., 2019), when it seemed that learning language could be a form of "*multitask learning*". This was particularly surprising to developers and researchers alike. Last year, the talks about "Sparks of artificial general intelligence: Early experiments with GPT-4" (Bubeck et al., 2023) began. The publication mentioned lists the extraordinarily impressive abilities, ranging from programming skills and mathematical competencies to composing music. However, all this should be taken with caution because, as Bender and Koller (2020) note, processing language is not equivalent to understanding language. Nonetheless, it's noteworthy that these systems have surpassed "common sense" for linguistic thought tasks over the years. They solve tasks that many thought could only be handled by a machine that thinks in the way we typically reserve for humans.

So if Bender and Koller (2020) are correct, which I assume, and these models do not *understand* in the sense of human understanding, then do we find a better term for what they do?

3 ONLY A TOOL?

In the aforementioned lectures, I advocated the following thesis:

> *Systems of generative AI can be arranged on a spectrum from tool to communication partner.*

This idea is thoroughly inspired by Elena Esposito and her book *Artificial Communication* (Esposito, 2022). I recalled her contribution as I listened to Björn Ommer's lecture at re:publica 2023. Ommer advocated the sensible view that this new technology should be seen as a tool, a view most in the technology sphere hold (see, e. g., Epstein et al., 2023).

As a side note, I should mention that Ommer is now striving to give a chance to smart algorithms and small models, thus also to democratization (Brühl, 2023)—an effort I can only support.

The reason for categorizing generative AI as a *tool* seems coherent: We should counteract the anthropomorphisation of the machine. Metaphors like "artificial intelligence", "hallucinating", "learning", and "teaching" obscure the fact that we deal with constructed, lifeless machines. In particular, the perception of human-like agency can undermine the recognition of creators, whose work underlies the system's output, and distract from the responsibility of developers and decision-makers when these systems cause harm. Therefore, we talk about generative AI as a tool to support creative designers instead of an actor capable of having its own intentions or authorship (Epstein et al., 2023). Of course, this does not mean that we do not need to adjust our understanding of authorship and intellectual property or protect this property with new regulations. Understanding generative AI as a tool suggests that it will have similar impacts as the camera and other tools before it. Just as the camera made portrait drawing partially obsolete but also created new opportunities, Generative AI will replace current methods of work with new ones—this is the general thesis.

Epstein et al. (2023) emphasize that this viewpoint leaves little room for the "creative" machine, which in turn raises the question of whether we place humans on too high or unique of a pedestal. Are artificial or biological processes, such as the formation of an ant track or flocking behaviour, excluded from creativity? At least, it's hard to deny that many processes that occur without our involvement exhibit a certain intelligence. Also, the strong emphasis on the individuality of artists is incompatible with systemic ways of thinking, which are considered essential for sustainable development (compare, e. g., Pasqualino, Jones, Monasterolo, & Phillips, 2015; Skene, 2020).

Looking back in time, we can repeatedly observe fears that have arisen through technological development. An interesting case is the introduction of the alphabet, which some thinkers at the time had seen as "the destruction of memory". In retrospect, this sounds quite absurd today. In a way, however, the critics were right. People lost the ability to remember epochal stories. But they were also wrong, because memory did not disappear. It was rather transferred from the "mind" to paper; from the human, as a biological being, into his environment.

Modern humans are situated in a network of essential tools and social, biological as well as ecological systems. A sharp separation between the atomic individual as a combination of body and "mind" and its environment seems increasingly difficult. In the spirit of a systems view, technology can be understood symbiotically: Societies irritate the evolution of technology, and technology irritates the evolution of society.

So, is generative AI then just a tool? Does it do the same as writing and the camera?

4 ARTIFICIAL COMMUNICATION

Looking back, one could say that the dilemma of the term *artificial intelligence* began with the so-called *Turing test*. The test proposes that a machine or algorithm can be considered "intelligent" if people cannot tell whether they are interacting with a person or an algorithm / machine. The test is based on intelligence tests and treats AI as a black box. It's worth mentioning that Turing's original version envisioned three players: a person with characteristic A, a person with characteristic B, and a questioner who must determine which of the individuals possesses characteristic A or B. Here, the role of person A might be played by a machine. Does the success rate change significantly if person A is mimicked by a machine?

The test doesn't reveal whether this intelligence is simulated (weak AI) or genuine (strong AI)—whatever "genuine" may mean. Many would claim that today's large language models would pass the test. Consequently, these models would be considered intelligent. At the same time, few developers of these systems would claim that these algorithms can actually understand or think, even if the technical terminology of machine learning suggests so. Such claims are hard to find outside of advertising texts. On the contrary, researchers and developers explicitly state that they are not trying to create human intelligence, and according to Esposito (2022) and Rosengrün (2021), they are so successful precisely because they have deviated from this path. They are doing what the engineers of airplanes did, who moved away from trying to construct birds.

The reason why language models may inevitably reach their limits is explained in a philosophical essay by Browning and LeCun (2022). The authors reiterate many old wisdoms. In a way, it's the misstep of the early Wittgenstein (1889–1951), which the late Wittgenstein corrected. Wittgenstein, as well as other thinkers of the 19th and 20th centuries, such as Russell (1872–1970) and Frege (1848–1925), believed at some point that language can express everything we can know, i. e., that theoretically all knowledge could be stored in a library. Wittgenstein basically destroyed such dreams and revised his own project by introducing his *language games*[4]. Later, it was Hubert Dreyfus (1929–2017) who, with moderate success, tried to alert the early AI developers to these limitations of symbolic systems (see, e. g., Dreyfus & Dreyfus, 1986)[5]. In the end, it can be said that Dreyfus contributed to the collapse of the symbolic AI era.

With artificial neural networks, the field partly moved away from the symbolic approach, but now it has arrived at a somewhat similar point again. As Browning and LeCun (2022) argue:

> "All representational schemas involve a compression of information about something, but what gets left in and left out in the compression varies. The representational schema of language struggles with more concrete information, such as describing irregular shapes, the motion of objects, the functioning of a complex mechanism or the nuanced brushwork of a painting."
> – (Browning & LeCun, 2022)

This thought sounds like a phenomenological argument, suggesting that we cannot easily separate the world and intelligence.

Instead of using a huge database of logically connected sentences as symbolic AI did, language models "learn" the role of words only in relation to other words. It's astonishing and insightful to see how far one can get with this approach. However, due to this linguistic limitation, many believe that an AI capable of actual planning will eventually need access to a *world model* and be able to simulate situations with it. If Dreyfus is right, this is impossible, at least on a human level. He did not believe that there is something like a world model in our head. Instead, according to Dreyfus, only the "real" world itself can be the model we seek. However, following the constructivist view[6], such as that inherited by Niklas Luhmann, there's a great chance he was

[4]Wittgenstein describes linguistic utterances as complex actions, where beyond the knowledge of words, an understanding of their function (in the real or constructed world) in a specific context is required.

[5]Dreyfus draws on the phenomenology of Husserl, Heidegger and Merleau-Ponty.

[6]Reality emerges as a result of the observers' construction. Of course, this does not mean that reality does not exist.

mistaken in this respect. Whether such a world model can be build and leads to an intelligent machine remains to be seen. Perhaps it requires a combination of symbolic and data-driven modelling, where the ability to perform symbolic manipulations can potentially be "learned" (Marcus, 2022)—however, starting with symbols seems to lead to the old problem the old AI systems suffered from.

If today's machines and chatbots are not intelligent or thinking entities, what are they doing? At this point, Esposito introduces Luhmann's concept of communication[7]. Defined constructively, according to Luhmann, communicators do not necessarily have to understand each other. Instead, the psychological system (the "mind") of the receiver is irritated by the content of the communication. The meaning of the communication is constructed by the receivers and is not objectively "out there" in some independent reality. What a text communicates to me may differ significantly from what the author intended to convey. Moreover, the interpretation by, for example, literary critics can be considered more meaningful and go beyond what the author originally intended to express. The critic's interpretation can be very different from the authors intention. Therefore, to be a talented artist, it is not necessarily required to understand the great meaning of ones own art since it is constructed by the observers.

Esposito introduces the concept of *artificial communication* to describe that generative AI is not intelligent but has acquired the ability to participate in communication[8] without understanding what is being communicated, just as paper represents a memory that cannot think. This way, she circumvents the challenging-to-define concept of intelligence.

The term *artificial communication* carries significantly less magic with it. There's less need for demystification by the developers of said systems and other entities, like us educators. I also believe that the term communication better explains or makes understandable why it can be effective, even though one of the communication partners may not grasp the meaning that the other party derives from the communication. In short, it seems irrelevant for effective communication whether it can be understood by both the sender and the receiver. What matters is only whether the content *makes sense* to the receiver. Actually, this sounds quite trivial but falls behind in discourse when we talk about "ChatGPT understanding what I mean". We assume that it must understand, otherwise it couldn't provide an answer from which I can *construct sense*. The phrase *"to understand"* should, if at all, only be used metaphorically.

In addition, the term *artificial communication* avoids ascribing too much control over the tool to humans and clarifies what we are dealing with and what we may need to prepare for: Neither the image of AI as a superhumanly intelligent being wanting to conquer the world, nor that of a tool fully controllable by us, represents, in my opinion, a particularly useful notion of generative AI.

[7]Luhmann defines the concept very broadly, for example, a payment or grading is a form of communication. For Luhmann, all social systems, such as the economic system or the education system, are communication systems.

[8]Communication can depend on different media such as images or language

5 (OUT OF) CONTROL

With the (unreliable) *communication partner*, one can push back against the concepts of *tool* and *intelligent being*. But what distinguishes Generative AI from a tool?

As Esposito (2022) rightly notes, we expect a tool to work (deterministically). Tools disappear from our perception; they become a kind of extension of our body. When writing this text, I do not think about the keyboard, but focus on my thoughts and how the individual letters appear on the screen in the area of my text editor. From a clock, I expect it to show the correct time. And from a camera, I expect it to produce realistic images. Only when the clock shows the wrong time, pressing the keys no longer makes letters appear, or the camera only produces black images, do I become aware of the tool again. Tools become visible when they do not function as we expect.

In contrast, communication is *contingent*, meaning

> "[...] something that is neither necessary nor impossible; something that can be as it is (was, will be), but can also be different. The concept thus denotes given things (to be experienced, expected, thought, fantasized) with regard to possible otherness; it denotes objects within the horizon of possible variations." – (Luhmann, 1987, S. 152)

This definition fits particularly well with the autoregressive nature of many models, such as all known large language models. Each outputted word narrows down what can come next. In the case of diffusion models, such as Dall-E, Stable Diffusion, each small step of reducing noise limits what is still possible.

When I communicate with someone, I do not know what answers I will get—otherwise, the communication would hardly be useful for gaining information. Also, my communication partner does not know how I will react, and we both are aware of this circumstance. Luhmann calls this *the problem of double contingency*. Unlike with a tool, I expect this contingency. I might be irritated if my partner says something odd. And there are many rules and behaviours to deal with the problem of double contingency, e. g. establishing a rule to drive on the right side of the road. However, the possibility of misunderstanding is always present but, the answers, as I construct them through the content of communication, are also not arbitrary, i. e., without any restriction. For instance, I can only make sense out of certain sentences.

Similarly, generated images, e. g., produced by Midjourney, are not a purely random result. At the same time, there's a realm of possibilities—a multidimensional probability distribution over a latent space, defined by the learned model parameters—that I can not escape from. For users, the systems are as opaque as the internal operations of human communication partners.

If the user can very precisely control what result the AI system delivers, the system moves towards *tool* on my imaginary scale. In the opposite direction lies the *communication partner*. If we move too far in this direction, the model's output becomes incomprehensible, and communication fails. Here we should probably distinguish between variations and surprises. I can design an algorithm that generates a different image for each run. If it merely produces variations of the same principle or generates nonsensical images, the irritation is too small or too large, respectively. Similar to an artwork, the result attracts attention if it *makes sense* and yet surprises the audience. The captivated anticipation for the result of the next prompt may arise from this property: I do not know what will happen, yet I have a good feeling for what cannot happen.

When using this technology, it means for me that we cannot have both: The *tool* will not generate surprises—aside from the surprise of its performance—and the *artificial communication partner* will not deliver exact results. Depending on the application and use, we can ask where the system should be on this spectrum. For instance, it initially sounds sensible that chatbots should only deliver reliable information—just the facts! But is that how effective communication works? Is the contingency of communication perhaps necessary or at least desirable? If we assume that we all *construct* (or make) our own sense, then a good communication partner should take this into account. And what about AI systems in the cultural sector? Should they perhaps represent an unreliable communication partner, who through "hallucinations" and other means offers the observer the opportunity to construct meaning, making contingency visible and thus challenging existing orders?

Another important relationship that falls short in the discussion about generative AI as a mere tool is the special position that large language models occupy. Even if they do not understand what makes sense at the other end, it might be possible that artificial and social systems could be coupled through the processing of language in the future. The first time in human history we can almost fluently speak with machines. This coupling could potentially resemble the coupling of social and psychological systems, as understood by Luhmann. Soon, the interface to various digital systems can likely be realized as a voice interface with all the advantages and disadvantages of the contingent character of communication. However, coupling also means that a symbiosis could form, similar to the symbiosis between social networks and their users.

Freed from the competitive idea between human and machine "intelligence", we can ask what we expect from a good communication partner. For example, there's *connectivity*, meaning my partner should communicate in a way that makes sense to me, at least to the extent that I can ask follow-up questions. From a good communication partner, I also expect explanations so I can understand and verify what the machine has done. It's less about looking into the brain of the other to understand the exact brain operations that led her to say what she said. Viewed through this lens, so-called "hallucinations" in large language models are not errors of an intelligent system but a helpful means to fulfil the purpose of communication: The algorithm participates in communication and tries to maintain it. In fact, upon inquiry, the misinformation usually comes to light, since the system is not intelligent and either conceals its error or even understands that it has made a mistake.

When discussing *explainable AI*, we should also consider that attempting to understand the internal operations might lead to a dead end. As it seems to me, only experts can make sense of such explanations. From a users perspective, communication could be the key, meaning that systems must be designed such that users can *make sense* from communicating with these systems. Systems should be able to explain themselves through communication.

Of course, the constructed sense could be harmful or dangerous. Furthermore, the objective window to the world certainly remains an impossible ideal. From Luhmann's constructivist perspective, there is no objective standpoint, as observation must always be selective, thus always creating a *blind spot*[9]. This certainly does not imply that there are no better or worse stories or that we are incapable of tolerating multiple stories. To accommodate this diversity, our new artificial communication partners should communicate different biases—different perspectives with different blind spots.

From the lack of intelligence in AI algorithms, we should not conclude that their use is unreasonable or ineffective; on the contrary. Their success is based on the fact that they do not think like humans. Therefore, I do not believe we need to fear superintelligence. Rather, there are concrete challenges that research and policy makers should focus on. Problems rather appear because these algorithms are not intelligent.

Artificial communication sounds less concerning than superintelligence until we realize that society is fundamentally about communication.

> "Precisely because generative AI intervenes in communication, we must be very concerned about its effects." – (Esposito, 2024, S. 49)

As a sociologist, Esposito particularly addresses *disinformation* and the spread of *fake news*. She makes clear that mass media have never been and will not be an objective window to the world. Instead, mass media construct their own specific world that becomes the *reference world of the public*. This world is not an arbitrary construction and keeps society restless. The danger is not only that false information is spread but that the reference world crumbles. In this case, no one can plausibly consider themselves informed or part of a community of people who refer to the same news (Esposito, 2024, S. 66), i. e. to the same "reality". On one hand, an open society must tolerate multiple worlds; on the other hand, a society in which all individuals live in their own world would arguably no longer be a society.

Esposito ends her book with a sober forecast: What we recognize as novelty in the present is, in a way, already old. Radical changes are unpredictable. She advises us to look back in history, to times when communication has changed—such as the development of language, writing, and the printing press. She suspects that the role and perspectives of humans will remain indispensable but will lose priority since it may no longer be necessary to consider them in communication (Esposito, 2024, S. 79).

> "The contribution of humans remains indispensable: The material from which data are derived will and must continue to be produced by humans." – (Esposito, 2024, S. 79)

[9]A metaphorical concept. Whenever we look at something, we must exclude everything else. We always need a perspective, and this perspective can in turn be observed (but not by us at the same time). Every perception is based on conditions that escape this very perception. The world as a whole can never be perceived.

REFERENCES

Bender, E. M., & Koller, A. (2020, July). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 5185–5198). Online: Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.463

Browning, J., & LeCun, Y. (2022). *AI and the limits of language.* https://www.noemamag.com/ai-and-the-limits-of-language/. (Accessed: 2024-02-16)

Brühl, J. (2023). *Wie ein Münchner KI-Professor gegen den Größenwahn der Branche kämpft.* https://www.sueddeutsche.de/wirtschaft/kuenstliche-intelligenz-ki-neurips-konferenz-bjoern-ommer-1.6318009. (Accessed: 2024-02-16)

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... Zhang, Y. (2023). *Sparks of artificial general intelligence: Early experiments with GPT-4.*

Dialogforum Generative KI und Hochschule. (2023). https://www.hm.edu/lehren/generative_ki/dialogforum.de.html. (Accessed: 2024-02-16)

Dreyfus, H. L., & Dreyfus, S. E. (1986). From Socrates to expert systems: The limits of calculative rationality. In C. Mitcham & A. Huning (Eds.), *Philosophy and technology ii: Information technology and computers in theory and practice* (pp. 111–130). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-009-4512-8_9

Epstein, Z., Hertzmann, A., Herman, L. M., Mahari, R., Frank, M. R., Groh, M., ... Russakovsky, O. (2023). Art and the science of generative AI. *Science*, *380*, 1110 - 1111. Retrieved from https://api.semanticscholar.org/CorpusID:259095707

Esposito, E. (2022). *Artificial communication*. The MIT Press. doi: 10.7551/mitpress/14189.001.0001

Esposito, E. (2024). *Kommunikation mit unverständlichen Maschinen*. Residenz Verlag.

Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). *ChatGPT is not all you need. A state of the art review of large generative ai models.* arXiv. doi: 10.48550/ARXIV.2301.04655

Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin.

Luhmann, N. (1987). *Soziale Systeme: Grundriß einer allgemeinen Theorie*. Suhrkamp.

Marcus, G. (2022). *Deep learning alone isn't getting us to human-like AI.* https://www.noemamag.com/deep-learning-alone-isnt-getting-us-to-human-like-ai/, note = Accessed: 2024-02-16.

Pasqualino, R., Jones, A., Monasterolo, I., & Phillips, A. (2015, 7 23). Understanding global systems today – a calibration of the world3-03 model between 1995 and 2012. *Sustainability*, *7*(8), 9864-9889. doi: 10.3390/su7089864

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners.* Retrieved from https://api.semanticscholar.org/CorpusID:160025533

Rosengrün, S. (2021). *Handbuch der Künstlichen Intelligenz* (G. Görz, U. Schmid, & T. Braun, Eds.). Berlin, Boston: De Gruyter Oldenbourg. doi: doi:10.1515/9783110659948

Skene, K. R. (2020, oct). No goal is an island: The implications of systems theory for the sustainable development goals. *Environment, Development and Sustainability*, *23*(7), 9993–10012. doi: 10.1007/s10668-020-01043-y

XR für den Mittelstand. (2023). https://www.ihk-muenchen.de/xr-mittelstand/. (Accessed: 2024-02-16)