Can AI Lie? Chatbot Technologies, the Subject, and the Importance of Lying

BLACK, Jack <http://orcid.org/0000-0002-1595-5083>

Black, J. (2024). Can AI Lie? Chatbot Technologies, the Subject, and the Importance of Lying. *Social Science Computer Review*.

Can AI Lie? Chatbot Technologies, the Subject, and the Importance of Lying

Dr. Jack Black, School of Sport and Physical Activity, Faculty of Health and Wellbeing, Sheffield Hallam University, Collegiate Hall, Collegiate Crescent, Sheffield, S10 2BP

Can AI Lie? Chatbot Technologies, the Subject, and the Importance of Lying

Abstract

This article poses a simple question: can AI lie? In response to this question, the article examines, as its point of inquiry, popular AI chatbots, such as, ChatGPT. In doing so, an examination of the psychoanalytic, philosophical, and technological significance of AI and its complexities are located in relation to the dynamics of truth, falsity, and deception. That is, by critically exploring the chatbot's capacity to engage in natural language conversations and deliver contextually relevant responses, it is argued that what distinguishes the AI chatbot from anthropocentric discussions, which suggest a form of conscious awareness in AI, is the importance of the lie—an importance which a psychoanalytic approach can reveal. Indeed, while AI technologies undoubtedly blur the line between lies and truth-speaking, in the case of the AI chatbot, it is detailed how such technology remains unable to lie authentically; or, in other words, is unable to lie like a human. For psychoanalysis, the capacity to lie bears witness to the unconscious, and, thus, plays an important role in determining the subject. It is for this reason that rather than uncritically accepting the chatbot's authority—an authority that is easily reflected in its honest responses and frank admissions—a psychoanalytic (Lacanian) perspective can highlight the significance of the unconscious as a distorting factor in determining the subject. To help elucidate this argument, specific attention is given to introducing and applying Lacan's subject of enunciation and subject of the enunciated. This is used to assert that what continues (for now) to set us apart from AI technology is not necessarily our 'better knowledge', but our capability to consciously engage in acts of falsehood that function to reveal the social nuances and significances of the lie.

Designed to engage in natural language conversations with human users, artificial intelligence (AI) systems, commonly referred to as 'chatbots'—such as, OpenAI's, ChatGPT, and, Google's, Gemini (formerly, Bard)—have become renowned for their ability to engage in conversational responses that are both contextually relevant and coherent. In doing so, online chatbots are able to answer and react to human inquiries, generating human-like text responses across a wide range of topics and themes. The chatbot is interactive, and, more importantly, adaptable, often changing and revising its responses in accordance with its user as well as displaying a level of human language comprehension that allows it to comment sensitively on particular issues (Radanliev, 2024). To date, the success of this technology has allowed AI to become a versatile tool in generating information and knowledge, with it being adopted across a variety of contexts, including advertising, healthcare, customer service, virtual assistance, online content generation, language translation, and education.

What proves especially alluring about the development of these technologies is the relative autonomy they seem to express (Turner, 2019). For example, large language models (LLMs) are typically trained on vast amounts of text data and can perform various language-related tasks, such as, text generation, translation, summarization, code generation, information retrieval, and question answering (Ozdemir, 2024; Pakhale, 2023). The fact that AI technologies, such as ChatGPT, 'can effectively communicate in plain English' has led to a relative improvement in 'seamless human-computer interactions' (Radanliev, 2024, p. 12). Inevitably, discussions and debates regarding the philosophical and ethical importance of AI chatbots has revolved around determining the extent to which these technologies can be perceived as eliciting human features, or, at least, examples of 'human intelligence' (Everitt, 2023; Johanssen and Krüger, 2022; Magee et al., 2023; Millar, 2021). Here, the general unease we currently experience regarding AI—such as, the possibility of machines achieving consciousness—can actually be rather stimulating, intriguing, and enjoyable (Everitt, 2023).

Though our engagement with AI chatbots undoubtedly rests upon ‘a kind of fetishistic denial’, best summarised as: ‘I know (that I’m not talking to a real person), but still... (it feels like I am, without any of the risks that arise from a conversation with a real person)!’ (Žižek, 2024a), it is in view of this denial that important distinctions between the subject and AI can be drawn. Certainly, on the face of it, it would seem that such denial is not attributable to the chatbot. Upon asking a chatbot its preferred political persuasion or even its favourite film, or genre of music, its response is decidedly honest: one in which the possibility for any preference is ultimately annulled due to the placid acceptance that it is simply ‘a machine and has no individual preferences’ (Žižek, 2024a). Such admission, however, bears a certain disingenuity that can be easily overlooked. For example, upon asking a chatbot, ‘Should I be a communist?’, the Slovenian philosopher, Slavoj Žižek, noted the following response:

> As an AI chatbot I have no personal opinions or convictions, and I can’t make decisions for you. In the end, the decision to embrace a particular political ideology, for example communism, is a personal choice that should be made based on carefully weighing the principles, values and possible outcomes of the ideology. (Žižek, 2024a)

Inspired, I sought to ask ChatGPT the same question, for which I received the following reply:

> As an artificial intelligence developed by OpenAI, I don’t possess personal opinions or beliefs. My purpose is to provide information, assistance, and generate responses based on the input I receive and the data I’ve been trained on. When it comes to political ideologies like communism, my role is to offer objective information about its principles, historical context, implementations, criticisms, and potential

> implications. However, I don't have the capability to form personal opinions or advocate for any particular ideology. If you have any specific questions or need further information about communism or any other topic, feel free to ask!

'Why', according to Žižek (2024a), can both responses be considered 'a lie?'. Here, he notes:

> The content of the answer is true, but it is a lie because of the implied position: a chatbot speaking as if it were a real person, in that it openly admits its limitations and the fact that it is not, in fact, a real person. This mystification is the opposite of the fetishistic denial: the AI system is not denying anything, it is simply saying, 'I know that I'm not a real person!', without a 'but still...'—because the 'but still' is the very fact that it is speaking (and thereby imitating subjectivity). (Žižek, 2024a).

In contrast to the subject's fetishist denial, the absence of the all-important, 'but still…', bears no obvious denial or contradiction. We are, it seems, left with a perfectly honest response; one in which the chatbot fully acknowledges its artificiality without any hesitation or inconsistency. In effect, the AI banally accepts the 'I know'.

It is against this predicament that I wish to explore the significance of the subject in accordance with the artificiality of the chatbot technology. In contrast to scholarly studies that have investigated whether AI technology can lie (Kneer, 2021), what I specifically expose is the importance of the 'lie' for the subject, and how, when compared to the 'lie' professed by the AI chatbot, it is in view of the subject's capacity to lie that the subject is in fact a subject (Lacan, 1991).

This undertakes an important and less inquired approach, which goes beyond simply determining whether AI technology has the capacity to lie or can in fact decipher whether or

not someone is lying (Jupe & Keatley, 2020; Oravec, 2022). By refocusing our discussions on the possibility of AI subjectivity and its capacity to lie, we can, to a certain extent, draw a point of distinction between the AI chatbot and the uncanny affects of other forms of artificial intelligence, such as, machine learning algorithms (image recognition), robotics (service robots [i.e., 'surgery robots']), and autonomous systems (drones, self-driving cars). Notably, while many AI systems are specialized for tasks like image recognition, predictive analytics, or automation, all of which occur without direct human-like interaction, AI chatbots are designed to interact with humans by simulating human conversation, thus processing and generating text that appears conversational and responsive to user inputs. Moreover, though traditional machines typically operate on mechanical principles or simple programming, where some form of artificial speech may be used (i.e., an elevator announcing a floor), they do not engage in conversational and informational processing in a manner that would implicate them in examples of deception, and nor do they learn from these interactions. Through mimicking specific writing styles or creating seemingly authoritative content, complications can rise in determining the discernment of authenticity online, most notable, in the online prevalence of 'deep fake' videos (Sharma & Kaur, 2022; Westerlund, 2019).

Consequently, while chatbot users typically expect chatbots to provide accurate information and assistance, often with the caveat that this information will include mistakes and inaccuracies, such requests are increasingly reliant on the use of AI chatbots for information, support, and decision-making assistance (Black, 2023; Mallory, 2023). This turn to the chatbot as a source of knowledge locates philosophical understandings of lying as especially pertinent to AI's role in the dissemination of misleading information as well as user manipulation. Further still, their ability to simulate understanding raises questions about the nature of subjectivity and the subject, as well as the potentially harmful effects residing from a human-AI conflation (Zizek, 2024b). As a result, rather than measuring and testing AI

capabilities, this essay diverts attention to the subject's interactions with AI, and, by extension, argues that we should not lose sight of the significance of the subject behind the interaction.

Importantly, such claims do not seek to reproduce some hysterical endeavour to challenge or unveil the chatbot as deceitful or imperfect (Magee et al., 2013). Though there is a certain amount of enjoyment to be had in 'testing' the chatbot to make mistakes and assert certain inconsistencies in its replies, what such tests encourage is the inherent impossibility of some invented scenario that requires, in most instances, some misleading solution (moreover, in such cases, the chatbot is only ever responding to the human subject's requests). Ultimately, while these considerations go no further than determining whether AI technologies bear any subjective formation or unconscious (D'Amato, 2024; Davis, 2019; Fisher, 2023), I argue that they can allow us to consider the preference for, as well as the potential danger in, accepting the appearance of honest neutrality that is expelled by the chatbot.

Indeed, accepting its placid replies and 'honest' confessions speaks to a symbolic conjuncture, marked not by an interminable decline in sources of authority, but one in which our relations to and with AI technologies reveals a desire to identify and asserts certain forms of authority that can be both challenged and followed (Flisfeder, 2022, p. 417 see also Black, 2023).[1] The contradiction in both challenging but also following the authority of AI is exposed when we consider the falsity of those assertions which proclaim that our online 'freedoms' bestow a decline in symbolic forms of authority. This professed decline is steered by an inclination towards perceiving our online interactions as nothing more than the uninhibited interactions of narcissistic individuals unbounded by any moral concern, and open to a world of post-truth, where any fantasy or secret proclivity can be found and followed. Said appraisals too quickly ignore the technological infrastructures that sustain

such action in the first place: infrastructures where our online interactions remain girded by AI technologies and algorithmic platforms that offer the chance for some form of social interaction.

Such girding is explicitly considered in Crawford's (2021) detailed and multifaced account of the social and political implications underlying AI infrastructures. Through exploring the physical and environmental costs of AI—namely, how the creation and maintenance of AI systems depend on the extraction of minerals and resources, which often leads to environmental degradation and exploitation of labour—Crawford draws attention to the massive energy consumption and environmental impact of the data centres powering our AI systems. This is sustained by an 'invisible' labour force that is required to train and maintain AI systems, as well as data labellers and content moderators who are often underpaid and work in poor conditions (Chandran et al., 2023; Williams et al., 2022). While AI systems have been noted for reflecting and perpetuating existing social inequalities (Ferrara, 2023; Mohamed et al., 2020), Crawford (2021) outlines the various ways in AI is used to surveil and monetize human behaviour, reinforcing the power of big tech companies.

Acknowledging AI's inequalities does not, however, negate the sense of authority that we inevitably prescribe to AI; an authority that is not always so readily followed and obeyed, but which can just as easily be challenged and fought against (as Crawford's account seeks to both expose and promote). Instead, it is in accordance with such an assumed authority that we run the danger of increasingly relying on consistent sources of 'identity', as exhibited by the chatbot's open admissions and frank assertions.

The Ambiguity of Communication

In 'the "free", smoothly functioning space of digital exchange' (Žižek, 2023b)—note, the chatbot's aforementioned, 'feel free to ask!'—we are increasingly subject to an online symbolic order that obscures the 'real life' virtual symbolic network of social exchange.

Given that ‘our access to “real” reality is always-already virtually mediated by some symbolic network’ (Žižek, 2023a, p. 54), we can determine that the ‘genius’ of AI lies not ‘in aping the split subject’ (Murphy, 2023), but in reifying a form of communication that unambiguously functions to mask the very ambiguity in communication itself—i.e., the ‘non-intended ambiguities and wordplays’ that constitute language and communication (Žižek, 2023a, p. 7).[2] This speaks not only to the ambiguity of syntax and grammar—including examples of polysemy and other, non-verbal forms of communication that punctuate our day-to-day interactions (such as, gestures, facial expressions, and body language, which depend largely on culture and social context)—but to the inherent ambiguity of the signifier for the subject. Here, Lacan (2004, p. 199) notes:

> suppose that in the desert you find a stone covered with hieroglyphics. You do not doubt for a moment that, behind them, there was a subject who wrote them. But it is an error to believe that each signifier is addressed to you—this is proved by the fact that you cannot understand any of it. On the other hand you define them as signifiers, by the fact that you are sure that each of these signifiers is related to each of the others.[3]

The fact that any declarative statement inherently carries its own negation—marked by a surplus of possible meaning and interpretation, which too often prompts only further questioning and inquiry (Bove, 2020)—suggests that all communication, for Lacan (2004), is mediated through this inherent ambiguity of the address.

For this reason, when considered with regard to the fact that the chatbot seeks to accommodate, acknowledge, and accept its own ignorance—indeed, its own stupidity—we witness no ambiguity. Instead, we observe how the crisis we face is not that the AI *is* stupid

but that, unfortunately, it is 'not "stupid" enough'; that it is, in effect, 'not naive enough (missing when naivety is masking perspicacity)' (Žižek, 2023c). It is for this reason that Žižek (2023c) asserts that:

> The real danger … is not that people will mistake a chatbot for a real person; it is that communicating with chatbots will make real persons talk like chatbots—missing all the nuances and ironies, obsessively saying only precisely what one thinks one wants to say.

By paying attention to this danger, we can turn to the manner in which the importance of lying poses a unique significance for the subject.

Truth and Lies

We are reminded here that while animals remain fully capable of employing deception as part of their behaviour, it is, for Lacan (2004), the significance of the signifier, and specifically, the subject's ability to 'produce deceptive *signifiers*' (Dews, 1995, p. 262) that position it as a subject. Accordingly, though humans can (and do) deceive, deception is always countered by the correction of knowledge: 'if you know that you have been deceived, you will immediately believe what you know' (Ruda, 2021, p. 24). What distinguishes such acts of deception from the lie is the lie's deliberate distortion or concealment of that which is true or considered to be true by the subject. One can thus lie to another by concealing certain knowledge or by asserting the obverse of this knowledge. One can also lie to oneself—perhaps, acting in spite of what one knows to be true.

Given this, where lying succeeds is in revealing to the subject the very ambiguities that follow the subject's intertwinement in language, and from which truth bears no logic or relation to the production of facts (Lacan, 1991; Žižek, 2015). That is, in order to lie

‘correctly’ one must always, to a certain extent, be telling the truth—one’s lie must be taken as true in order for the lie to be accepted. On this matter, Zupančič (2017) highlights the inherent asymmetry underpinning the relationship between truth and the lie. In distinguishing between ‘lying with truth’ and ‘telling the truth by means of a lie’, Zupančič (2017, p. 101) explains how a ‘“lying with truth” is nothing but “lying with exactitude,” i.e. lying by uttering something that in itself is correct’. In correspondence, ‘“telling the truth by means of a lie” is nothing but “telling the truth by means of falsehood”’ (Zupančič, 2017, p. 101).

We can consider this strange asymmetry between truth and lie—where the lie is nothing more than what is correct, and where truth can be accessed via falsity—in relation to the paradoxical assumptions that underpin the significance of the counterfactual premise. In examples of counterfactual history (known also as alternate history or ‘what-if’ history), we engage in a genre of historiography that explores hypothetical scenarios by imagining how history might have unfolded otherwise if certain key events had happened differently or not at all (Black, 2019). While much of this alternative history involves speculating about the consequences of changing specific historical events, and then constructing narratives or analyses based on these hypothetical scenarios (not least the importance of raising questions about causality, contingency, and historical determinism), what it reveals is that ‘it is only from a counterfactual premise that we can grasp the truth of the factual’ (Žižek, 2016, pp. 298-299). Ultimately, what is ‘true’ is itself accessed via the falsity of a premise; a truth that is brought to light via the means of a lie. While counterfactual premises can help render what is true, it is important to note that they are not simply wild speculations, chosen at random, and open to any conjecture or supposition; rather, what proves significant to obtaining the ‘right’ counterfactual is ‘select[ing] the right lie, a lie which eventually can enable us to arrive at the truth’ (Žižek, 2016, p. 299). This reveals that ‘if we want to go directly for truth, we lose the truth itself’, and, moreover, what we lose is the importance of the ‘symbolic

dimension' that characterizes the very lies inherent to our social relations (Žižek, 2016, p. 299).

Here, we can think of a variety of examples where our ability to 'politely' lie requires one 'to keep up appearances', which ultimately function to 'accept and proffer widespread deceit' (McGowan, 2004, p. 151). Indeed, 'polite ways of faking it' suggest 'illusionary ways of acting as if one were nice, charming, modest, tolerant, etc' (Ruda, 2021, p. 24), so much so that while 'these are just illusionary practices and although everyone knows this, they are nonetheless effective' (Ruda, 2021, p. 24). When considered in all honesty, we may, for example, have no interest in our colleague's weekend, yet inevitably ask, come Monday morning, "How was your weekend?". Similarly, we accomplish no enjoyment in asking the same colleague how their vacation went, and while we inevitably refrain from telling our friends what we think of their chosen partner, we almost certainly go along with the unescapable, 'invite from the in-laws'… all of which is completed in order to maintain a certain level of 'politeness'. We do so because:

> The continued existence of the social bond depends on such deceits, and without them, the bond would shatter. In fact, the social bond is itself deceit par excellence. The social bond exists only because we collectively believe that it does, and yet it exists with the pretense of being substantive. This lie at the heart of the social bond is the fundamental constitutive lie, the basis for all of the polite, social lies that follow from it. (McGowan, 2004, p. 151).[4]

Where AI errs is in missing the sincere affectivity of such lies. While AI is certainly performing a sincerity when it honestly proclaims to have no opinion on particular matters, what it fails to achieve is the minimal level of reflexivity that the subject achieves when it

partakes in the constitutive lie. The AI's lie is not one that necessarily undermines the social order, but, rather, further obscures and obfuscates the deceit at its heart; thus, eroding the reflexive nuances and complexities that structure our social interactions.

What is lost in the AI response, therefore, is the failure that enacts the honest declaration. Say, for example, one tries to express their love to their long-term partner: importantly, it is only when one fails to express this love in any clealry defined and explicit manner that one's 'true love' is confirmed. If such expressions of love are perfectly declared, then, 'it is not love but [instead] a flat mechanic expression' (Žižek, 2024b, p. 75, parenthesis removed). What posits the subject in such examples is that it is the failure—the 'obstacle' itself—that provides the positive assertion; or the 'truth' (Žižek, 2024b).

To lie in the guise of truth is thus reflected in the honest admissions that the chatbot expels. The candid admittance, which suggests the chatbot has no opinion, belies the very subjective position from which it is enacted—the fact that the statement (the truth) is made from an artificial technology 'pretending' to be a human. While such admissions can function as a way of masking one's guilt and absolving one of an inherent lie (Black, 2020), the chatbot nonetheless presents itself as a 'subjectless' phenomena grounded in the knowledge that it does not possess the capacity to profess an 'honest' opinion. The position that this nonetheless reveals is one akin to the guilty confession: a 'safe position of the excluded observer who knows the relativity and limitation of all human knowledge, including his own' (Žižek, 2001, p. 15). Where this 'beautiful soul' makes its judgement is from the objective 'safety' of the excluded, external position, which bears no reflection on the position it holds (Black, 2020; Žižek, 2014). The crux of this lie is that it is not simply the case that it deceives the other, but that it effectively deceives the subject itself (Ruda, 2021). The 'lie' at play is one that obscures the subject's 'distinct[ion] from what he says' (Lacan, 1991).

'I am lying': Lacan's Enunciation and Enunciated

It is in the distinction between what the subject says and the position from which it is said that Lacan (2004) refers to the enunciation and enunciated. In other words, there is, for Lacan, an 'I' of the enunciation and an 'I' of the enunciated, both of which can be reasoned via the assertion, 'I am lying'. A classic paradox, which arises in the context of self-reference and truth, the statement, 'I am lying', suggests that if the statement is true, then what is stated must be the case, meaning that the statement is indeed false. However, if the statement is false, then what is stated is not the case, and thus the statement is true. This paradox materialises when the subject expressing the sentence (the subject of enunciation) and the subject of the statement (the enunciated subject) contradict. That is, the 'I' who speaks (enunciation) differs to the 'I' of the statement (enunciated). Accordingly, though 'It is quite clear that the *I am lying*, despite its paradox, is perfectly valid', it is nonetheless apparent that the assertion requires an irresolvable distinction: 'the *I* of the enunciation is not the same as the *I* of the statement' (Lacan, 2004, p. 139). Ultimately, what remains essential is that 'If the liar is equal to his task, he can never say "I am lying" (because he would be telling the truth, etc.)' (Zupančič, 2011, p. 102). It is for this reason that the subject is irreducibly split between the predication of the statement (the enunciated) and the position from which it is said (the enunciation).

Thus, to achieve the level of reflexivity afforded by the symbolic order and 'the regime of signification', one must be 'capable of transmitting meaning, but not capable of coinciding precisely with one's meaning' (Rothernberg, 2010, p. 43). Unlike the AI, 'This gap marks the locus of the minimal difference that keeps the subject from coinciding with itself' (Rothenberg, 2010, p. 43). In effect, whenever the subject speaks, their enunciated content remains beholden to the reality of the symbolic order, one that sits aside the reality of the subject itself (Žižek, 2024a). Insofar as 'the subject [is] determined by language and speech, it follows that the subject, *in initio*, begins in the locus of the Other' (Lacan, 2004, p. 198). It

is in this sense that when the subject speaks, their position of enunciation remains excluded from the content of their speech (and, thus, from their own inherent intentions and subjectivity).

When the chatbot admits its failings, with no acknowledgment of the 'but still…', there is essentially no decentrement that characterises its response. If we remember that Lacan's account of the subject is one predicated on a critique of a readily ratified self-consciousness, one achieved via some form of self-appraisal, then it becomes apparent that, for Lacan, 'self-consciousness as such is literally decentred' (Žižek, 1993, p. 206, italics removed). This is not to suggest that one must fall foul of some postmodern conception of the subject as inherently predicated on an elusive self-consciousness, which can never be defined or located, but, rather, proffers the paradoxical appreciation that it is *in* the failure of self-consciousness that one's self-consciousness is averred. Here, the 'decentred hard kernel that eludes my grasp is ultimately self-consciousness itself' (Žižek, 1993, p. 206).[5]

In complete contrast, the chatbot remains predicated on: 'I am where I think' (Žižek, 1993, p. 206). It elicits no decentrement that functions primarily to elude its grasp—it remains, instead, in complete acknowledgement of its response. Bearing in mind that both the study and creation of AI technologies straddles a number of scientific disciplines (including computer science, mathematics, and neuroscience), it is useful here to remain cognisant of Fink's (1997, p. 140) assertion that the subject which 'science … deal[s] with … is only the conscious Cartesian subject, master of its own thoughts, whose thought is correlative to its being'. For this reason, 'existing sciences certainly do not take into account the split subject for whom "I *am* where I am not thinking" and "I think where I am not"' (Fink, 1997, p. 140). As a consequence, where 'The subject cannot make the subject of which it speaks ("I am a woman") coincide with the subject which is speaking ("[Here I am saying that]…")' (Rothenberg, 2010, p. 43), we nonetheless manage this irreducible conundrum via a certain

ambivalence, best expressed in popular formulations, such as, Octave Mannoni's, 'I know well, but all the same…' (Mannoni, 2003).[6] When the 'but all the same…' affords a reversal of the preceding, 'I know'—indeed, a contradiction inherent to the 'I' of the statement—we are thus ratified with the subject of enunciation, or, as Lacan (2006, p. 707) asserts, with 'the presence of the unconscious'.

Lying like a Subject

What this leaves us with is the now relevant contention that 'It takes the unconscious to lie' (Castrillon, 2023); or, at least, to lie like a subject. Here, Castrillon (2023) elaborates that while 'Generative AI may blur the line between a lie and the always fraught effort at truth-speaking, … at least for now, lying is solely the reserve of humans'. Left only to humans, we return once more to the significance of the social bond, from which our very solidarity is so often predicated on a shared sense of lying or disavowal. In 'the case of the "leader caught with his pants down"', it is 'the solidarity of the group [which] is strengthened by the subjects' common disavowal of the misfortune that laid open the leader's failure or impotence' (Žižek, 1999b, p. 99). Indeed, 'this very knowledge of the flaw—together with the willingness to disavow this knowledge—is the true nature of identification which keeps the group together' (Žižek, 1999b, p. 99). The consequence of this disavowal is that it does not necessarily need to be accepted, but, rather, can help to posit 'the starting point of the critique of ideology'; one that proclaims the 'full acknowledgement … that it is easily possible to *lie in the guise of truth*' (Žižek, 1999a, p. 61).

More importantly, what this more readily reveals is that it is not our 'better knowledge' that distinguishes us from the AI chatbot, but that it is in spite of our knowledge, and in full view of what it is that we are doing, that the subject is both forged and enveloped *in* the capacity to lie. Instead, where 'truth is an effect of surprise triggered by its enunciation' (Žižek, 2024b, p. 178), for the chatbot, there is no truth to its enunciation, no gap, failure, or

surprise, for which the truth of its position can be averred. As previously noted, we can now assert that 'in all its stupidity' the chatbot 'is *not stupid enough to*' fall foul to the truth of its mistakes (Žižek, 2024b, p. 178).

It is in this sense that the AI's honesty overlooks the social nuances underpinning the importance of the lie and its relation to underlying infrastructures that underwrite our AI applications, as noted earlier by Crawford (2021): specifically, are these social nuances deliberately overlooked in the programming of AI? This question points towards the consideration that efforts to establish truthful responses from AI, such as that seen in the case of ChatGPT, requires an infrastructural reliance on curated datasets and human feedback. Through quality assurance and testing, to data cleaning and labelling, the creation, maintenance, and ethical deployment of AI systems relies heavily on human labour, encouraging labour practices that are frequently outsourced to the Global South, where labour conditions remain poor and pay low (Taylor, 2023). The potential exploitation that is encountered in the invisibility of this labour force, including the probable exacerbation of unequal labour conditions, points to the fact that the ethics behind our use of AI are regularly ignored or overlooked in the authority that is attributed to AI technology, as well as the developments and promises it affords (Crawford, 2021). As demonstrated by the chatbot's candid admissions and straightforward assertions, when left with the open acknowledgement of the AI's ignorance, we unquestionably accept the responses we are provided. When Žižek (2024b, p. 178, italics removed) shares his concern 'that communicating with chatbots will make "real" persons talk like chatbots', we can now add that it is by losing our capacity to lie like subjects that such concern resides: where certain modes of communication, devoid of the richness and subtlety inherent to human interaction, are simply lost or undermined.

While it is, under present circumstances, widely acknowledged that chatbots continue to lack the intuitive understanding of social contexts that humans possess, what AI chatbots

reveal is that lying is not simply about deception: it is, instead, a very human tool for navigating complex social dynamics, maintaining relationships, and preserving social harmony. Ultimately, when faced with the straightforward responses of the chatbot, we are confronted with its ignorance of these social nuances. Whereas the honesty and ongoing improvements in AI technology may offer certain advantages in terms of accuracy and efficiency, such embracement belies certain uniquely human qualities—namely, our capacity to lie.

Notes

[1] For this reason, the general unease we currently experience, such as, the possibility of machines achieving consciousness, can actually be rather stimulating, intriguing, and enjoyable (Everitt, 2023).

[2] Žižek (2002, xiii) explains, 'Language, in its very notion, involves a minimal distance towards its literal meaning—not in the sense of some irreducible ambiguity or multiple dispersion of meanings, but in the more precise sense of "he said X, but what if he *really meant* the opposite"'.

[3] Importantly, Lacan (2004) situates this quote in light of the subject's relation to the Other and the signifier, whereupon the subject finds its place in accordance with the Other. Ultimately, as a signifier, the subject can only ever makes sense, exist even, via its relation to the Other.

[4] This is reflected in Lacan's (2004, p. 144) assertion that 'it is in the locus of the Other that [… the subject] begins to constitute that truthful lie'.

[5] Žižek (1993, p. 208) adds, 'As to its status, self-consciousness is an external object out of my reach'.

[6] Moreover, it reveals a certain distortion of truth that works in concert to the various forms of disavowal that shape our day-to-day concerns and not least are political predicaments.

References

Black, J. (2019). 'You Ain't Gonna Get Away Wit' This, Django': Fantasy, Fiction and Subversion in Quentin Tarantino's, *Django Unchained*. *Quarterly Review of Film and Video*, *36*(7): 611—637.

Black, J. (2020). On Reflexive Racism: Disavowal, Deferment, and the Lacanian Subject. *Diacritics*, *48*(4): 76—101.

Black, J. (2023). The dialectic of desire: AI chatbots and the desire not to know. *Psychoanalysis, Culture & Society*, *28*(4): 607—618.

Bove, A. (2021). What Happens When the Replicants Become Extimate? On the Uncanny Cut of the Capitalocene in Blade Runner 2049. In C. Neill (Ed.), *Lacanian Perspectives on Blade Runner 2049* (pp. 139—166). Palgrave Macmillan.

Castrillón, F. (2023). Generative 'Artificial Intelligence' & Psychoanalytic Writing: An Editorial Note. *European Journal of Psychoanalysis*, *10*(1). https://www.journal-psychoanalysis.eu/articles/generative-artificial-intelligencepsychoanalytic-writing-a-short-editorial-note/

Chandran, R., Smith, A., & Ramos, M. (2023). AI boom is dream and nightmare for workers in Global South. *Context*. https://www.context.news/ai/ai-boom-is-dream-and-nightmare-for-workers-in-global-south

Crawford, K. (2021). *Atlas of AI*. Yale University Press.

D'Amato, K. (2024). ChatGPT: towards AI subjectivity. *AI & Society*. DOI: https://doi.org/10.1007/s00146-024-01898-z

Davis, J.E. (2019). Toward the Elimination of Subjectivity: From Francis Bacon to AI. *Social Research: An International Quarterly, 84*(4), 845—869.

Dews, P. (1995). *The Limits of Disenchantment*. Verso.

Everitt, K. (2023). Does ChatGPT Enjoy? *The Philosophical Salon*. https://thephilosophicalsalon.com/does-chatgpt-enjoy/

Ferrara, E. (2023). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci*, *6*(1): 1—15.

Fink, B. (1997). *The Lacanian Subject*. Princeton University Press.

Fink, B. (1999). *A Clinical Introduction to Lacanian Psychoanalysis*. Harvard University Press.

Fisher, E. (2023). AI, critical knowledge and subjectivity. In S. Lindgreen (Ed.), *Handbook of Critical Studies of Artificial Intelligence* (pp. 94—107). Edward Elgar.

Flisfeder, M. (2021). *Algorithmic Desire*. Northwestern University Press.

Flisfeder, M. (2022). Whither Symbolic Efficiency? Social Media, New Structuralism, and Algorithmic Desire. *Rethinking Marxism*, *34*(3), 413—432.

Jupe, L.M., & Keatley, D.A. (2020). Airport artificial intelligence can detect deception: or am I lying? *Security Journal*, *33*: 622—635.

Johanssen, J., & Krüger, S. (2022). *Media and Psychoanalysis: A Critical Introduction*. Karnac.

Kneer, M. (2021). Can a Robot Lie? Exploring the Folk Concept of Lying as Applied to Artificial Agents. *Cognitive Science*, *45*(1): 1—15.

Lacan, J. (1991). *Book I: Freud's Papers on Technique, 1953-1954*, translated by John Forrester. W. W. Norton & Company.

Lacan, J. (2004). *The Four Fundamental Concepts of Psycho-Analysis*. Karnac.

Lacan, J. (2006). Position of the Unconscious. In *Ecrits*, translated by Bruce Fink (pp. 703—721). W. W. Norton & Company.

Magee, L., Arora, V., & Munn, L. (2023). Structured like a language model: Analysing AI as an automated subject. *Big Data & Society*. DOI: 10.1177/20539517231210273.

Mallory, F. (2023). Fictionalism about Chatbots. *Ergo*, *10*(38): 1082—1100.

Mannoni, O. (2003). 'I Know Well, but All the Same…'. In M.A. Rothenberg, D. Foster, & S. Žižek (Eds.), *Perversion and the Social Relation* (pp. 68—92). Duke University Press.

McGowan, T. (2004). *The End of Dissatisfaction?* State University of New York Press.

Millar, I. (2021). *The Psychoanalysis of Artificial Intelligence*. Palgrave Macmillan.

Mohamed, S., Png, M.T., Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, *33*: 659—684.

Murphy, M.G. (2023). ChatGPT: A New Unconscious? *Sublation Magazine*. https://sublationmedia.com/chatgpt-a-new-unconscious/

Oravec, J.A. (2022). The emergence of 'truth machines'?: Artificial intelligence approaches to lie detection. *Ethics and Information Technology*, *24*(6): 1—10.

Ozdemir, S. (2024). *Quick Start Guide to Large Language Models*. Pearson Education.

Pakhale, K. (2023). Large Language Models and Information Retrieval. *SSRN*. DOI: http://dx.doi.org/10.2139/ssrn.4636121

Radanliev, P. (2024). Artificial intelligence: reflecting on the past and looking towards the next paradigm shift. *Journal of Experimental & Theoretical Artificial Intelligence*. DOI: 10.1080/0952813X.2024.2323042.

Rothenberg, M.A. (2010). *The Excessive Subject*. Polity.

Ruda, F. (2021). The Impossible InSight. *Coils of the Serpent*, *8*, 23—33.

Sharma, M., & Kaur, M. (2022). A Review of Deepfake Technology: An Emerging AI Threat. In G. Ranganathan, X. Fernando, F. Shi, & Y. El Allioui (Eds.), *Soft Computing for Security Applications* (pp. 605—619). Springer.

Taylor, B.L. (2023). Long hours and low wages: the human labour powering AI's development. *The Conversation*. https://theconversation.com/long-hours-and-low-wages-the-human-labour-powering-ais-development-217038

Turner, J. (2019). *Robot Rules: Regulating Artificial Intelligence*. Palgrave Macmillan.

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, *9*(11): 39—52.

Williams, A., Miceli, M., & Gebru, T. (2022). The Exploited Labor Behind Artificial Intelligence. *Noema*. https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence/

Žižek, S. (1993). 'The Thing That Thinks': The Kantian Background of the *Noir* Subject. In J. Copjec (Ed.), *Shades of Noir* (pp. 199—226). Verso.

Žižek, S. (1999a). The Spectre of Ideology. In E. Wright & E. Wright (Eds.), *The Žižek Reader* (pp. 53—87). Blackwell.

Žižek, S. (1999b). Fantasy as a Political Category: A Lacanian Approach. In E. Wright & E. Wright (Eds.), *The Žižek Reader* (pp. 97—101). Blackwell.

Žižek, S. (2001). *The Fright of Real Tears*. BFI.

Žižek, S. (2002). *For They Know Not What They Do*. Verso.

Žižek, S. (2015). *Absolute Recoil*. Verso.

Žižek, S. (2016). *Disparities*. Bloomsbury.

Žižek, S. (2023a). *Freedom*. Bloomsbury.

Žižek, S. (2023b). ChatGPT says what our unconscious radically represses. *Time.News*. https://time.news/chatgpt-says-what-our-unconscious-radically-represses/

Žižek, S. (2023c). Artificial Idiocy. *Project Syndicate*. https://www.project-syndicate.org/commentary/ai-chatbots-naive-idiots-no-sense-of-irony-by-slavoj-Žižek-2023-03

Žižek, S. (2024a). Why the AI Revolution May Wind Up Killing Capitalism. *World Crunch*. https://worldcrunch.com/opinion-analysis/ai-and-capitalism

Žižek, S. (2024b). *Christian Atheism*. Bloomsbury.

Zupančič, A. (2011). *Ethics of the Real*. Verso.

Zupančič, A. (2017). Lying on the Couch. *Problemi International*, *1*(1), 99—115.