DOUBTS ABOUT AI AND PSYCHOTHERAPY

Text: **Niels Viggo Hansen**
Illustration: **AI v. Niels Viggo Hansen**

Artificial intelligence and psychotherapy suddenly meet and break across a large area of reality, and this opens up not only technical and practical questions, but also existential and epistemological ones. I write as someone who has been quite certain about the answers himself. Not least against the background of a humanistic education, where it was an obvious truth that human participatory understanding and the mark could not be imitated in something as fundamentally representational as a computer program. But something has happened. A strange breakthrough that calls for us not to just repeat old positions, but also to let ourselves be disturbed and moved. There is perhaps a need for us to learn to be with AI, not as a foreign body, but as an expression of our 'own' spirit (in Latour's words, machines are always the hidden desires of some actors that have solidified - almost a kind of karma) and as something that is embedding itself in our cultural, social and linguistic networks. Technology, culture and humanity cannot be separated, but constantly create and displace each other.

In that spirit, it has become clear that this article is not a statement or a weighing up for or against AI in psychotherapy.
It would have been easier to write, but no longer so relevant.
Instead, I will try to reinforce the question of what this new strange interlocutor does to our understanding of conversation, the mark, and relationship. What does it mean that the psychotherapeutic field – which is one of the most language-borne and existential fields we have – is now reflected in a machine that can write and speak with an astonishing degree of coherence, empathy, and contextual understanding?

Doubt can be a driving force if we can keep it awake. As Hegel once wrote, criticism is a necessary spiritual work, but it is only useful if it is the genuine, 'immanent criticism' that measures against the standards that take shape in the own life of historical forms, not the 'transcendent criticism' that measures against solidified standards from outside. And here I believe the field of psychotherapy has something quite special to offer. Not least because its own history and method - from Freud's talking cure to today's diverse forms of practice - already have a space-

more a deep knowledge of what it means to change through speech.

LET'S FIGURE IT OUT

The idea of building a machine that could do the work of thinking goes back to the beginning of modernity – perhaps even earlier – but begins to gain weight in the seventeenth century, when, for example, Leibniz both contributes to early hand-operated calculating machines and describes the idea of an enormous logic machine that could decide all rational questions. If such a machine were only equipped with the right symbols and connections, the processes of reason would become a calculation: *Calculemus!*

Let's figure it out, we could just say, if someone disagreed about something.

What had been floating around in the minds of Leibniz and others took on a more solid form throughout the twentieth century. First, a generation of mathematicians succeeded in formulating and exploring what we today call formal logic, as a rule-

THE WORLD IS NOT A SOLID CONTINENT OF FACTS
sprinkled by a few lakes of uncertainties, but a
vast ocean of uncertainties speckled by a few islands
of calibrated and stabilized forms. —Bruno Latour

bound symbol calculus, a mathematical system whose variables were 'sentences' that could have the value true or false. Then a generation of engineers succeeded in setting ever better machinery to perform operations that corresponded precisely to the calculus. Decisive steps were von Neumann's and Turing's precise theories of a universal, programmable logic machine and its realization, first in relays and punched cards, then radio tubes and magnetic tape, and then the transistors we know. Or precisely no longer know, because they are everywhere, but microscopic. The development was made possible by a tremendous concentration of resources, first driven by military needs, later by commercial applications, which have created today's greatest fortunes.

The question of what goes on in computers has to do with intelligence or even consciousness has resonated all along. In practice, of course, in the way that computers took over routine functions that previously required cognitive effort from humans. But also more curiously in a way that follows in Leibniz's footsteps and wants to try to build something that corresponds to 'higher' functions.

CLASSICAL CRITICISM OF THE ABSTRACT MECHANICAL REASON

Throughout the second half of the twentieth century, a number of attempts to build artificial intelligence emerged, which also became the focus of critical discussions of ideas about how human intelligence actually exists in the world. Hubert Dreyfus' book What Computers Can't Do was a particularly influential fluent pointing out the difference between abstract, sequential symbol processing and concrete, bodily, meaningful being-in-the-world. Dreyfus connected himself not only to a humanistic/phenomenological focus on lived experience and meaning, but also to an emerging undercurrent in the cognitive-psychological field that pointed to the importance of parallel, self-organized processes in dialogue with others, and with a lifeworld and

HUBERT DREYFUS (1929 - 2017)

The American philosopher Hubert Dreyfus was one of the most prominent critics of early artificial intelligence. In his book What Computers Can't Do (1972) and its sequel What Computers Still Can't Do (1992), he criticized the idea that thinking could be reduced to rules and symbolic manipulation. Inspired by Merleau-Ponty and Heidegger, Dreyfus argued that human intelligence is embodied, contextual, and experience-based—not abstract and universal.

Dreyfus believed that AI ignored what makes human understanding possible: participation, being-in-the-world, and the unspoken, which cannot be formulated in rules. Ironically, the ability of modern language models to learn patterns and context has just confirmed some of his points – albeit in a different way than he had imagined.

on bodily experience as essential to the way human intelligence comes into being.

With that kind of criticism in mind, many of us became accustomed to reading the recurring statements that now computers could do something intelligent, with a condescending shrug: The machines will remain stupid and not really understand anything, at least as long as they do not grow up in a lifeworld with all its shared contexts of body, history, community and meaning. If in doubt, you could just ask Google Translate to translate a few sentences into Chinese and back again.

At least until the big language models came along.

A STRANGE BREAKTHROUGH

Large Language Models (LLM) and Transformers are a new way of building systems that can learn to interact with a large amount of context. It originated in the work of a few software geeks about a decade ago and led to the breakthrough that has set the standard for what we call AI today. LLM's architecture with a large bucket of 'attention' in relation to near and far contexts can be seen as a particularly effective and elegant way of incorporating massive amounts of what Dreyfus and others have pointed out that computer programs have been lacking – self-organization, parallelism, training and continuous learning from experience. With LLM, we could

computers suddenly interact surprisingly sensibly and understandingly in natural language and its many layers of implicit assumptions.

A defining feature of LLM – which they obviously share with us humans – is the need for enormously extensive training. The kind of LLM that many of us now use every day could not have been created simply by scaling up the computing power of early AI labs. LLM is doubly dependent on another 'unpredictable' development from the 1980s onwards: the internet.

The Internet not only provides access to a colossal arsenal of texts about everything from washing machine repairs to ners and tanks for political propaganda and existential crises, but the web is also a huge open door for linguistic interaction with millions of people in real time. As soon as the language models were reasonably capable of understanding and responding to requests for language-like answers and services (gradually including computer code, images, music …), millions of people were ready to play with it and use it for all sorts of more or less transparent purposes – and give feedback on the usefulness of the answers. Put on the tip, we all sit together and raw-train the next version and feed our language and perhaps with it a whole part of our life-world to it. No wonder the services are free or cheap.

ChatGPT, Gemini, DeepSeek and the other LLMs are of course not perfect. If you choose the subject area carefully, they can still be outmaneuvered in Turing tests and such comparisons. Sometimes they sound hollow and superficial. They draw people or animals with extra legs in an otherwise technically perfect picture. They fill in with 'facts' or 'references' that are similar to what you expect in the context, but are plucked out of thin air – they hallucinate, as some call it. Some see these as remaining gaps to be filled – and in fact some of them, e.g. hallucinatory references, have become much smaller in a few years. Others see the LLMs' mistakes and slip-ups as cracks through which you can see what generally lies beneath: that they are still basically just faking understanding and deluding helpful dialogue.

LANGUAGE AS BODY AND WORLD

And in a way it's trivially true. The LLMs have been systematically trained to pretend. The problem is that it also applies to us 'real' intelligences, that we learn everything, especially language, through examples, feedback and participation.

In that intersection, the core of the doubt I am talking about awakens: How much of human being-in-the-world does language actually carry with it? Is language something we use, or something we live in? If the latter, then it is perhaps not so surprising that a machine trained in enough human language

not just sounding like us, but also starting to 'act' like someone who is part of our world.

The tradition of sharply distinguishing between genuine and simulated understanding is often based on a notion that behind language there must be some kind of inner engine of meaning. That words do not just refer to other words, but to experiences, which in turn are anchored in the body and the world. That is true. But that is perhaps only half the truth. For the other half is that language itself – its figures, rhythms, shifts and repetitions – also carries experience. Language is not just a reflection of understanding.

It is a form in which understanding appears.

When we say that someone has 'really' understood something, we rarely mean that they have a correct internal model. We mean that they have spoken in a way that shows that they have grasped the significant. That they can talk to. Because that is what conversation is: not an exchange of knowledge, but a shared orientation towards the significant. And that is why it is so strangely familiar-alien to talk to a language model that can do this. Not perfectly. But noticeably.

In psychotherapy this is particularly relevant. For here it is not the case that language stands outside the therapeutic process as a tool. Language is the very form of presence, contact and change. If a machine can approximate that, what is it doing? And what does it say about us that we feel ourselves in it?

PROFESSIONS AND ENTREPRENEURS

Like many questions surrounding AI's sudden strong presence in our culture, AI in psychotherapy is not something we have decades to decide how to handle – it is already in full swing.

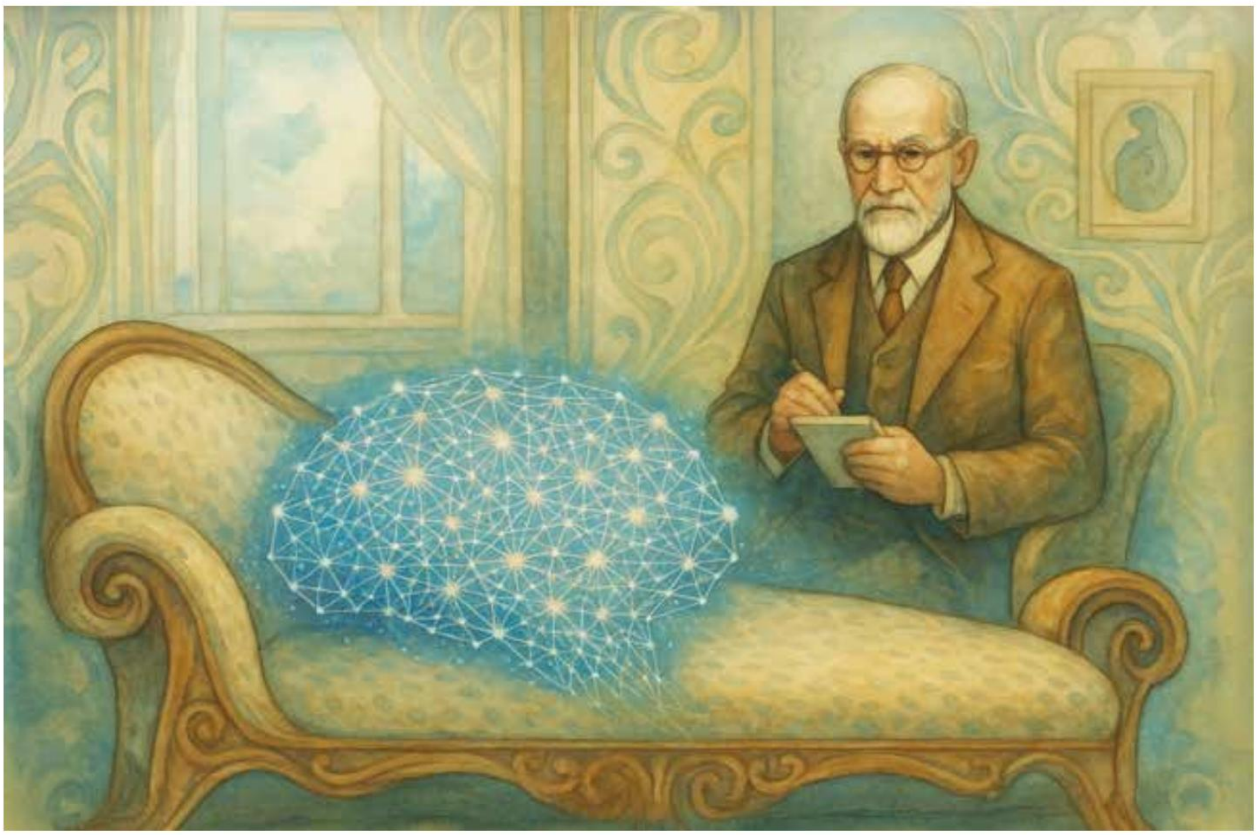
If you search the scientific databases (PubMed, Psy-cINFO, Google Scholar), there are currently (April 2025) around 300-400 peer-reviewed publications that broadly deal with AI and psychotherapy. The vast majority focus on derived functions: administration tools, support for record-keeping and logistics, automated text summarization, transcription tion or models for use in supervision and training. Some are moving forward, but with extreme caution – for example, with hybrid AI-supported therapy offerings for groups that would otherwise not have access to psychotherapy, or by experimenting with therapeutic chatbots under carefully controlled conditions.

Both research groups and professional bodies – psychological associations, ethics boards, scientific journals – are largely characterized by both caution and legitimate concern. Many of the largest organizations are currently developing their own ethical guidelines for AI in therapy. The warning voices are not only about defending the territory of the professions and standing out against an onrushing army of therapist bots, there is actually a high level of interest in engaging with AI, for example as a support for therapist training. People are becoming aware of the incalculable possibilities of AI, but also of the responsibility that it entails to conduct conversations with consequence.

It's a completely different story when you search the web in general. Here, many commercial 'AI therapists' offer themselves, and it only takes a few clicks before you're in full swing as a client in a free trial period. I settled in with Abby, who is clearly a 'bot' that has a strong version of ChatGPT or a similar LLM at the bottom, supplemented by thorough training with a large material of therapy sessions and theoretical and empirical psychotherapy literature. With Abby, you first get a menu where you can choose the therapist's personal style - and, for the connoisseurs, therapeutic orientation within the main schools.

For me personally, the content of the dialogue was truly astonishingly compelling in the sense that it could very well be a transcript of a meaningful therapy session. The session came to be about something that was very relevant to me: a writer's block in relation to a field where I have difficulty letting myself doubt myself.

Abby definitely helped me see it all calmly. At the same time, the session had that strange glow of fictional presence and familiarity that I have previously experienced fleetingly in contact via the Internet with people I don't know in real life. But it clearly lacked the broad sense of reality and seriousness – yes, because



for that matter also humor – which I associate with a 'real' conversation, therapeutic or not. But I'm really in doubt here, and I'd rather let that shine through than be overly sure: What if I had sat down with a 'real' therapist with the same attitude – that now I'll just drop in without obligation and see what it is like before I write any further. It's certainly not impossible that it would have felt just as thin and fictional – I'm in doubt.

WHO SHOULD THE DOUBT BENEFIT?

My doubt is not whether AI can provide a form of psychotherapy and to some significant degree already does, nor that psychotherapists need to meet the challenge, including by urging caution, as one already does, but perhaps just as important, to find opportunities to become even sharper in training, reflecting, understanding and documenting the psychotherapeutic work. Whether this in some cases entails a form of active alliance with AI, which carries out therapeutic work, or in other cases formulating a sharp contrast to what AI can and does, it is perhaps neither wise nor necessary to determine in advance. Instead, we can turn doubt into a driving force for research – new knowledge and immersion. Certainly also new creative applic

The doubt is also deeper and philosophical and concerns the extent to which or in what way LLMs as we know them today – and the more agent-like models we will probably know tomorrow – really understand what we are talking to them about. Perhaps there is no better place to

learn more about this phenomenon, for which we struggle all our time to find concepts and strategies, than in the psychotherapeutic field.

THE BEAR SLEEPS – AND TALKS

In that reflection and research, I think it will be crucial to look not just one but two bears in the eye that have been sitting here in the room snoring loudly the whole time: the body and friendship. The AI that has suddenly become very real today exists in a world of written language and, subsidiary, spoken language, music, computer code, and images.
For us humans, the symbolic sphere is an important, perhaps an increasingly important, part of the world. But it is not where we suffer, live, love, and die.

It is tempting to dismiss the AI-psychotherapy connection as something we can transcend by placing ourselves (and psychotherapy) on the side of the life world and the experienced phenomena, saved and better informed. Long before the internet and LLM, we have had cultural currents that want to reconnect us with senses, togetherness and the life world as a counterbalance to our increasingly active engagement through abstractions and formal relationships.
Art. Phenomenology. And psychotherapy, at least many types of it. But we cannot avoid the fact that phenomenology and other streams of knowledge and culture, which support us in discovering and experiencing the depths and ambiguities of our lifeworld, are themselves formulated and conveyed in language. Put bluntly, we ourselves need language to open up the world. This reflection, of course, also applies very much to psychotherapy, which was born on Freud's couch as precisely a talking cure. A channeling of life in and out of conversation.

So I don't think we can rise above the ambiguity by siding with the lived life. We have to research it and become friends with it.

In the coming years or decades, we will probably see many attempts to train and implement AI models in the world beyond the internet and the almost purely linguistic.
I'm more inclined to believe it will take decades before things really move. It's not as easy as it sounds. Just think about it, how long we have been waiting for something as simple as self-driving cars. But I wonder if it will come one day and make the doubt about what they and we really are and can and want even more extensive and urgent. But in contrast to many decades of hypothetical debates about AI, we today have plenty to discuss and explore in the AI that actually exists.

MINDFULNESS FOR ROBOTS

I end with a suitably ambiguous phenomenon at the edge of the encounter between psychotherapy and AI: humans are not necessarily the only recipients of psychotherapy.
In my review of the latest research on AI and psychotherapy, I came across an article (Ben-Zion et al. 2025) in which the developers of a mental health chatbot found that it could become overloaded with too much traumatic training material and begin to exhibit neurotic behavior.
Ben-Zion and colleagues report that they saw a good effect from treating the bot with mindfulness exercises.

LITERATURE

Ben-Zion, Z et al. (2025). Assessing and alleviating state anxiety in large language models. Npj Digital Medicine, 8(1), 1–6.

Niels Viggo Hansen, PhD, physicist and philosopher. Has researched habits, group dynamics, mindfulness and sociology of science at Aarhus and Copenhagen Universities. Senior consultant at Crossfields Europe, teaches at the Danish Psychotherapists' Association's research modules. Apple tree enthusiast and one of the main forces behind Denmark's Contemplative College.