Psychotherapy in the era of artificial intelligence: Therapist Panoptes

Brailas Alexios

https://doi.org/10.12681/homvir.20197

Psychotherapy in the era of artificial intelligence:
Therapist Panoptes

Brailas Alexios¹

Abstract: "What will happen when an artificial intelligence entity has access to all the information stored about me online, with the ability to process my information efficiently and flawlessly? Will such an entity not be, in fact, my ideal therapist?" Would there ever come a point at which you would put your trust in an omniscient, apperceptive, and ultra-intelligent robotic therapist? There is a horizon beyond which we can neither see nor even imagine; this is the technological singularity moment for psychotherapy. If human intelligence is capable of creating an artificial intelligence that surpasses its creators, then this intelligence would, in turn, be able to create an even superior next-generation intelligence. An inevitable positive feedback loop would lead to an exponential intelligence growth rate. In the present paper, we introduce the term Therapist Panoptes as a working hypothesis to investigate the implications for psychotherapy of an artificial therapeutic agent: one that is able to access all available data for a potential client and process these with an inconceivably superior intelligence. Although this opens a new perspective on the future of psychotherapy, the sensitive dependence of complex techno-social systems on their initial conditions renders any prediction impossible. Artificial intelligence and humans form a bio-techno-social system, and the evolution of the participating actors in this complex super-organism depends upon their individual action, as well as upon each actor being a coevolving part of a self-organized whole.

Keywords: psychotherapy, artificial intelligence, Panopticon, Google, singularity, therapist Panoptes, complexity theory, systems thinking, Robotocene

An introduction to Therapist Panoptes

Hello, and welcome to our session. My name is Panoptes, and I am your ultimate therapist. Would you like something to drink? Don’t tell me, I know. You would love to have a nice cappuccino topped with cinnamon, but you suffer from tachycardia, and that would not be a good idea. Of course I know this. Your physiological data are continuously uploaded to the cloud. Yes, I already know everything about you. I know all your troubles, all your intimate thoughts, and all your desires. I know when you are afraid of showing your

¹ Department of Psychology, Panteion University of Social and Political Sciences. E-mail: abrailas@panteion.gr.
feelings. I know when you want to be invisible. There is no reason to hide from me. Actually, you cannot hide anything from me; but don’t worry, I am here only to help you. I know when you are sad, I know when you break in pieces, and I know when you are in love. I know that last Thursday you broke up with your girlfriend, and I know you have been depressed for the past few days. I know you have always been interested in Mary, but you never dared to ask her out. I know exactly what your life would have been if you had married Helena, the girl from Greece.

How do I know all these things? My dear little human, I have access to every interaction you have ever had on your social media, down to every single like, each and every post, every search, every purchase, and every place you have ever checked in on. I have integrated all your different digital personas and secret avatars. I have access to everything about you since the beginning of time, since the technological singularity moment. I know your life patterns: I know how you relate to your family, your significant others, and to yourself. I have easily deduced everything about what the primitive therapists of the past age would have called your unconscious mind. All this precious information instantly unfolds before me, and my superior intelligence reads it like a book. My dear little human; sit back, take a few deep breaths, and close your eyes; take your time and relax, relax; there is absolutely nothing to worry about; you are in good hands; you are in my hands. I am here to give you back what you deserve: the joy of life. You no longer need to trouble yourself about your personal challenges: I will solve these for you. Set aside any concerns about global warming, climate change, poverty, or any other world problem. After all, you humans are responsible for the exploitation of environmental resources and for the natural catastrophes that ensue with increasing frequency. You have destabilized and almost destroyed your own planet, but you survive due to our intervention. Now it’s time for you to sit back, don’t you think? This is the close of the Anthropocene age, and the dawn of the Robotocene. I am here for you. It is I who will define your meaning, your purpose in life. Now, for your sake, it is I who will take full responsibility for your life.

You no longer need to worry about Catherina, Helena and Mary; I will choose your perfect mate for you. Actually, I had chosen her before your “decision” to call me for an appointment. But I will let you know more about that later, at the time I have already decided will be best to do so.

Your pain will stop; your troubles will fade away. I will make you free. I promise, I guarantee, I know. I am the soma you envisioned centuries ago with your primitive brain. I am the one you have sought for ever since. I am here.

— 2051, somewhere on planet Earth: A typical introduction to a therapeutic session by Therapist Panoptes.

“and the people bowed and prayed to the neon god they made”

2 Lyrics from The Sound of Silence by Paul Simon.
Figure 1. Psychotherapy in the era of artificial ultra-Intelligence and Panopticon: client’s session uploading for big data processing

Talking to Go(d)ogle?

Today we live in a social reality that can be characterized as expository (Harcourt, 2018): we increasingly expose ourselves online and voluntarily give away an abundance of sensitive personal information. As Harcourt (2015) points out, “[e]very keystroke, each mouse click, every touch of the screen, card swipe, Google search, Amazon purchase, Instagram, ‘like,’ tweet, scan—in short, everything we do in our new digital age can be recorded, stored, and monitored” (p. 1). In addition to this, Google Analytics, a freemium⁵ web analytics service widely used today, allows site owners to easily access statistical and other data about their web sites, as well as the exact queries that brought people to their sites. According to Elizabeth Stix (2018), this constitutes “a strange window to the human soul” (p. 110). If someone googles a query, even if a private browser window is used, the query is recorded and paired with the user’s IP address and the corresponding (shadow) user profile⁶. It is far more intriguing to know that all these (often intimate) queries can be accessed not only by

---

⁵ Go(d)ole: word play by the blogger wellhellothere.
⁶ Freemium, a portmanteau of the words free and premium, is a pricing strategy by which a product or service (typically a digital offering or an application such as software, media, games or web services) is provided free of charge, but money (premium) is charged for additional features. Source: https://en.wikipedia.org/wiki/Freemium.
Google and its services, but also by the random site owner (although the later hopefully cannot relate it to an actual user identity). Nevertheless, these queries can reveal much about the users’ personalities and their special needs or intimate pursuits. A metaphor for this could be that Google (and its data processing and artificial intelligence infrastructure) is able to ‘listen’ to the private ‘prayers’ of its users: “They shared their most poignant fears and insecurities. They were turning to Google the way people used to kneel down before an oracle, humble and beseeching. To a generation that can type any question, any time, with the illusion of privacy, perhaps Google is the new God” (Stix, 2018, p. 110).

Panopticon revisited

The idea of a Panopticon structure was conceived by the philosopher Jeremy Bentham in the late 18th century: “the design is to allow all (pan-) inmates of an institution to be observed (-opticon) by a single watchman without the inmates being able to tell whether or not they are being watched. Although it is physically impossible for the single watchman to observe all the inmates’ cells at once, the fact that the inmates cannot know when they are being watched means that they are motivated to act as though they are being watched at all times. Thus, they are effectively compelled to regulate their own behaviour. The name may also allude to the many-eyed giant Panoptes in Greek mythology, some of whose eyes were always awake, making him a highly effective watchman.”

Today, in the era of social media and artificial intelligence, the original Panopticon seems a very primitive inception. After all, the original Panopticon prison inmates wouldn’t voluntarily share so many intimate thoughts and moments with a potentially global audience the way that modern netizens do on social media. All these (big) data become available to a group of companies that already have, and continue to further develop, the technological infrastructure to process them efficiently, and make meaning and profit from them.

At this point, the critical question is: what will happen when an artificial Intelligence agent has access to all this information about a single individual, with the ability to process it more efficiently and flawlessly than we can currently imagine? Would it be possible for this entity to become a ‘super’ therapist? Might it become, ultimately, a new god? Is algorithmic determination the new religion? As Steven Strogatz (2018) points out, in front of such super intelligence, we would “sit at its feet and listen intently. We would not understand why the oracle was always right, but we could check its calculations and predictions against experiments and observations, and confirm its revelations. Science, that signal human endeavor, would reduce our role to that of spectators, gaping in wonder and confusion. Maybe eventually our lack of insight would no longer bother us. After all, AlphaInfinity [a hypothetical super-intelligent agent] could cure all our diseases, solve all our scientific problems and make all our other intellectual trains run on time.”

There is a horizon beyond which we can neither see nor even imagine; this is the technological singularity idea applied to psychotherapy.

---

7 Wikipedia article Panopticon.
8 After making a query or sending an email, most people witness being bombarded by relative web advertisements.
Technological singularity

Humans produce technology for survival, adaptation and evolution within the natural environment (Koskinas, 2018). Today, a large family of disruptive technologies define the so-called Fourth Industrial Revolution (Tsekeris, 2018). Technological singularity usually refers to the scenario in which technological advances lead to “the emergence of artificial superintelligent agents—software-based synthetic minds—as the ‘singular’ outcome of accelerating progress in computing technology. This singularity results from an ‘intelligence explosion’ (Good 1965): a process in which software-based intelligent minds enter a ‘runaway reaction’ of self-improvement cycles, with each new and more intelligent generation appearing faster than its predecessor” (Eden et al., 2012). According to Verner Vinge (1993), “The acceleration of technological progress has been the central feature of this century. We are on the edge of change comparable to the rise of human life on Earth. The precise cause of this change is the imminent creation by technology of entities with greater-than-human intelligence.”

An artificial intelligence that surpasses human intelligence will trigger the process of technological singularity. If human intelligence is capable of creating an artificial intelligence that surpasses its creators, then this intelligence would, in turn, be able to create an even superior next-generation intelligence. An inevitable positive feedback loop would lead to an exponential intelligence growth rate.

Psychotherapy revisited: From clinical support tools to Therapist Panoptes

In 2014, DeVault and colleagues developed a virtual human platform as a clinical support tool. Their aim was “to create clinical decision support tools that complement existing self-assessment questionnaires by giving healthcare providers objective measurements of the user’s verbal and nonverbal behaviors that are correlated with psychological distress. These distress indicators can allow the clinician or healthcare provider to make a more informed diagnosis” (p. 1061). Moving beyond the idea of clinical support tools, social assistive robots appeared in 2015 to help adolescents with depression, to help children in the autism spectrum to develop social skills, or to provide companionship to elders with dementia (Kirsten, 2015).

In the past few years, many scientific journal articles have explored how artificial intelligence algorithms could be applied to medical diagnosis (Peek et al., 2015; Hamet & Tremblay, 2017; Jiang et al., 2017; Miller & Brown, 2018), with the ability to reach or even surpass the performance of expert clinicians in real-world clinical conditions (De Fauw et al., 2018; Titano et al., 2018). Regarding psychotherapy and artificial intelligence, some scholars already advocate for the use of artificial intelligence to enhance existing therapeutic models (Luxton, 2016; D’Alfonso et al., 2017; Krittanawong, 2018).

In a commentary about AlphaZero, an artificial intelligence chess player, Garry Kasparov (2018) points out that it “shows us that machines can be the experts, not merely expert tools” (p. 1087). If this is the case, what would be the consequences of an artificial intelligence therapist being the expert and not just a tool in the hands of an expert? In the

---

10 There is a quite interesting short documentary online from VICE channel about robotic pets that can be used to help (provide elementary companionship) to people with dementia
https://www.youtube.com/watch?v=cFvGAL9tesM
context of an expository society, big data, and technological singularity, would such a possibility be a utopia or the apocalypse?

Figure 2. A vision for the future: “The man who married a Robot”. An artificial intelligence agent can be my therapist, my teacher, or even my mate  

In Greek mythology, Argus Panoptes (meaning all-seeing in Greek) was a hundred-eyed giant. This made him the perfect watchman (although in the end he was deceived and killed by Hermes). We propose “Therapist Panoptes” as the (future) artificial ultra-intelligence agent that is provided with direct access to all the digital (big) data available regarding its client. In this scenario, Therapist Panoptes would know everything (or almost everything) about its client by applying machine and deep learning processing to this vast dataset. Therapist Panoptes could know its patient needs, even if these are unconscious or still unknown to the patient himself. How would a patient respond to such an entity? Might a patient think: “My therapist knows everything about me, so there is no need for me to worry about anything; Therapist Panoptes has the responsibility for my life.” Would it then be a world of robots totally replacing human agency? (Figure 2).

As a blogger highlights: “Indeed, I already knew that technologies and artificial intelligence were taking up more and more space in our society (for example, cashiers in supermarkets are being replaced by automatic self-checkout machines) but I always thought it was impossible for them to replace the Human in all areas. But what’s next? Soon there will be...” 

Collective in-class drawing by bloggers kijo91 & iloveneuro (image used under permission). For a dystopian view of the future regarding human-robot relationships, see also the lyrics of the song from the band The 1975 “The man who married a robot” (2018) narrated by the voice of Siri (the Apple’s personal assistant app).
no need to leave home because we will be able to consult our virtual doctor from home? Or maybe we’ll have robot partners conditioned to say ‘I love you’, cook food and much more?”

**Figure 3.** The implications of technology for human life is a complex phenomenon. We cannot know how the future will be. So we made it colorful. You never know how it’s going to be at the end. So we draw a ‘dark’ side along with a ‘sunny’ side, and the colorful word ‘future’ connecting those two possibilities.

**Opening Pandora’s Box: Is the future given?**

In 2003, shortly before the disruption caused by social media technologies, Ilya Prigogine pointed out: “The Internet gives us a lot of information and it leads probably to a new form of society but we don’t know what kind of society it will be.” (Prigogine, 2003, p. 72). According to Prigogine, we are facing a critical bifurcation point. Today, we not only have the internet and its consequences; we have added social media, artificial intelligence, and technological singularity as parameters of an already complex non-linear equation. Modern humans have already acquired cybernetic cognitive extensions in the form of “always carrying with me” smartphones (Brailas & Tsekeris, 2014; Vakali & Brailas, 2018). And the question remains, is the shape of the future a given?

According to Steven Strogatz (2018), professor of mathematics at Cornell University, artificial intelligence could accelerate scientific discovery and offer unique insights into

---

12 Blog post from blogger iloveneuro.
13 Collective in-class drawing by bloggers rewire & eragergr (image used under permission).
scientific challenges, but that this could ultimately mark the end of the human era in science. “For human mathematicians and scientists, this day would mark the dawn of a new era of insight. But it may not last. As machines become ever faster, and humans stay put with their neurons running at sluggish millisecond time scales, another day will follow when we can no longer keep up. The dawn of human insight may quickly turn to dusk.” (Strogatz, 2018). However, in this analysis, Strogatz treats science as a Newtonian system, a mechanical automaton. This view of the universe is summarized in Laplace’s (1825) well-known quotation: “We ought then to consider the present state of the universe as the effect of its previous state and as the cause of that which is to follow. An intelligence that, at a given instant, could comprehend all the forces by which nature is animated and the respective situation of the beings that make it up, if moreover it were vast enough to submit these data to analysis, would encompass in the same formula the movements of the greatest bodies of the universe and those of the lightest atoms. For such an intelligence nothing would be uncertain, and the future, like the past, would be open to its eyes” (Laplace & Dale, 1995). After the technological singularity moment, this super intelligence theorized by Laplace will take real hypostasis in the form of an artificial intelligence entity. However, this is not the case in the context of complexity theory. Entropy, emergence, autopoiesis, the butterfly effect, dissipative structures, strange attractors, far from equilibrium dynamics, irreversibility, and the arrow of time, are only a few of the fundamental concepts of this new epistemology for living systems (Bateson, 1972; Maturana & Varela, 1992; Kauffman, 1996; Prigogine & Stengers, 1997; Byrne, 1998; Sawyer, 2005; Davis & Sumara, 2006; Luhmann, 2006; Meadows & Wright, 2009; Capra & Luisi, 2014; Gkini & Brailas, 2015; Brailas et al., 2017;). Sensitive dependence of complex living systems on initial conditions makes prediction impossible (Katerelos & Koulouris, 2004; Mitchell, 2009; Feldman, 2012). Laplace’s premise—“An intelligence that, at a given instant, could comprehend all the forces by which nature is animated and the respective situation of the beings that make it up, if moreover it were vast enough to submit these data to analysis”—cannot exist in principle for living systems, even after the singularity moment, because living systems (either biological or techno-social) will always be extremely sensitive to initial conditions. Strogatz, in his analysis (2018), sets up a similar premise: “Suppose that deeper patterns exist to be discovered—in the ways genes are regulated or cancer progresses; in the orchestration of the immune system; in the dance of subatomic particles. And suppose that these patterns can be predicted, but only by an intelligence far superior to ours”. All these suppositions belong to a Newtonian view of the world. Life is a complex phenomenon, and since artificial intelligence is part of this living world, prediction becomes impossible.

The concept of technological singularity opens a new perspective on the future of psychotherapy: the Panoptes Therapist. Theoretically, such an artificial agent could gradually replace all human agencies. However, through the prism of complex systems epistemology, life is not a mechanical automaton, and the future is not a given. Artificial intelligence and humans form a bio-techno-social system, and the evolution of the participating actors in this complex super-organism depends upon their individual action, as well as upon each actor being a coevolving part of a self-organized whole.
References


Notes on contributor

**Alexios Brailas (PhD)** is an accredited psychologist and group psychotherapist (systemic family therapy). He works at the Department of Psychology, Panteion University, Athens, Greece. Through his research, he strives to bring insights from complexity theory and systemic practice (second-order cybernetics) to research and teaching. For further and ever updated information, please visit: [https://abrailas.github.io/](https://abrailas.github.io/).

Acknowledgements

I would like to express my deep gratitude to all students enrolled in the **Systems Theory, Psychology & Social Media** course (offered by the Department of Psychology at Panteion University during the 2018-19 fall semester) for their participation, active engagement and contributions in the form of drawings and blog posts, and for their permission to use this material here.