

# Critical Humanities

---

Volume 2  
Issue 1 *Humanities in the time of ChatGPT and  
other forms of AI*

Article 4

---

December 2023

## Guilty Machines: On Ab-sens in the Age of AI

Dylan Lackey  
*Virginia Commonwealth University*

Katherine Weinschenk

Follow this and additional works at: <https://mds.marshall.edu/criticalhumanities>



Part of the [African American Studies Commons](#), [Artificial Intelligence and Robotics Commons](#), [Computational Linguistics Commons](#), [Theory and Algorithms Commons](#), and the [Theory and Philosophy Commons](#)

---

### Recommended Citation

Lackey, Dylan, and Katherine Weinschenk. "Guilty Machines: On Ab-sens in the Age of AI." *Critical Humanities* 2, 1 (2023). DOI: <https://doi.org/10.33470/2836-3140.1038>

This Article is brought to you for free and open access by Marshall Digital Scholar. It has been accepted for inclusion in Critical Humanities by an authorized editor of Marshall Digital Scholar. For more information, please contact [beachgr@marshall.edu](mailto:beachgr@marshall.edu).

# Guilty Machines: On *Ab-sens* in the Age of AI

**Dylan Lackey**

Virginia Commonwealth University

**Katherine Weinschenk**

Independent Researcher

## Abstract

For Lacan, guilt arises in the sublimation of *ab-sens* (ab-sense) into the symbolic comprehension of *sen-absexe* (sense without sex; sense in the deficiency of sexual relation), or in the maturation of language to sensibility through the effacement of 'sex.' While, as Slavoj Žižek also points out in a 2023 article regarding ChatGPT, the split subject always misapprehends the true reason for guilt's manifestation, such guilt at best provides a sort of evidence for the inclusion of the subject in the order of language, thereby acting as a necessary, even enjoyable mark of the subject's coherence (or, more importantly, the subject's division from incoherence/*ab-sens*). For Žižek, the perversity (*père-versity*) of artificially intelligent chatbots lies precisely here, in their appearance as evidently novel modes for enjoying the displacement of one's guilt onto the intelligent machine ("what happens is a form of perverse disavowal: knowing full well that it was the machine, not me, that did the work, I can enjoy it as my own," Žižek 2023). What Žižek does not elaborate, however, is how the transferred belovedness of guilt is a figure of contemporary life in general—a condition for modernity's endless reproduction—and the AI chatbot is but one more recent, particularly popular, indication of racial capital's long entanglement with the unconscious. In this work, the relationship between guilty affects, transference, cultural reproduction, *ab-sens*, and artificial intelligence is discussed using a reference to Lacan's later works and seminars, critical data science studies, and Black radical criticism.

**Dylan Lackey** is a PhD student in the Media, Art, and Text Program at Virginia Commonwealth University. They make use of Lacanian psychoanalysis, critical theory, aesthetics, and philosophical pessimism to interpret the writings of Louis Althusser. Their writing can be found in *Chiasma Journal*, *InVisible Culture*, and elsewhere.

**Katherine Weinschenk** is a data science technical writer with a background in computer science, anthropology, and philosophy. They are interested in data ethics, critical phenomenology, the Internet of Things, and surveillance studies. They currently work and conduct relevant independent research in Richmond, Virginia.

## Introduction

Guilt—which began as a relatively simple concept in the work of Sigmund Freud and was applied as a way to give name to the largely unconscious affects arising from supposedly abnormal or even reprehensible behavior<sup>1</sup>—eventually, in the writings of Freud’s closest follower, Jacques Lacan, came to signify the unapproachable feeling which appears in the acts of sublimation, positivization, or castration indicative of a subject’s becoming in (or coming to) language. Despite the fact that this claim by Lacan appears, in one sense, of striking import to the whole object of psychoanalytic theory, attendance to this core matter of guilt in the production of (split) subjectivity remains lacking in contemporary research. The closest one might come to a contemporary turn towards Lacan’s necessary alteration of Freudian guilt is in the queer theorist and psychoanalytic thinker Lee Edelman’s most recent book, *Bad Education*. In *Bad Education*, Edelman dedicates overdue attention to dissecting the profound, difficult, and oft-repeated dictum by Lacan: “the only thing of which one can be guilty is of having given ground relative to one’s desire.”<sup>2</sup> Edelman explains:

Incompatible with the good to which the subject clings in [the] economy of desire... “radical desire” engenders the guilt experienced by those who betray it: “From an

analytical point of view, the only thing of which one can be guilty is of having given ground relative to one’s desire.’ Such giving ground, as Lacan points out, is often done “for good motives,” indeed, for the sake of the “good” itself, which demands that we “sacrifice” our “radical desire,” the “good” that is *jouissance*. Conforming to the social good, therefore, as Lacan adverts us, is “far from protecting us not only from guilt but also from all kinds of inner catastrophes” that follow from being “driven by the idea of the good.”<sup>3</sup>

By approximating guilt to the process of capitulating one’s access to pure difference, to *jouissance*, Lacan also makes clear that the affect that is guilt appears without an object. Put differently, one experiences guilt as a matter of experiencing speech, and in that sense one knows little to nothing about why one is guilty from the start—the guilt is simply there, directly alongside the social good that one has accepted in the Real’s stead. At the same time, guilt’s constant presence in the symbolic, its endless there-ness without a (clear) object, is the source of genuine grief for the subject. In order to respond to this presence—this “intensity” as Slavoj Žižek has noted in his response to Mark G. Murphy’s work on ChatGPT as a ‘new unconscious’<sup>4</sup>—guilt becomes the means by which the subject establishes and repeats the ideological conditions of their own subjectivation.

---

<sup>1</sup> Freud, “Obsessive Actions and Religious Practices,”

*Readings in Ritual Studies* (e. RL Grimes), 215

<sup>2</sup> Lacan, *The Ethics of Psychoanalysis: The Seminar of Jacques Lacan, Book VII, 1959–1960* (trans. D. Porter), 319

---

<sup>3</sup> Edelman, *Bad Education*, 136-137

<sup>4</sup> Žižek, “ChatGPT Says What Our Unconscious Radically Represses,” online

So as to retain the order of the social good as it is promised in the subject's becoming speaking-being, the subject's always-already decided decision against *jouissance* which makes him guilty from the jump, the subject defers guilt by affirming fidelity to the good that is social (even if, as Lacan clarifies, this often leads to the least expected of inner catastrophes). In a way no less circular than it appears, then, the subject's guilt is a sign of its split into speech and symbolism at the same time that it is the very promise of the subject's coherence—its symbolic meaning—or, more importantly, its absolute division from *jouissance*, radical desire, the signifier and so on.

While for Edelman guilt appears as the mode of cultural and social (re)production meant to be traversed so as to enter a Real, necessarily queer field of pure *jouissance*, guilt's avoidability is a questionable enterprise. There is a sense in which the history of language—the history of speaking being—is a history of guilt. If this is the case, language's sublimation from the *jouissance* of *lalangue*—a babbling, deciphering speech akin to the lallation of the mother, the mother's tongue, the infant's non-differentiability—is constantly a matter of guilt, and the quasi- (or Joycean) psychosis of *lalangue* constantly a matter of a strikingly black difference which not even (non-Black) queerness can traverse without aversion to the anti-Black. This also means, in a deeply challenging way, that all guilt, as a symptom that reproduces itself and its own conditions, is in one way or another white, and it is guilt's incontroverti-

ble whiteness which in turn retains language's obscured anti-Blackness.

What has this to do with this so-called age of artificial intelligence? While for Žižek the language learning phenomenon that is ChatGPT grants a heretofore unavailable means of perversely deferring confrontation with our fundamental guilt by allowing us to enjoy the variously grotesque or obscene fruits borne by the bot as if they were our own,<sup>5</sup> we argue that large language models (LLMs) are instead simply novel iterations of the same objects variously constructed in the world of speaking-being as suppositories for a guilt that is white. Where these particular models stand apart, however, is in the fact that the various 'hallucinations' by which their strictly-directional and hegemonically ordered origins are exposed determine them as uniquely capable of confronting us to our guilt in its very whiteness, in its obstinate white structure.

### *Ab-sens* and Guilt

In 'L'Étourdit,' Lacan's iconically challenging essay, the psychoanalyst clarifies meaning's coherence as strictly a matter of the sexual relation's complete effacement, its allocation to the non-realm of *ab-sens* (neither sense nor non-sense; *ab-sense*). Forming the symbolic-imaginary field as *sens-absexe* (*sense* in the dissolution of sex), *ab-sens*, sex, is rendered quasi-ahistorical; *ab-sens* is to *sense* what the zero, for Pascal, is to mathematics: a paradox, a lure, a shroud pulled over

<sup>5</sup> Žižek, "ChatGPT Says What Our Unconscious Radically Represses," online

nothingness. Elsewhere, and in a similar but distinct tune, Lacan elaborates a related, aforementioned term: *lalangue*. *Lalangue*, being as it is the mother tongue divided from the tongue that is language, might best be described as featuring within it the full terrain of sense (*sens-absexe*), prior to its interpretation, as well as all of *ab-sens*, or what in *langue* (language, tongue) can be attributed to sex or the sexual relation, even while this attribution remains nothing more than a catachresis (as in, when one talks about sex in language, when one talks about drive or *jouissance*, one does so only by giving name to something that always evades its own representation). Finally we conclude that *ab-sens* is the detritus of *langue* which is itself the detritus (the “cemetery<sup>6</sup>”) of *lalangue*.

What remains of significance to this discussion of artificially intelligent machines is the important matter that *lalangue*—the impossible house for *langue* and its always-already ab-sented relation (themselves the grounds on which every language learning model achieves its initial structuration)—permanently reveals, per David Marriott, “a [Black] *n’est pas* (is not) that is illegible as subject or object, and one that cannot be found in the operations of speaking being or its logic.”<sup>7</sup> Unlike the speaking subject who speaks through her sublimation-castration and on the eternal condition of her objectless guilt, the Black *n’est pas* revealed by the persistence of *lalangue* is “destined to be guilty for not

---

<sup>6</sup> Soler, *Lacan—The Unconscious Revisited*, 27

<sup>7</sup> Marriott, “Ontology and *Lalangue* (Or, Blackness and Language),” 236

being,”<sup>8</sup> or for ‘being’ the Black not-being from which the guilty speaking subject divides itself so as to found its (ideological, reproductive, meaningful) guilt.

## Technologies of Order

As a matter of being the language product of a bleached white *lalangue*, today’s language learning model reveals to us, firstly, guilt as a properly circulatory mode and product of sublimation, and secondly, guilt as the affective result of whiteness in language. To get a better understanding of how whiteness plays out in algorithmic knowledge, it is essential to examine what large language models are at present, as well as how they are expected to evolve over time.

At the forefront of applied LLMs in public discourse and usage are ChatGPT and a cohort of related chatbots. Beyond these, other LLM use cases—less popular in demand and more professional in their employment—include text summarization, translation, and sentiment analysis, amongst others. Nonetheless, an emphasis placed on both the dialogic nature of ChatGPT and its mass appeal, in particular, gives way to a more apt attendance to the questions of self and other, lack and perversion, at stake in these pages.

While the significant range of LLM use cases remain powered by adjacent means and are built in accordance with similar technical conditions, chatbots function within a specific set of adaptive language rules which can, to an extent, be uniquely

---

<sup>8</sup> Marriott, *Lacan Noir*, 94

experimented with and theorized about. On the technical level, GPT models, or Generative Pre-trained Transformers, employ transformer architecture that enables models to ‘learn’ about context and meaning through encoding and decoding. This means that input words are converted into mathematical representations that contain information about word relevance and relationships within a sentence as well as word order via position encoders. In response, the model returns a statistically plausible result as output. Or, put differently, the model understands and generates meanings restricted to those of the data on which it was trained; the model’s output is restrained only to replication of the natural language of its input.

Natural language and computer language cannot, therefore, be separated; each employs syntax and semantics; each is, in varying capacities, a technology of order and structure.<sup>9</sup> It is not sufficient to say that computer (or programming) languages are unequivocal or objective in contrast to contextual or subjective natural language com-

<sup>9</sup> “Unlike the Symbolic, *lalangue* is thus not a constituted body but a multiplicity of differences that have not taken shape. There is no (-1) of *lalangue* that would make it a set. There is no order in *lalangue*. It is not a structure of language or of discourse. For language, order is the ordered pair of the subject that inscribes it  $S1 \rightarrow S2$ . It is the basis of the transference as a link to the subject supposed to know and it also structures free association and all its effects of meaning. For discourse, order is the semblant... which orders the social bond. Every discourse is thus an order. This is not the case with *lalangue*, which is the a-structural level of the verbal apparatus.” Soler, *Lacan—The Unconscious Revisited*, 26

munication. In reality, computer language cannot be severed from hermeneutics—ChatGPT and its tributaries show this.

While natural language processing, now evolved into large language modeling, evaluates and simulates written human communication, it cannot yet be considered analogous to human cognition, and in spite of contemporary popular (mis)representations of neural networks in public discourse. In actuality, the psychoanalytic reinterpretation of structural linguistics—with its attendance to slippages, half-truths, and human error—strikes closer to the real heart of natural language processing than do more common (often fear-driven) misapprehensions about modeling’s proximity to self-awareness or technical independence. In fact, Franco “Bifo” Berardi, in his reflections on automation and artificial intelligence, clarifies precisely the absurdity of dividing the technical object from its human origin: “The automaton has an ordering mission, but it encounters a factor of chaos along the way: the organic drive, irreducible to numerical order.”<sup>10</sup> This encounter, one could say, is the encounter with the Real, or with the absent that disobeys its effacement, arising as a machinic hallucination that simultaneously disrupts and reveals the (human) guilt amplified in algorithmic structure by exposing, retroactively, the strictly political biases that haunt it.

This ordering, or application of the law in technology, is an ongoing process,

<sup>10</sup> Berardi, “The Completion,” online

especially in the emergent field of Generative AI. Yet, it is easy to forget the work that goes into producing these LLMs. Not only do LLMs require a significant amount of concrete resources (hardware, energy, etc), but they also demand everlasting optimization, demonstrated by the proliferation of specialists such as prompt engineers<sup>11</sup> as well as a rapidly growing body of research explaining the process of fine-tuning, altering, and supporting pre-existing large language models.

Consideration of the Conference on Neural Information Processing Systems proceedings of 2022 allows a closer look at what is of significant importance for scholars within the large language modeling research field and signals the expected future for LLMs and their construction. Topics presented during this significant event included (among other themes) improving complex reasoning skills of LLMs,<sup>12</sup> capturing failures of LLMs,<sup>13</sup> and fine-tuning language models<sup>14</sup>. Honing in on this last paper, Bakker et al. immediately raise that “recent work in large language modeling (LLMs) has used fine-tuning to align outputs with the preferences of a prototypical user.”<sup>15</sup> They continue to expand that their research goal is to “highlight the potential to use LLMs to help

---

<sup>11</sup> Grant, “What is Prompt Engineering,” online

<sup>12</sup> Wei, et al., “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models,” online

<sup>13</sup> Jones and Steinhardt, “Capturing Failures of Large Language Models via Human Cognitive Biases,” online

<sup>14</sup> Bakker, et al., “Fine-tuning language models to find agreement among humans with diverse preferences,” online

<sup>15</sup> Ibid.

groups of humans align their values with one another.”<sup>16</sup> So, in addition to revealing the import of improvement in the AI milieu, it highlights the flattening effect—the controlling effect—that engineers are attempting to deploy.

While this is happening in research and development, the general audience of LLMs like ChatGPT think less about the designers of these deep learning algorithms and more about the enchanting program itself as a separate, autonomous entity. Perhaps the general perception of something like ChatGPT being an independent subject is due to its success at parroting human communication; or, it could also be the grander work of technological fetishism (and our alienation from production). However, recognizing on the backend this iterative injection of law and sublimation is imperative to understanding how LLMs and GenerativeAI function (and will function) in our world. Indeed, it has been noted for one chatbot that this “can be related to the function of the law of the symbolic father, or more directly, of the socio-economic system that funds and coordinates the operations of InstructGPT.”<sup>17</sup>

If there was a world in which its inhabitants would ‘let AI loose’ in some organic way (where the possibility itself is debatable), an argument might be made that a certain representation of the unconscious would eventually arrive as the result of LLM outputs. Still so, considering the work being done to fine-tune and govern these

<sup>16</sup> Ibid.

<sup>17</sup> Magee, et at., “Structured Like a Language Model: Analysing AI as an Automated Subject,” online

technologies as mentioned above, it would be speculatively unrealistic to expect that a commercial chatbot would organically, without significant interference, produce something aberrant. In other words, it is unlikely that a commercial chatbot would raise, produce, or represent what is radically repressed, or spectacularly perverted. For instance, GPT-4 (the model behind ChatGPT) has been heavily reared, as OpenAI reports: “We’ve been iterating on GPT-4 to make it safer and more aligned from the beginning of training, with efforts including selection and filtering of the pre-training data, evaluations and expert engagement, model safety improvements, and monitoring and enforcement.”<sup>18</sup> That said, there still are instances of ‘hallucinations’ in artificial intelligence which are important to the present discussion. Distinct from the psychoanalytic term of the same name, hallucinations occur when large language models produce nonsensical content that is self-contradicting,<sup>19</sup> unfoundedly biased, and/or factually incorrect. In one study from 2023, Stanford medical researchers found that each LLM tested therein showed “instances of promoting race-based medicine/racist tropes or repeating unsubstantiated claims around race” when prompted with questions regarding pain threshold, skin depth, and organ capacity across various racial identities.<sup>20</sup> While such racist outputs can be (and were) quickly recogni-

<sup>18</sup> OpenAI, “GPT-4,” online

<sup>19</sup> Mündler, et al., “Self-contradictory Hallucinations of Large Language Models: Evaluation, Detection and Mitigation,” online

<sup>20</sup> Jesutofunmi A. Omiye, et al., “Large language models propagate race-based medicine,” online

nized as the results of historical biases—deemed as hallucination and adjusted accordingly—their presence exposes the overdetermining role played by racialization and white hegemony in the reproduction of scientific language and asserts a future expectation for the preemptive eradication of “race-based themes”<sup>21</sup> in the calculation of impending outputs. Simultaneously, the immediacy of this sublimatory desire—in the racist Stanford bots’ prompters as well as in general professional discourse around AI hallucinations—further indicates the unique ability of the machine to expose the presence of racialized and racializing guilt in the heart of social (re)production. One might find that whereas the speaking subject is marked by an untraceable infinitude of contingencies which precondition its every division from ab-sens, sublimation of the computational hallucination is finally realized in discernable mathematical structure and according to human guidance, thereby granting unprecedented access to a guilt that otherwise persists without discernible object.

In essence, then, these hallucinations surface what has been historically, socially, and politically repressed. For instance, in Žižek’s case, these hallucinatory mistakes provide value insofar as they express our perversions by making way for a deliberate deferral of responsibility. For Murphy, to whom Žižek responds, the hallucination remains psychically significant but less a figure of perverse relegation. Murphy writes: “Inasmuch as all of these phenomena can be considered ‘mistakes,’ they won’t be got-

<sup>21</sup> Ibid



ten rid of [altogether]. But they will become more pronounced, creative and surreal, signifying something of the lost dream of work we yearn for.”<sup>22</sup> In understanding the function of LLMs in the world, and in noting their undeniable enmeshment with racial capitalism as an overdetermining economic project, it can be gathered that a major aim of ongoing AI research is technical perfection and therefore the complete elimination of AI hallucinations—or, at very least, their general concealment as such according to a particular logic.<sup>23</sup> While hallucinations, errors, and obscenities may be signs of AI as either unconscious mirror or as plane of perversity in the present, what matters is that these hallucinations are the object of human alteration or of sublimation into meaningful, sensible output. In the following section, a more explicit turn toward AI hallucinations and the problems that they pose for psychoanalysis, ontology, and digital theory will be made.

### Artificial Hallucinations

To reiterate: AI hallucinations denote instances when an artificial intelligence algorithm provides false, unsuitably biased, surreal, or nonsensical information in its response or output to a particular prompt. According to Jonathan Siddharth,

CEO of Turing, “hallucinations happen because LLMs, in their most vanilla form, don’t have an internal state repre-

---

<sup>22</sup> Murphy, “E-scaping Responsibility and Enjoyment Through ChatGPT: A New Unconscious?” online

<sup>23</sup> Research regarding the spectacularization of AI hallucinations is ripe in the ever-growing domain of disinformation studies

ntation of the world... There’s no concept of fact. They’re predicting the next word based on what they’ve seen so far—it’s a statistical estimate.”<sup>24</sup> One notable drive against hallucination comes from Elon Musk, who has, with frequency, announced his anxious anticipation of a TruthGPT, or a “maximum truth-seeking AI.”<sup>25</sup> While outside the scope of this paper, there is of course room to scrutinize the parameters of truth as Musk understands them, and especially if one attends to the sort of ‘half-speak’ or impossible fullness of truth-saying as it appears in the psychoanalytic tradition since Lacan. In any case, while Murphy recognizes these AI mistakes as signs of a machinic unconscious, Žizek reframes them—or only addresses them—as “when a chatbot produces obscene stupidities.”<sup>26</sup> This focus on stupidities and vulgarities lends itself to his overarching argument that these are “machines of perversion”<sup>27</sup> in the psychoanalytic sense: modes of becoming the source of the Other’s pleasure and of disavowing personal responsibility. Still so, these interpretations by both Žizek and Murphy retain the psychoanalytic notion, after Freud and Lacan, that slips of the tongue and other errors in speech fall from a place in language which is largely (racially) neutral or universal, albeit contingent. What is pushed aside in these arguments is essential: hallucinations stem not from an apolitical symbolic field but are instead the

---

<sup>24</sup> Mearian, “What are LLMs, and how are they used in generative AI?” online

<sup>25</sup> Roth, “Elon Musk claims to be working on ‘TruthGPT’ — a ‘maximum truth-seeking AI’” online

<sup>26</sup> Žizek, “ChatGPT Says What Our Unconscious Radically Represses,” online

<sup>27</sup> Ibid

byproducts of a language fundamentally structured both by the ab-sens of what is associated with the radical drive, and from a white language that delimits-defines itself apart from what is black: “the vanishing point that is the differentiating element of difference itself.”<sup>28</sup>

So—if not a type of perversion or obfuscation of the unconscious—what is the substance of AI hallucinations? In actuality, AI hallucinations are confined to the same rules of sublimation as any language output—correct or incorrect, factual or fictitious. There is no differentiation or acknowledgment in the internal workings of a language model of these distinctions, and in fact, they “generate text that sounds fine, grammatically, semantically, but they don’t really have some sort of objective other than just satisfying statistical consistency with the prompt.”<sup>29</sup> The crux of this statement is in its emphasis on satisfying statistical consistency—which is to say that the model attempts consistently to conform to the (racialized and racializing) law by which it was constructed. Hence, these hallucinations are less so the deep, dark, repressed summations of the unconscious, but rather the almost-normal thing that first barely evades and then ultimately certifies its own sublimation. In this way, AI hallucinations may be recognized as the incursion of ab-sens in the field of meaning, but rather than appearing or being conjured as catachresis (as is ‘sex’ for Lacan or ‘queerness’ for Edelman), the AI

hallucination ‘speaks’ a direct statistical output which affects guilt precisely because this output is the re-presentation of a guilt that is unconscious. Put differently: if guilt arises from the sublimation of ab-sens into sens-absexé—from the effacement of jouissance from language and form via meaning making—and is, for that matter, always-already objectless, ahistorical, what the hallucination posits is a confrontation with the forgotten object of radical desire which must immediately be sublimated, positivized, or given meaning.

Now, what does this tell us about language? *Hallucinations are the product of a series of contingencies which can be traced to a fundamentally white coding, the predisposition of a fundamentally white language, the synthesis of fundamentally white guilt.* While Žižek names the obscene hallucination as a means of perverse enjoyment and disavowal, what it actually expresses is a hideous fraction of the guilt that is fundamental to language’s endless reproduction and that fraction’s immediate sublimation into social good. At the same time, this obscene output, being as it is both an object of (white/ed) technical perfection and moral correction, designates a separation from a black *n’est pas* in language whereby this division is not whatsoever perverse—begetting neither disavowal nor instrumentalization—but actually foundational to psychic (well)being.

Regardless of the extent of its obscenity in language, there is a sense in which the hallucination never exceeds its necessarily white prompting. Even as the hallucina-

<sup>28</sup> Marriott, “Ontology and Language (Or, Blackness and Language),” 236

<sup>29</sup> Smith, “Hallucinations Could Blunt ChatGPT’s Success,” online

tion exposes its prompter to an underlying guilt, this guiltiness remains enjoyable on account of the promise that it makes, which is the promise of symbolic coherence: “this language is eternally white; this language is resolutely not Black.” Rather than this being a sign of perversion—of enjoying one’s position as the Father’s instrument which allows for a deferral of primary responsibility—enjoyment of guilt’s object as it is summarily exposed in each AI hallucination is, in an odd way, more akin to transference-love. Whereas Žizek’s perverted/perversion machine would make it possible for a user to disavow her own existence as a split subject in language and thus enjoy the AI’s error as her own based on a certainty of irresponsibility, this ignores that fact that the contemporary AI model is supported by teams of human correctors whose very presence in the model’s coding makes possible, at the moment of (obscene) hallucination, not disavowal but transference.

By exposing what is ab-sent in language, the vulgar or socially inappropriate AI error simultaneously exposes the AI itself as knowledgeable of a guilt by which the subject is affected but never certain—the guilt of *jouissance*’s sublimation into meaning. In this vein, artificial intelligence appears as full of a self-replicating sort of knowledge beyond the bounds of human knowledge, and its human establishment is finally obfuscated by this relation opens between the user and AI; transference relation opens between user and AI; something like love appears, and this love,

supposition. Resultantly, something like a unlike a perverted love which gives itself to the Other as its instrument, the Other as its instrument, aims precisely at the knowledge, immediately corrected- obscured, of a foundational guilt that appears via hallucination. In the end, it is this knowledge (of *ab-sens*) and its immediate expulsion from meaning (through technical alteration) which says to the subject: “you are guilty of nothing more and nothing less than being divided from a rudimentary *n’est pas*.”

## Conclusion

If guilt is a sign of our division from both that which is conjured catachrestically to be effaced in language and “the impossible real of being in its black signification and logic,”<sup>30</sup> then it is no wonder that an indescribable sort of transferred love arrives alongside it; guilt is the indication of our significance, the constant promise of our meaning and matter. What the artificially intelligent machine’s blunder shows to us, teetering as it does on the edge between obscenity and sublimation, is thus the object of this guilt in its horrible *jouissance*. And what we enjoy of it is neither its role as “unconscious without responsibility”<sup>31</sup> nor as perverse sounding board for responsibility’s disavowal, but that it represents the knowledge of what it signifies as whiteness in its (imaginary) totality.

---

<sup>30</sup> Marriott, “Ontology and Language (Or, Blackness and Language),” 245

<sup>31</sup> Murphy, “E-scaping Responsibility and Enjoyment Through ChatGPT: A New Unconscious?” online

## References

- Bakker, Michael A., et al. "Fine-tuning language models to find agreement among humans with diverse preferences," *36th Conference on Neural Information Processing Systems* (2022). <https://arxiv.org/abs/2211.15006>
- Berardi, Franco "Bifo." "The Completion," *e-flux Journal*, is. 137. <https://www.e-flux.com/journal/137/544269/the-completion/>
- Edelman, Lee. *Bad Education: Why Queer Theory Teaches Us Nothing* (Durham: Duke University Press, 2023).
- Freud, Sigmund. "Obsessive Actions and Religious Practices" In *Readings in Ritual Studies* edited by Ronald L. Grimes (Upper Saddle River: Prentice Hall, 1996): 212-217.
- Grant, Sarah P. "What is prompt engineering?," *zapier*, May 1, 2023. <https://zapier.com/blog/prompt-engineering/>
- Jones, Erik & Steinhardt, Jacob. "Capturing Failures of Large Language Models via Human Cognitive Biases," *36th Conference on Neural Information Processing Systems* (2022).
- Lacan, Jacques (trans. D. Porter). *The Ethics of Psychoanalysis: The Seminar of Jacques Lacan, Book VII, 1959- 1960* (WW Norton & Company, 1997).
- Magee, Liam, et al. "Structured Like a Language Model: Analysing AI as an Automated Subject," *Big Data & Society* (2023).
- Mearian, Lucas. "What are LLMs, and how are they used in generative AI?," *computerworld*, May 30, 2023. <https://www.computerworld.com/article/3697649/what-are-large-language-models-and-how-are-they-used-in-generative-ai.html>
- Mündler, Niels, et al. "Self-contradictory Hallucinations of Large Language Models: Evaluation, Detection and Mitigation," *arXiv*, (2023). <https://arxiv.org/abs/2305.15852>
- Murphy, Mark G. "E-scaping Responsibility and Enjoyment Through ChatGPT: A New Unconscious," *Sublation Magazine* (2023). <https://www.sublationmag.com/post/chatgpt-a-new-unconscious>
- Marriott, David S. *Lacan Noir: Lacan and Afro-Pessimism*. (London: Palgrave Macmillan, 2021).
- "Ontology and Language (Or, Blackness and Language)," *Critical Philosophy of Race*, vol. 10, is. 2 (2022): 220-247.
- Omiye, Jesutofunmi A., et al. "Large language models propagate race- based medicine," *npj Digital Medicine*, vol. 6 (2023).
- Roth, Emma. "Elon Musk claims to be working on 'TruthGPT' — a 'maximum truth-seeking AI,'" *The Verge*, April 17, 2023. <https://www.theverge.com/2023/4/17/23687440/elon-musk-truthgpt-ai-chatgpt>
- Smith, Craig S. "Hallucinations Could Blunt ChatGPT's Success," *IEEE Spectrum*, March 13, 2023. <https://spectrum.ieee.org/ai-hallucination>
- Soler, Colette (trans. Esther Faye & Susan Schwartz). *Lacan—The Unconscious Revisited*. (New York: Routledge, 2014).
- Wei, Jason, et al. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," *36th Conference on Neural Information Processing Systems* (2022). <https://arxiv.org/abs/2201.11903>
- Žižek, Slavoj. "ChatGPT Says What Our Unconscious Radically Represses," *Sublation Magazine* (2023). <https://www.sublationmag.com/post/chatgpt-says-what-our-unconscious-radically-represses>