

## Chatbots, Robots, and the Ethics of Automating Psychotherapy

By Eric B. Litwack\*

*Recent developments in artificial intelligence—AI—have caused considerable discussion among both philosophers of technology and psychotherapists. In particular, the question of whether or not new forms of AI will complement or even replace traditional psychotherapists has emerged as a major contemporary debate. This debate is not entirely new, as it has its origins in the Turing Test of 1950, and an early psychotherapy chatbot named Eliza, developed in 1966 at MIT. However, recent developments in AI technology, coupled with long waiting lists and variable access to psychotherapists have raised the question of machine psychotherapists in an urgent manner. Already, there are psychotherapy apps that one can download onto a standard smartphone and use in lieu of a human psychotherapist. In the near future, this simulacrum of a human therapist may be enhanced by the use of android therapists, programmed to duplicate the knowledge and behavior of human therapists instantly, appealing to convenience or self-gratification through technology. This raises a host of ethical questions: can such beings be equally effective, and if so, ought we to reason in a consequentialist manner in their favor so as to increase accessibility and reduce costs through technology? Would there be a psychological difference between automated and potentially anthropomorphic therapy and genuine human therapy, if only a subtle one? Even if a chatbot or robot therapist is transparently such, is there an element of emotional manipulation and potential dishonesty in this interaction? Can the security of clients and their data be thus secured? I will argue that key aspects of chatbot psychotherapy present major ethical and clinical challenges in these areas, although transparent forms of it should not be legally banned.*

**Keywords:** chatbots, technological momentum, automation, psychotherapy, trust, anthropomorphism, data

### Introduction

In this paper, I would like to point to some ethical questions already emerging in the recent development of chatbot psychotherapy. There will be some interdisciplinary content in this, so as to do justice to the full range and depth of challenges that psychotherapists, applied ethicists and philosophers of technology face in this matter. This also reflects my general thinking, as I am both a practicing psychotherapist and a philosopher.

The ethical questions that strike me as especially urgent around this technology relate to professional trust, data privacy and client security,

---

\*Honorary Research Fellow, School of Languages and Cultures (Germanic Studies), University of Sheffield & Faculty, Syracuse University London, (Philosophy and Information Studies), UK.

anthropomorphism, and an excessive degree of instant gratification. The broader philosophical and historical question of technological momentum also raises an important question around chatbot therapy, namely: to what extent are we nudging ourselves into an excessively narrow and harmful technological fix which we may find difficult to correct in the future?

Speaking on Independence Day, 1830, the American statesman Edward Everett declared:

It is the spirit of a free country which animates and gives energy to its labor, which puts the mass in action, gives it motive and intensity, makes it inventive, sends it off in new directions, subdues to its commands all the powers of nature, and enlists in its service an army of machines, that they do all but think and talk. (Cited in Nye 2007, Location 1939, Kindle Edition).

Now the machines may be, on some accounts, virtually thinking, and they certainly are talking.

The ethical and general technological aspects of this topic are thus at once a matter of urgent and perennial concern. This is because interrogations about the risks and values attendant to technology are at least as old as Greek myth and philosophy, and as recent as ChatGPT. It may therefore be of interest to go back even further than the 19<sup>th</sup> century, and relate two Greek myths.

## **Two Greek Myths for the 21<sup>st</sup> Century**

A Greek myth that has always especially struck me as prescient is the myth of Icarus and the sun (Graves 2017, Chapter Ninety-Two). Imprisoned by King Minos of Crete, the artist and engineer Daedalus and his son Icarus attempted to escape with the help of what could be seen as an early form of aviation technology: wings of wax. This initially seemed an efficient and clever type of tool use, until Icarus disregarded his father's warning and flew too close to the sun, melting his wax wings. He then drowned in the Aegean Sea.

Also noteworthy from ancient Greek thought is Plato's recounting of The Tale of Theuth the Egyptian inventor of writing. In *Phaedrus*, Plato recounts how Theuth showed his craft to the god Thamus, who then foretold the decline of memory, linked to the move from an oral to a written culture (Plato 1995).

In both of these Greek myths, the ancients warn us of the dangers of *hubris* and the reckless use of technology, as well as the potential damage of over-dependence on technology. They urge us to handle them with care. Today, as AI and robotics advance at nothing short of spectacular speed, we would do well to heed their counsel and consider where we are going and how we are likely to get there.

Here, I will bracket the important topic of machine consciousness in philosophy of mind. However, I will briefly state that there is a very plausible and strong consensus that consciousness or sentience has not, to date, been instantiated in a machine, and let us assume that this is correct. It therefore must follow that current chatbot therapy involves sentient human beings who are seeking the care

of non-sentient computer programs, which they will generally understand to be the case. Although, on some accounts, machine consciousness may emerge during the course of this century, I believe that the highly probable debate and general controversy around this very process will negate the satisfactory resolution of the problems here discussed. In this article, I will limit myself to some aspects of a topic that greatly concerns me both as a philosopher and a practicing psychotherapist: the ethics of the potential use of Chatbots as replacements for psychotherapists; in other words, the automation of mental health care.

### **The General Automation Scene**

We are living in an era in which automation is progressing at remarkable speed. There is much here both for philosophers of technology and psychotherapists to ponder. By way of general historical and social scientific background, a few key observations: One estimate from the Chief Economist of the Bank of England indicates that as many as half of all UK jobs will soon be automated (Elliot 2015, Litwack 2021). A very significant number of CEOs, engineers, and state agencies see the optimal development of automation as one of the key challenges of our time, and they are wise to do so. However it occurs, it has already changed the way we see and engage in the worlds of technology and work.

There is a lively and broad-ranging debate on the full ethical and philosophical implications of rapid and extensive automation. Optimists, such as Silicon Valley entrepreneur and analyst Jerry Kaplan (Kaplan 2015), hold that automation offers great promise for the fulfilment of human aspirations, and that technological unemployment can be offset both by worker retraining in other sectors, and a well-designed unemployment insurance scheme. Pessimists such as Martin Ford (Ford 2021) are concerned that the speed and breadth of contemporary automation may well overwhelm us over the coming decades, should we not be exceptionally careful. He believes that a basic minimal income scheme is called for to offset this trend.

Part of this perspective stems from concerns over unmanageable technological unemployment, as well as a growing sense that we may be dominated or even destroyed by superintelligent machines of our own creation. Furthermore, some pessimists, such as the philosopher and motorcycle mechanic Matthew Crawford, are concerned that the rise of the machines will lead to the acceleration of an already lamentable tendency towards declining mechanical and craft skills in industrial society (Crawford 2015). One might also speak here of a narcissistic need for instant gratification, of not earning what one has by building it. That isn't always the case; consider the benefits of high-tech improvements in areas such as theoretical science, modern medicine, and public health. But with reference to some skilful activities linked to human interaction, practical skills, and understanding, Crawford is right to remind us that faster and easier is not always better.

### **Current AI Psychotherapy: Technological Momentum and Nudging?**

What is the situation with regards to psychotherapy and the current Chatbot controversy? There are already several specific chatbot therapy programs available; some are free—like Woebot—and others have a paywall. One of the first ethical questions that presents itself is: is the specific roll-out of this technology directing or nudging us in a direction that may be harmful or distorting?

By “nudging”, I am following Thaler and Sunnstein’s account of how technologies can maximise the odds of certain choices, whether good or bad, by their very form or application (Thaler and Sunnstein 2021). This may well be the case with some psychotherapy chatbots.

Several chatbot therapy developers claim to use the common therapeutic approach of cognitive-behavioural therapy (CBT), which is already widely preferred by state and private health care providers, in part due to its short duration and lower cost than long-term psychodynamic therapy. The current pattern of chatbot therapy programs already indicates a further narrowing of therapeutic range, as psychodynamic approaches to talk therapy may be harder to program at a high and sophisticated level using even the most sophisticated contemporary AI. What if, as is likely, psychodynamic approaches to psychotherapy are to be considered preferable for some clients in mental distress? General CBT nudging would then be a clear harm.

If we accept an eclectic model of psychotherapeutic technique that allows for a high degree of client choice, this tendency could represent a case of client nudging through technological momentum, to use a notion from the historian of technology, Hughes (1994). This insightful concept holds that the availability of a particular technology, with few if any saleable and infrastructural alternatives, implicitly directs client or consumer choice in a narrowing manner. This could be seen as a large scale and powerful form of nudging.

Hughes offered this notion as a hybrid alternative both to a radical social constructivism and a rigid technological determinism. He means by the former term a perspective on technology according to which there is little if any necessary causation in technology, it being the product of consumer or users’ choices and modification by producers in response to these choices. By the latter, he means the idea that technology is the primary, or even the sole driver of history.

The former perspective, he thought, sees technological development as more open-ended than it actually is, and the latter perspective sees it as more strictly caused and controlled than is the case. A comparison from automotive history may prove useful here.

It is likely a little-known fact today that most of the earliest late nineteenth century vehicle engines were either electric or steam-engines; only about twenty percent of American cars were gasoline-based in 1900 (Appleyard 2022, p. 18). So, in a sense, the current move towards replacing the internal combustion engine with electric vehicles is actually a return to a *past* technology, now seen as the better road not travelled. We didn’t know about future climate change and dependency on foreign energy supplies in the early twentieth century. Had we

known about these problems, I am inclined to believe that incentives would have been set up so as to nudge the perpetuation of electric (and possibly steam) vehicles, thereby discouraging the further development of internal combustion engine machines. As such, the massive automotive infrastructure established in the early twentieth century around the internal combustion car might never have occurred. This infrastructure has included service centres, oil wells, engine parts, and targeted advertising.

Reading this as technological momentum doesn't entail seeing it as a matter of fully free consumer choice among equally viable alternatives along the lines of social constructivism, nor does it entail seeing it as an inevitable determinism, no matter what choices were made by social organisations and consumers. Rather, it entails that it is not surprising that one technology came to prevail over others, given its support from a growing infrastructure and its interaction with manufacturers and consumers.

A further key factor in the technological momentum of chatbot therapy may prove to be the mental health care crisis that we are now in, in which numerous prospective psychotherapy clients encounter long waiting lists for therapy and/or treatment costs that they cannot afford. Open access chatbots are already available instantly and at no cost, and even the paywall programs will likely prove to be significantly cheaper than a course of treatment with a human therapist. These chatbots are, of course, available for consultation 24-7, which is not conducive to the development of client autonomy and a mature sense of deferred gratification. Both autonomy and the capacity to defer gratification are necessary components of mental health and maturity, in my view. Their ubiquity within the process is more beneficial than the lack of a waiting period to begin psychotherapy. In effect, clients may be thereby nudged and enabled in narcissistic and dependent behaviours during their use of this technology.

To be fair to CBT, it may be eventually proven to the satisfaction of all psychotherapists to be incomparably superior to its psychodynamic rivals, but this is far from being the case today, and I am doubtful that it ever will be. If contemporary chatbot technology nudges us towards CBT, we may discover that its future popularity facilitates less than optimal results for psychotherapy for the widest possible range of clients.

### **Trust and Psychotherapy Ethics**

A further and fundamental ethical question surrounding chatbot therapy is the question of trust. Trust in professional care providers of all types is fundamental to the very exercise of care, as argued by bioethicists such as Pellegrino (1991, pp. 69–92). If we do not have a full fiduciary relationship with the professional with whom we work, the work itself is not likely to succeed due to mutual suspicion and a coldly legalistic framework, he holds. Trust, in Pellegrino's sense, ought to be considered a necessary condition for psychotherapy, and there must be some of it on both sides of the psychotherapeutic dyad.

How can even a modicum of mutual trust be present, when the therapist is known *not to be even a sentient human being*? Trust entails intentionality and consciousness for its full meaning. Worse, still, what if there is an element of sheer deception and manipulation in the provision of chatbot therapy not employing text-messaging as is currently standard, but a realistic humanoid visual figure that may be difficult or even impossible to distinguish from a sentient human being with his or her own memories, experiences, and emotions upon which to draw?

I take this to be the worst-case scenario, in which the client is wilfully deceived as to the ontology or identity of his or her psychotherapist. This would imply robotic perfection, beyond what Mori (1970) has termed the “uncanny valley” in which we have more of a sense of the eerie or uncanny with robots that are only slightly short of perfection, as opposed to clearly mechanical ones, or perfect ones. This well-known phenomenon in the psychology of technology is likely caused by blurring the line between the human and the mechanical, which would be the very point of advancing chatbot therapy.

It is especially important to note that on any reasonable account of professional ethics, not just procedural, but what might be henceforth termed “full ontological disclosure” is required in the therapeutic process. This means that the client has a right to know whether he or she is talking to a human being or to an AI. One could well here draw on Beauchamp and Childress’ classic “Four Principles” approach to health care ethics, and in particular its first principle of respect for the personal autonomy of the client (Beauchamp and Childress 2019, Chapter Four). A lack of full ontological disclosure can only be seen as a violation of such required professional respect. The reasonable legal or professional requirement of a watermark for such technologies would be an effective means of preventing this violation of principle through informed consent and an indication of robotic identity in all interactions, like recording in progress signals.

Furthermore, it is worth noting here that maximising access to psychotherapy at any cost, which would seem to be fundamental to any strongly consequentialist approach to this general topic, would seem to vitiate this fundamental character of client autonomy, as well as being subject to the reasonable objections from the perspective of anti-utilitarian critiques of radical consequentialism (Smart and Williams 1973). A full treatment of this interesting topic lies beyond the parameters of this article.

The current concern about deep fakes underlines the fundamental values of truth and trust, and this applies not just to media ethics, but to a wide range of social and political questions, including the ethics of psychotherapy (Citron 2019). Should we fail to be careful in this matter, the potential for fraud and deception across the board is obvious.

None of this latter possibility is currently fully possible for psychotherapy, but according to some analysts, the current state of chatbot technology may prove to be preparatory to flawless deep fakes and superintelligent AIs that will make our current technology look nothing short of primitive. Should this technology reach new proportions such that we have reason for believing that it is as sentient as we are, or even beyond us, then this controversy will take on new ethical and psychological dimensions. I believe that we would then find ourselves faced with

a disturbing dilemma of our own creation: is the entity before me conscious or not? If it is conscious and intelligent, that is, capable of intentional states, feelings, and any other psychological phenomena that we normally attribute to human beings, we would likely be inclined to attribute genuine moral and social status to it. We would then see it not only as a true psychotherapist, but as a person as well, with all that this attribution implies with reference to trust, empathy, and intentionality.

### **The Brilliance and Limitations of the Turing Test**

The current controversy over whether or not any program has already passed the Turing Test, and to what extent this clever 1950 thought experiment is sufficiently comprehensive, may thus prove to be only the beginning of a problem of psychological verification, as much as one of moral ontology. We may not, because of the problem of other minds, achieve in this matter a reasonable consensus on whether or not we are dealing with a person, let alone a true psychotherapist. And just as some will be satisfied with an operational model of consciousness, others will not be satisfied that the machine's light is really on, no matter how many tasks an advanced machine with impressive chatbot functionality can perform.

In this matter, I am in agreement with Marcus and Davis (Marcus and Davis 2019, Chapter Eight) that for all of its operational cleverness, the Turing Test measures the capacity of a programmable machine to *simulate* consciousness, rather than offering grounds for the certain *ascription* of consciousness to such a machine.

Many of us will likely want to attribute some degree of sentience to advanced chatbot therapists, if only because doing so will justify our confiding our deepest secrets to them, especially if they are even more realistic than contemporary programs. This confirmation bias may lead to delusionality and excessive anthropomorphism. Nothing is truly conscious just because we want it to be. I would take that to be a narcissistic delusion, which is likely to be reinforced by 24-7 availability and reduced or even zero cost. It could distort our social relations generally, as well as the psychotherapeutic process. In effect, we would thus be blurring the distinction between truth and reality in a culture already in the grips of a tendency towards post-truth and advancing deep fake technology (McIntyre 2018, *passim*).

### **Psychotherapy Should Not Be Excessively Easy**

On this supposedly beneficial therapeutic process, perhaps a clinical word, with reference to psychodynamic approaches. Psychotherapy involves real emotional *work*. Not just for the client, but for the therapist. Psychoanalytic approaches stress the importance of the transference and countertransference, of the enactment of the childhood complexes of the client in his or her behaviour

towards the therapist, as well as the therapist's assimilation of perceptions and feelings, in general perspectives on the client. Non-psychoanalytic forms of psychodynamic therapy tend to focus on a more present-oriented notion; the working relationship between client and therapist.

Furthermore, in attachment psychotherapy, there is an underlining of the importance of an attunement between therapist and client, a meeting of two conscious beings in the therapeutic relationship (Holmes 2014, Chapter Seven).

On any psychodynamic account, replicating this relationship is a necessary condition for the full equivalence between human and chatbot therapists. This, at its full extent, would seem to imply a relationship between two conscious beings. If that is not present, we should probably coin a neologism for the overall process. It might be of at least limited help to some, but it would not be complete psychotherapy.

One possible constructive use for chatbots and companion robots is currently being researched. Studies have been done on the use of companion robots with neurodiverse, and in particular, autistic children (Lucaciu 2013). That may well be an effective and morally unproblematic use of robot and AI technologies, with the proviso that the final goal of the work will be a graduation to more comfortable social relations with human beings.

### **The Automated Therapeutic Seduction**

In her 2015 book, *Reclaiming Conversation*, psychologist Sherry Turkle warns us of the potential for delusionality and manipulation in our interactions with advanced and lifelike technologies. We may come, she thinks, to see companion robots such as Paro the seal, a companion robot that looks and even "behaves" almost like a real animal, as akin to living beings worthy of care and capable of giving it (Turkle 2015, p. 349). However, in the end, Turkle believes that we are here being emotionally seduced or manipulated by what is merely a programmed simulacrum. There is indeed great potential for manipulation and delusionality in this, and Turkle is correct to go on to claim that it is all too easy to provide elderly and disabled patients with such a cute machine for reasons both of cost and convenience. This is being done now, mainly in Japan (Piore 2014).

There is a deep psychological concern in this matter, which I believe can be understood in attachment terms. If, following John Bowlby, we hold that the avoidance of security-producing and exploratory contact with our fellow human beings is a form of developmental disorder, then we are unlikely to be swayed by appeals to a technology of this sort being an inevitable wave of the future (Bowlby 1988, *passim*).

Given that the development of client autonomy and a high degree of relational trust is central to the practice of psychotherapy, one must also ask if these qualities can emerge in our interaction with non-sentient robots and AIs in the same way as in our interaction with conscious human professionals. The hard work of psychotherapy has its ups and downs during an individual's course of therapy. There will be smooth sessions and challenging ones, containing moments



of joy, despair, grief, and anger. Sessions will be missed or arrived at late, illness or technical problems will occur periodically. These must be jointly navigated by therapist and client for months or years, as personality changes in the client are facilitated by facing and dealing with the very difficult imperfections of the process. Could such a volatile and subtle background process be programmed into a future sentient AI? Even if so, how much delusionality and emotional manipulation can be justified along the way? All of this in addition to the likely unresolvable problem of attaining a clear and effective consensus on the question of machine consciousness.

### **In the Shadow of ELIZA**

Once more, philosophers and psychotherapists have much to learn from history. There was an early forerunner of the current and possible future situation. As far back as 1966, Joseph Weizenbaum tested the world's first psychotherapy chatbot, the original ELIZA (Weizenbaum 2015).

Weizenbaum's ELIZA was a standard computer console of its era—and therefore not remotely humanoid. Furthermore, its programming could be described as a very simple version of client-centred therapy. And yet, much to Weizenbaum's own concern, his laboratory subjects often rapidly anthropomorphised it, knowing that it had no more consciousness than a toaster. They confided deep personal problems to it, to the point of requesting that Weizenbaum leave the room in order to do so. For this reason, he became an articulate and passionate critic of what he took to be the irresponsible anthropomorphising of computer and AI technology (Schanze 2010). Weizenbaum, one of the world's greatest computer scientists, believed that we should not share our intimate secrets with a computer.

How much more temptation to anthropomorphise is there with the latest chatbots, such as ChatGPT? And if, as is likely, first deep fake avatars and eventually in the room realistic androids or holograms can provide a simulacrum of psychotherapy algorithmically, then we will be reminded of the fact that sometimes yesterday's science fiction becomes tomorrow's science fact. Future life will look increasingly like a Philip K. Dick story, and many of us will not be entirely pleased with such simulacra.

### **Deep Data Breaches**

I have spoken of sharing our intimate secrets. With chatbot therapy, this entails the use and potential sharing of the most intimate data that anyone can generate. There are privacy commitments to be expected, but data breaches do occur regularly, and so does hacking. This both returns us to the question of trust in professionals, and it raises the ethical and legal matters of client autonomy and liability for violations of privacy. Should there be a psychotherapy data breach

through negligence or hacking, who is responsible? The programmers? The company for whom they work? The possibly sentient AI itself?

Similar concerns are raised by the possibility, already alleged (Bharade 2023), that psychotherapy apps have nudged users to suicide, or potentially to criminality.

There is a parallel here to the ongoing debate over ethical and legal responsibility with reference to autonomous vehicles, and although road fatalities are not at stake in chatbot therapy, we should nonetheless agree that intimate secrets should remain between psychotherapists and their clients. Furthermore, if AIs have indeed nudged users to suicide or criminality, these allegations ought to be investigated and if substantiated, prosecuted along lines comparable to the policies applied to human therapists. However, this would seem especially challenging, given the complex levels of potential design here—much like with the ongoing debate over autonomous vehicles and legal responsibility. One possible solution would be to have software engineers and designers sign off legal responsibility to the companies for which they have produced the apps, with final quality and safety control to be done by said companies which would then be liable for harm due to risks deemed excessive or negligent.

Furthermore, will people feel comfortable seeking the help that they require, if they have good grounds for trusting chatbot therapists less than people? If they trust them more, that would seem to imply social anxiety that should be treated by members of human society. In psychoanalytic and attachment terms, a pattern of avoidance would thus be enabled by computer technology.

Psychotherapists, are flawed, like all human beings. We are not fool proof systems. However, we cannot be hacked for nefarious purposes, nor is there a tendency among us to sell clients' intimate secrets for profit to insurance companies or, in some jurisdictions, to government agencies. Once massive profit enters the picture, we are at risk of what Shoshana Zuboff has termed 'surveillance capitalism', in which our data is problematically shared for profit. Or it could be used, notably by authoritarian states, in order to control and punish their citizens. This, too, is already well underway, with the potential for totalitarianism beyond the imagination of history's worst dictators. One might add the risk of extortion in cases in which unscrupulous hackers could threaten to expose psychotherapy clients' intimate secrets if they do not send funds immediately.

I do realise that this may seem to be dystopian doomsaying, but I am here inclined to agree with Ludwig Wittgenstein that:

the work of the philosopher consists in assembling reminders for a particular purpose (Wittgenstein 1976, Part I, Section 126).

If we hold this to be true, at least when controversial matters are involved, then we not only can risk playing Cassandra, we must do so. So much is at stake here that moral responsibility entails caution of the sort that has motivated the recent call for a moratorium of at least six months—not a cessation—on chatbot development so as to explore optimal governance procedures that would include legislation and advanced security measures (Future of Life Institute 2023). Leading computer scientists and Silicon Valley entrepreneurs are among those

who have signed this petition, which acknowledges both the promise of AI and its potential perils for the future of humanity.

With reference to transparently chatbot psychotherapy, we would do well to remain in this spirit. Let us acknowledge its potentially beneficial aspects, as well as its indicated pitfalls, while advocating greatly improved governance and ethical judgment. Low or even no cost therapy, and shorter (but not necessarily no) waiting time to begin a course of therapy are indeed advantageous developments, even if we are talking about a stop gap measure that is in need of empirical study concerning outcomes.

With reference to legislation, an all-out ban on the *transparently* virtual varieties of psychotherapy would be excessively draconian within a democratic society. We have to allow for a reasonable range of choices that may be far less than optimal, as long as harm is contained, and reducible.

That is likely the case here, given the foreknowledge of the psychotherapy clients of what they are talking to—full ontological disclosure—and the potential for the wide dissemination of critical information encouraging richer and more authentically human forms of this work. Sometimes we must look to the state to prohibit severe harm, but the object of our discussion here may be a more moderate harm, or settling for a less than optimal activity as a technological quick fix. In that sense, transparently virtual forms of psychotherapy may be, at best, comparable to vitamin supplementation in lieu of a well-balanced diet. Not as good, but better than severe malnutrition, thanks to advanced technology. Should more than anecdotal evidence emerge indicating that this form of the technology is worse than that, then state regulation should certainly be considered. The trick then would be to counteract the potentially huge underground economy that would accompany the prohibition or severe curtailment of the technology. It may well not be worth an attempt at strict prohibition.

Having said this, it is my belief that much more must be done to provide access to psychodynamic therapy of various schools with human beings at reasonable cost and in reasonable time. This may not be easy. But the ethical and psychological aspects of this problem necessitate a valiant effort. Nudging or even requiring psychotherapists to provide some concessionary rates will likely be part of this strategy, as will the training of more professional therapists.

Above all else, we should be cognisant of the need to distinguish between sharing intimate secrets with a trained and conscious human being who has a genuine professional and personal concern for us, with what (for now) is in the end an obvious simulacrum with no more consciousness than a toaster.

## References

- Appleyard J (2022) *The Car: The Rise and Fall of the Machine that Made the Modern World*. London: Weidenfeld & Nicolson.
- Beauchamp T, Childress J (2019) *Principles of Biomedical Ethics*. New York: Oxford University Press.

- Bharade J (2023) *A Widow is Accusing an AI Chatbot of Being a Reason Why her Husband Killed Himself*. Business Insider. Available at: <https://www.businessinsider.com/widow-accuses-ai-chatbot-reason-husband-kill-himself-2023-4>.
- Bowlby, J. (1988) *A Secure Base*. Abingdon: Routledge.
- Citron D (2019) *How Deep Fakes Undermine Truth and Threaten Democracy*. TEDSummit. Available at: [https://www.ted.com/talks/danielle\\_citron\\_how\\_deepfakes\\_undermine\\_truth\\_and\\_threaten\\_democracy](https://www.ted.com/talks/danielle_citron_how_deepfakes_undermine_truth_and_threaten_democracy).
- Crawford M (2009) *Shop Class as Soulcraft: An Inquiry into the Value of Work*. London: Penguin Press.
- Elliot L (2015) *Robots Threaten 15m UK Jobs, Says Bank of England's Chief Economist*. Available at: [www.theguardian.com/business/2015/nov/12/robots-threaten-low-paid-jobs-says-bank-of-england-chief-economist](http://www.theguardian.com/business/2015/nov/12/robots-threaten-low-paid-jobs-says-bank-of-england-chief-economist).
- Ford M (2021) *Rule of the Robots: How Artificial Intelligence Will Transform Everything*. London: Basic Books.
- Future of Life Institute (2023) *Pause Giant AI Experiments: An Open Letter*. Available at: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.
- Graves, R. (2017) *The Greek Myths*. London: Penguin Press.
- Holmes J (2014) *John Bowlby and Attachment Theory*. Hove: Routledge.
- Hughes TP (1994) Technological Momentum. In L Marx, MR Smith (eds.), *Does Technology Drive History? The Dilemma of Technological Determinism*. Cambridge: The MIT Press.
- Kaplan J (2015) *Humans Need Not Apply*. New Haven: Yale University Press.
- Litwack EB (2021) Bertrand Russell on Automation: How Necessary are Human Beings? *Bulletin of the Bertrand Russell Society* 163: 61-69. Available at: <https://bertrandrussellsocietyorg.files.wordpress.com/2022/01/32e20-brs-bulletin-spring-2021.pdf>.
- Lucaciu I (2013) *Robots: The Answer for Treating Children with Autism Spectrum Disorder?* Available at: <http://www.theneuroethicsblog.com/2013/07/robots-answer-for-treating-children.html>.
- Marcus G, Davis E (2019) *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Pantheon Books.
- Marx L, Smith MR (1994) *Does Technology Drive History? The Dilemma of Technological Determinism*. Cambridge: The MIT Press.
- McIntyre L (2018) *Post-Truth*. Cambridge: The MIT Press.
- Mori M (2012) *The Uncanny Valley*. Available at: <https://spectrum.ieee.org/the-uncanny-valley>.
- Nye D (2007) *Technology Matters*. Cambridge: The MIT Press.
- Pellegrino ER, Veatch RM, Langan JP (1991) *Ethics, Trust, and the Professions*. Washington DC: Georgetown University Press.
- Piore A (2014) Will Your Next Best Friend Be a Robot? *Popular Mechanics*. Available at: <https://www.popsci.com/article/technology/will-your-next-best-friend-be-robot/>.
- Plato (1995) *Phaedrus*. Indianapolis: Hackett Publishing Company.
- Schanze J (2010) *Plug & Prey*. Maschafilm. Available at: <http://www.plugandpray-film.de/en/>.
- Smart JJC, Williams B (1973) *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.
- Thaler R, Sunstein C (2021) *Nudge: The Final Edition*. London: Allen Lane.
- Turkle S (2015) *Reclaiming Conversation: The Power of Talk in a Digital Age*. New York: Penguin Books.
- Weizenbaum J (2015) *Islands in the Cyberstream: Seeking Havens of Reason in a Programmed Society*. Sacramento: Litwin Books.
- Wittgenstein L (1976) *Philosophical Investigations*. Oxford: Basil Blackwell.

Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.