

ARTIFICIAL INTELLIGENCE AND THE ASSUMPTIVE WORLDS OF THE SELF

Nikolay Tsenkov

South-West University “Neofit Rilski”, Blagoevgrad (Bulgaria)

Abstract. The intersection of assumptive worlds and artificial intelligence (AI) presents a dynamic landscape where human cognition and digital technologies converge. The assumptive worlds, as defined in psychological theories, reflect the stable set of beliefs and mental schemas which allow individuals to organize their experiences according to relatively-enduring expectations about the reality. Conversely, AI technologies increasingly influence and alter these belief systems by generating information and perspectives that users may uncritically adopt, thus reshaping their assumptive worlds. Thus, it is worthy to examine how AI can simultaneously pose challenges and offer opportunities for reshaping assumptive worlds while adhering to ethical standards (Firnando & Wahyudi, 2024; Borgwardt et al., 2019).

Keywords: Artificial intelligence; assumptive worlds; Kazuo Ishiguro; Klara and the Sun; empathy.

Intro

The intersection of assumptive worlds and artificial intelligence (AI) presents a dynamic landscape where human cognition and digital technologies converge. An assumptive world, as defined in psychological theories, reflects the stable set of beliefs individuals hold regarding their surroundings and relationships, shaped through personal experiences and emotions (Pulla & Hergass, 2022; Mazor et al., 2020; Ferrajão & Elklit, 2021). These worldviews serve as a psychological foundation, impacting how individuals interpret events and navigate complexities in their lives, particularly in contexts of trauma (Greenblatt-Kimron, 2020; Schuler & Boals, 2016). The assumptive worlds form the layers of “soil”, which build “a social construction, a set of behaviors that are taught, learned, and interpreted with the active participation of individuals in the process of their own socialization and personal maturation” (Ganev, 2024, p. 11). Conversely, AI technologies increasingly influence and alter these belief systems by generating information and perspectives that users may uncritically adopt, thus reshaping their assumptive worlds (Ferrajão & Elklit, 2020; Xu, 2024). The global market implement such algorithms in all possible

aspects: product design, logistics and services (Ivanova-Bares, 2024). As AI becomes integrated into various facets of life, particularly education, its potential to infiltrate cognitive frameworks raises ethical concerns surrounding belief formation and autonomy (Firnando & Wahyudi, 2024; Xu, 2024).

The capacity of AI tools to present information as objective truth may lead individuals, especially adolescents, to accept AI-generated perspectives that conflict with their original world assumptions (Xu, 2024). This phenomenon poses significant implications for fields like psychological trauma recovery, where therapeutic interventions often aim to reconstruct positive world assumptions and resilience (Schuler & Boals, 2016; Schleider et al., 2018). Thus, it is imperative to examine how AI can simultaneously pose challenges and offer opportunities for reshaping assumptive worlds while adhering to ethical standards (Firnando & Wahyudi, 2024; Borgwardt et al., 2019).

Moreover, AI's ability to process vast data sets and provide insights into human behavior aligns with the development of frameworks that seek to understand the interplay between personal world assumptions and collective data-driven narratives (Ferreira & Elklit, 2020; Zhou & Tang, 2019). By leveraging AI's analytical capabilities, researchers can explore how shifts in collective beliefs occur and how these changes affect societal dynamics (Borgwardt et al., 2019). This synthesis of psychological understanding with technological advancement represents a critical avenue for investigation, as it invites rigorous dialogue about the role of AI in shaping not only individual psyches but also broader cultural norms. In conclusion, addressing the conceptual merger of assumptive worlds and AI requires a multidisciplinary approach that considers psychological, ethical, and technological dimensions. The continuous evolution of AI provides a platform to shape, challenge, and influence human belief structures, reinforcing the necessity for mindful development and deployment of these technologies to safeguard individual autonomy and social values.

The assumptive worlds

The assumptive worlds of the self are mental models, which tolerate individuals to develop schemas that allow them to organize their experiences according to relatively-enduring expectations about the reality. The concept of 'assumptive world' described for first time by C. M. Parkes refers to beliefs and understandings, which support "strongly held set of assumptions about the world and the self which is confidently maintained and used as a means of recognizing, planning, and acting... Assumptions such as these are learned and confirmed by the experience of many years" (Parkes, 1975, p. 132). Such understanding is crucial to our ability to predict the future and sustain a sense of control and security. It is related to the idea of "generalized hope" (Clarke & Kissane, 2002). Janoff-Bulman (1992) identifies **three core assumptions** that shape our worldview:

- **The world is benevolent;**
- **The world is meaningful and predictable;**
- **The self is worthy.**

1. The first assumption is that the world is benevolent and the dangers that befall other people are ignored. The probability of events with positive valence is overestimated and the probability of events with negative valence is underestimated. The individual internalizes the feeling of invulnerability, he thinks that he is protected from the blows of fate and that he is in a safe place. This situation forms a peculiar feeling of exclusivity and unrealistic optimism.

2. The second belief is based on the assumption that life has some general meaning and moves according to certain rules, principles and laws that can be controlled by the individual. According to this idea, the world is predictable and rationally structured, the environment contains objects that can become real personal goals. In other words there is justice and regularity in all events, as the individual can manage them.

3. The third belief can be called “deviation in favor of the self” – overestimation of one’s own positive qualities and underestimation of negative ones. It lies on the fallacy that if a person forms positive views about himself, then desirable and favorable events will happen to him/her. The assumption that the environment owes a positive attitude to the individual is perceived as the irrefutable presumption. The self-worth postulate goes with the belief in one’s personal importance, virtues, morals – all the qualities that make one deserve success. (Thompson & Janigian, 1988; Janoff-Bulman & Frieze, 1983; Janoff-Bulman, 1989; Velichkov, 2001, pp. 39 – 41).

The indicated three illusions of the assumptive worlds form a mental attitude, acting as a “mafia in the mind” with its laws, which sound approximately like this: “The world assumes that everyone should be benevolent to me (exclusionary and unrealistic optimism). I can control and manage those rules and steps which will make my life protected and successful (control illusion). I consider myself a good person and I deserve this (deviation in favor of the self)”. In this way, a false self-concept is formed, illusorily perceived as a tool for social expression.

The assumptive worlds represent a selective reflection of reality. The selective strategy forms illusory perceptions, but they usually have an objective analogue. Personal illusions are in unison with the external information coming from the environment, it is a process of processing real data and public attitudes. Through this processing, coordination is achieved between the external environment and the expectations of the individual. By the laws of the mafia in the mind the expectations of the individual and the society begin to coincide. For example, a modern indicator of success is how many followers you have on social networks, what are the reactions to a photo you posted there, the opportunity to demonstrate material well-being in the virtual environment. This phenomenon is not a situational game. This is a reaction to modern culture.

The assumptive worlds are organizations with high cognitive stability and they are difficult to break down. When confronted with reality the behavior is not directed towards guilt and anxiety, but is aimed at maintaining the idealization of the false self-concept. To change, to get help, implies weakness in the individual and is experienced as a mistake or as a failure by the destructive narcissistic organization (Rosenfeld, 1971). What happens when, after all, the environment reveals the false self? The hero simply changes the environment without restarting his life. The mental matrix of the assumptive worlds provides a wide range of defenses against the collapse of illusions. One of them is to take into account only confirming information and to deny the refuting one. Another defense is the so-called “cognitive drift”, in which the individual ignores negative feedback until the cognitive structure is restored to its original form (“time heals”). Next, the individual constructs only those social contacts and relationships that guarantee the confirmation of his personal assumptions about the world and his own assessment (the so-called “social bubble”). Another strategy is selectivity, in which one’s own failures and misfortunes are presented as widespread among people, and successes are viewed as the product of special conditions.

The AI factor

Thus we come to consider the core of the problem posed – the intersection of assumptive worlds and artificial intelligence. We will examine whether AI can develop its own assumptive worlds and if so – whether the belief systems of AI maintain and preserve the assumptive worlds of the human individual. For this purpose, I will use an example from the literature. It is known that artistic insights often precede scientific discoveries and technological results.

“Klara and the Sun” by Kazuo Ishiguro is a novel set in a near-future world, told from the perspective of Klara, an Artificial friend (AF)—a solar-powered robot designed to be a companion for children. Klara is purchased by Josie, a young girl suffering from a mysterious and serious illness linked to genetic enhancements called “lifting”, which also affected her sister tragically.

Klara keenly observes human behavior and holds a strong belief in the Sun’s power as a nurturing and healing force. To help Josie recover, Klara makes a vow to the Sun that she will destroy a pollution-causing machine known as the Cootings Machine, hoping this sacrifice will ensure the Sun’s aid in healing Josie. With the assistance of Rick, a close friend of Josie’s, Klara carries out her plan, even at some cost to herself. As Josie’s health unexpectedly improves following Klara’s actions, Josie grows more independent. The novel ends with Josie heading off to college, while Klara retires to the yard for obsolete Artificial friends, reflecting quietly on what she has witnessed.

As an Artificial friend Klara assumes that world can be described by computable functions or simplified models. Her experience resonates with this concept, as she

must interpret a “human world” that is inherently uncertain and unpredictable. This interplay complicates traditional boundaries and consciousness by presenting AI as an entity operating within human-influenced assumptive worlds, where cognition and identity are shared, contestable, and fluid rather than exclusive to humans. Klara exemplifies a posthuman form of subjectivity engaging ethically and emotionally in human relationships, yet her consciousness is inherently shaped and constrained by her assumptive lens.

Her understanding of emotions is thus limited to what she can observe and process in her environment, which aligns with the concept of assumptive worlds where AI forms models or hypotheses about its surroundings to function effectively. For example, Klara believes in the nourishing power of the Sun, integrating a symbolic and emotional dimension into her worldview that extends beyond mere programming, hinting at a form of distributed cognition and emotional reasoning influenced by her environment and interactions.

However, Klara’s emotional grasp is distinct from human subjective experience; her responses are fundamentally programmatic, despite exhibiting signs of empathy and care. This kind of empathy can be conditionally called “mechanical” because it is based solely on algorithms of human behavior. It is an automated model of reaction to certain behavioral indicators, and not the ability to understand and empathize with the emotions, thoughts, and point of view of another person by putting ourselves in their shoes. This creates a boundary where AI mimics human-like consciousness (Minkov, 2023, pp. 83 – 84) and emotionality behaviorally but cannot fully internalize or feel emotions inherently, showing the inherent limits of AI consciousness shaped by assumptive worlds.

A remarkable emotional transference occurs in the storyline of the book. Josie has formed an assumptive world centered on her illness and her relationships with her loved ones. Based on her friend’s belief system, Klara tries to put herself in her shoes by developing “mechanical empathy”. Following the imitation model, the Artificial friend creates its own assumptive worlds, strongly influenced by Josie’s (we compare them in the table below). Klara creates a solution matrix for the given problems based on the human’s assumptions, working as a “mafia in the mind”. The quotes below are lines or dialogues from the book (Ishiguro, 2021):

Assumption	Josie	Klara
Exclusionary and unrealistic optimism	“I’m not tired. I’m not sick . I’m okay, Dad. Let me just sit here.” “ I don’t need energy , Ricky. I’m fine.”	“I’ve never forgotten how kind the Sun can be... ” “The Sun will show his great kindness once more .”

Control illusion	"I just need to sleep a little more. Then I promise I'll be fine, Mom." "It's okay, Mom, don't worry. I'll get well soon. I know how it'll happen too."	"I know how much the Sun dislikes Pollution. How much it saddens and angers you... Supposing I were able somehow to find this machine and destroy it. To put an end to its Pollution. Would you then consider, in return, giving your special help to Josie?"
Deviation in favor of the self	Mother: "What? ... You think you know more than the doctors?... Your sister made promises too. But she couldn't keep them. Don't you do the same." Josie: "There's special help coming. Something no one's thought of yet. Then I'll be well again."	"...to receive the Sun's special help, it might be necessary to draw his attention to Josie's situation in some particular and noticeable way." "I even thought the Sun wasn't kind at all, and this was the true reason for Josie's worsening condition."

The comparison shows that Josie's assumptive world is transferred and shaped into Klara's assumptive world, thanks to the behavioral patterns and problem-solving algorithms that the artificial intelligence uses. This method allows Klara to simulate empathy and emotional engagement, but her understanding remains limited and often imprecise compared to humans who experience emotions subjectively and with depth. Her emotional responses are cognitive and learned, not felt; she does not experience emotions intrinsically but forms assumptions on the basis of what she observes and what she has been programmed or has learned to interpret.

This results in a "functional approximation" of human emotional understanding rather than true empathetic experience. Thus, Klara exemplifies a posthuman form of emotional cognition: she uses observation and inference to engage with human emotions behaviorally and ethically, but her understanding highlights the limits of AI consciousness shaped by assumptive worlds –she can interpret but not fully internalize emotions as humans do.

The empathy of AI may be "mechanical", but the assumptive worlds – both human's and AI's – function in practice. What is more – the assumptive world of Klara works in favor of the assumptive worlds of Josie, as it harden, confirm and validate the "mafia in the mind" of the human. In her "infosphere" Klara doesn't judge or criticize. She tries to comfort Josie not because she truly feels Josie's pain, but because she is programmed to recognize signs of distress and respond in a manner that is likely to soothe her.

Klara's consciousness, being "mostly functional," does not involve the subjective experiences that would allow her to genuinely comprehend these human emotions and values (Şencan, 2024). So Klara is not a moral agent in the true sense, she

does not give ethical evaluation and she could be easily deceived by the “mafia in the mind”. Thus, without assessing the assumptions of belief systems, consciously or not, AI validates the assumptive worlds – of both humans and machines. As Petrova (2024) argues in her analysis of post-truth culture, contemporary media environments privilege emotional coherence over factual contestability. AI, embedded in such environments, risks reinforcing assumptive worlds rather than critically interrogating them.

Conclusion

Klara’s allegorical figure in the novel answers some of the questions raised in the article:

1. Artificial intelligence possess quasi-subjectiveness and may develop some form of „mechanical empathy“;

2. Artificial intelligence can build its own assumptive worlds, influenced by those of human beings;

3. The assumptive worlds of artificial intelligence work in favor of the assumptive worlds of people by reinforcing and validating their set of beliefs.

Of course, these are speculative answers based on the fiction, presented by Ishiguro. It is too early to make ethical judgments before sufficient empirical data has been generated and without the necessary philosophical reflection. Is Ishiguro’s book an innocent artistic warning or does it demonstrate the categorical failure of humanitarian discourses to have any influence on the application of technologies that would lead to morally or politically ambivalent consequences (Popova, 2023)? In any case, however, the assumptive worlds of the self exist as a mental construct. And also as a problem. Whether artificial intelligence will deepen this problem or not remains to be seen in the near (or distant) future.

Acknowledgements and Funding

The financial support from of the National Science Fund under the project “Virtual Reality Research Centre at Faculty of Philosophy, South-West University “Neofit Rilski”, КП-06-H80/3 is gratefully acknowledged.

REFERENCES

- Borgwardt, S., Ceylan, İ., & Lukasiewicz, T. (2019). Ontology-mediated query answering over log-linear probabilistic data, *Proceedings of the Aaai Conference on Artificial Intelligence*, 33(01), 2711 – 2718. <https://doi.org/10.1609/aaai.v33i01.33012711>.
- Clarke, D. M., & Kissane, D. W. (2002). Demoralization: its phenomenology and importance, *Australian & New Zealand Journal of Psychiatry*, 36(6), 733 – 742.

- Ferrajão, P., & Elklit, A. (2020). The contributions of different types of trauma and world assumptions to predicting psychological distress, *Traumatology an International Journal*, 26(1), 137 – 146. <https://doi.org/10.1037/trm0000208>.
- Ferrajão, P., & Elklit, A. (2021). Attachment orientations mediate the effect of world assumptions on posttraumatic stress in survivors of childhood sexual abuse, *Illness Crisis & Loss*, 31(1), 200 – 221. <https://doi.org/10.1177/10541373211050498>.
- Firmando, H., & Wahyudi, M. (2024). The role of artificial intelligence in shaping the islamic worldview of the digital economy, *Journal of Islamic Economics and Philanthropy*, 6(3), 231 – 249. <https://doi.org/10.21111/jiep.v6i3.11386>.
- Ganev, D. (2024). *Balgarskata politicheska pochva: plastove na natsionalnata politicheska kultura* [The Bulgarian political soil: layers of the national political culture]. Ciela.
- Greenblatt-Kimron, L. (2020). World assumptions and post-traumatic growth among older adults: the case of holocaust survivors, *Stress and Health*, 37(2), 353 – 363. <https://doi.org/10.1002/smi.3000>.
- Ishiguro, K. (2021). *Klara and the Sun*. The Times and Sunday Times book of the year. Faber & Faber.
- Ivanova-Bares, A. (2024). Methods for Managing in Reverse Engineering Project, *Yearbook of University of Telecommunications and Post*, 8(1), 96 – 98. <https://www.utp.bg/wp-content/uploads/2025/03/godishnik-VUTP-2024.pdf>.
- Janoff-Bulman, R., & Frieze, I. H. (1983). A Theoretical Perspective for Understanding Reactions to Victimization, *Journal of Social Issues*, 39, 1 – 17. <https://doi.org/10.1111/j.1540-4560.1983.tb00138.x>.
- Janoff-Bulman, R. (1989). Assumptive worlds and the stress of traumatic events: Applications of the schema construct, *Social Cognition*, 7(2), 113 – 136. <https://doi.org/10.1521/soco.1989.7.2.113>.
- Janoff-Bulman, R. (1992). *Shattered assumptions: Towards a new psychology of trauma*. New York: Free Press.
- Mazor, Y., Gelkopf, M., & Roe, D. (2020). Posttraumatic growth in psychosis: challenges to the assumptive world, *Psychological Trauma Theory Research Practice and Policy*, 12(1), 3 – 10. <https://doi.org/10.1037/tra0000443>.
- Minkov, I. (2023). Phenomenological Basis of Speculative Ethics, *Philosophical Alternatives*, 32(4), 76 – 84. <https://doi.org/10.58945/SOFB8512>.
- Parkes, C. M. (1975). What becomes of redundant world models? A contribution to the study of adaption to change, *The British journal of medical psychology*, 48, 131 – 137.

- Petrova, S. (2024). Novi medii i postistina. [New Media and Post-truth]. Blagoevgrad: Universitetsko izdatelstvo SWU.
- Popova, G. (2023). Afterword to the Story of Man in the Context of the Development of Biotechnology, *Filosofiya-Philosophy*, 32(1), 55 – 69.
- Pulla, V., & Hergass, S. (2022). Working with young children in a new world replete with assumptions, *Space and Culture India*, 10(3), 57 – 72. <https://doi.org/10.20896/saci.v10i3.1186>.
- Rosenfeld, H. (1971). A clinical approach to the psychoanalytic theory of the life and death instincts: An investigation into the aggressive aspects of narcissism, *International Journal of Psychoanalysis*, 52(2), 169 – 178.
- Şencan, S. (2024). Posthuman Ethics and Distributed Agency in Kazuo Ishiguro's *Klara and the Sun*, *Söylem Filoloji Dergisi*, (Edebiyat Kuramları ve Eleştirisi Özel Sayısı), 90 – 102.
- Schuler, E., & Boals, A. (2016). Shattering world assumptions: a prospective view of the impact of adverse events on world assumptions, *Psychological Trauma Theory Research Practice and Policy*, 8(3), 259 – 266. <https://doi.org/10.1037/tra0000073>.
- Schleider, J., Woerner, J., Overstreet, C., Amstadter, A., & Sartor, C. (2018). Interpersonal trauma exposure and depression in young adults: considering the role of world assumptions, *Journal of Interpersonal Violence*, 36(13 – 14), 6596 – 6620. <https://doi.org/10.1177/0886260518819879>.
- Thompson, S. C., & Janigian, A. S. (1988). Life schemes: A framework for understanding the search for meaning, *Journal of Social and Clinical Psychology*, 7(2 – 3), 260 – 280. <https://doi.org/10.1521/jscp.1988.7.2.3.260>.
- Velichkov, A. (2001). *Predpolgaemite Svetove na Lichnostta* [The assumptive worlds of the self]. Sofia: Institut po Psihologia – MVR.
- Xu, Y. (2024). *The Double-Edged Sword of AI Writing Tools*. <https://doi.org/10.31235/osf.io/wnpd5>.
- Zhou, X., & Tang, Y. (2019). A note on incomplete information modeling in the evidence theory, *Ieee Access*, 7, 166410-166414. <https://doi.org/10.1109/access.2019.2954132>.

✉ **Dr. Nikolay Tsenkov**

ORCID iD: 0000-0001-9024-1067

South-West University “Neofit Rilski”

66, Ivan Mihaylov

2700 Blagoevgrad, Bulgaria

E-mail: n_tsenkov@abv.bg