Turkish Journal of Psychiatry 2024;35(4):340–342

Türk Psikiyatri Dergisi

Turkish Journal of Psychiatry

Letter to the Editor

On the 'Hallucinations' of Artificial Intelligence and the Hallucination Experience in Human

Dear Editor,

We have read Mahmut Özer's editorial letter titled "Does Artificial Intelligence Have Hallucinations?" published in your journal. It is well-known that large language models based on artificial intelligence (AI), such as ChatGPT, recently made widely available, are not immune to errors and can generate incorrect information (Ayhan, 2023). As Mr. Özer explains clearly in his article, the term "AI hallucinations" is often used to define these different types of "false" or "fabricated" responses (Özer, 2024). Given that the concept of "intelligence," which pertains to the human mind, serves as an analogy in the development of this technology, it is understandable that attempts are made to explain its malfunction (pathology) using a term related to the human mind. However, as Mr. Özer also pointed out in his article, the production of false or fabricated information by AI does not fully overlap with the concept of hallucinations, a neuropsychiatric phenomenon. Nevertheless, we believe that addressing the hallucinatory experience within a neuropsychiatric context in parallel with this article would offer a more comprehensive understanding to readers of the journal.

The term "hallucination" entered the psychiatric literature towards the end of the 18th century, and the briefest definition of the term was made as "perception without an object" (Esquirol, 1822; Telles-Correia et al., 2015). In its widely accepted form today, hallucinations are defined as "sensory experiences that occur in the absence of an external stimulus from the relevant sensory organ, have a sufficient sense of reality to resemble actual perception, occur in a waking state, and are experienced without the individual having direct or voluntary control over them" (David, 2004). In the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5), which helps establish a common language among psychiatrists, hallucinations are similarly described as perception-like experiences that occur without an external stimulus, with the following characteristics: "They are vivid and clear, have the full force and impact of normal perceptions, and are not under voluntary control" (American Psychiatric Association, 2013).

An important feature indicated by these definitions is that hallucinations occur spontaneously, even intrusively, and are not under the individual's voluntary control. Although there may be triggering conditions, they do not emerge directly in response to any preceding input. Another key characteristic of hallucinations, which differs significantly from what is defined in AI, is that they are subjective experiences. For example, when considering auditory verbal hallucinations (AVH), the content of the heard voice, its physical characteristics, the attributed source's characteristics (gender, age, status, etc.), the duration, frequency, and intensity of the hallucinations, and the emotions experienced by the person in response to these experiences all vary from individual to individual. Furthermore, all of these are features experienced only by the individual, and they can only be expressed to the outside world as an output through communication with others (Parnas et al., 2024).

Hallucinations can occur in all sensory modalities, including hearing, sight, touch, smell, taste, and even proprioceptive senses, and sometimes in multiple modalities simultaneously (Montagnese et al., 2021). The hallucinatory experience has been associated with numerous clinical conditions, and it has also been reported that 10-15% of the population mayexperience hallucinatory-like phenomena (Sommer et al., 2008a). Major clinical conditions associated with hallucinations include psychiatric disorders such as schizophrenia spectrum disorders and bipolar disorder, neurodegenerative diseases like Lewy Body Dementia (LBD),

the effects of drugs and substances, epilepsy, brain tumors, and other neurological and medical conditions, as well as sensory disorders such as hearing or vision impairments. While some phenomenological features are statistically more prominent in different diagnoses (e.g., visual hallucinations being much more common in LBD), none of them is considered pathognomonic for any particular diagnosis (Waters & Fernyhough, 2017).

Many hypotheses have been proposed regarding the causes and processes of hallucinations from past to present, but a definitive and universally accepted mechanism has yet to be clarified. With the introduction of hallucinations into psychiatric literature, a debate emerged about whether their origin is perceptual or cognitive in nature (Telles-Correia et al., 2015). Considering the heterogeneous nature of the hallucinatory experience, it is believed that multiple processes with overlapping features may play a role. For instance, the cognitive mechanisms underlying musical auditory hallucinations resulting from medication side effects, where the individual retains full insight, are likely to differ from those of auditory verbal hallucinations that are threatening in nature and interpreted through delusional explanations. Among the various types of hallucinations, auditory verbal hallucinations (AVH) are the most extensively studied in terms of their formation processes, and several hypotheses have been proposed regarding their development (Barber et al., 2021).

The first hypothesis regarding the cognitive model of auditory verbal hallucinations (AVH) is based on a reduction in top-down inhibition. It is proposed that dysfunction of the prefrontal cortex, which is responsible for inhibition, leads to hyperactivation in functional networks, including the auditory cortex, resulting in abnormal auditory signals (Waters et al., 2012). When these abnormal neural activations produce auditory signals that exceed the perception threshold, they may generate unexpectedly intense (hyper-salient) sensory information.

Additionally, deficits in source monitoring led to these stimuli being perceived as foreign and distinct from the individual's internal mental processes, concluding that their source is external. According to this hypothesis, the content of AVHs is shaped by factors such as perceptual expectations, mental imagery, and prior experiences or knowledge (e.g., memories), which contribute to the formation of a unique and highly personalized sense of reality.

Another relatively consistent hypothesis in the explanation of auditory verbal hallucinations (AVH) is the "memory intrusion" hypothesis (Waters et al., 2006). According to this hypothesis, with a reduction in cognitive inhibitory control, material related to long-term memory (memory fragments) is activated in a repetitive, unexpected, and intrusive manner, disconnected from its context, and becomes involved in language processing, thus experienced as AVH (Ćurčić-Blake et al., 2017). There is clear evidence that in patients with AVH, memory networks abnormally interact with language areas (Ćurčić-Blake et al., 2017). It has been suggested that this explains the intrusive and repetitive nature of hallucinations. In support of this, in individuals diagnosed with schizophrenia, hippocampal deactivation was observed just before the onset of AVH, which points to a memory discharge (Hoffman et al., 2008). Additionally, source-monitoring deficits accompanying memory intrusions lead to these intrusions being perceived as external stimuli rather than internal processes.

Finally, the most prominent hypothesis in explaining auditory verbal hallucinations (AVH) today is the attribution of inner speech to an external source (Langland-Hassan, 2016). Inner speech is the process of talking to oneself in order to perform functions such as planning and verbal rehearsal. In this case, a self-monitoring error occurs, and the individual mistakenly attributes their own inner voice to an external source. As a result, the common points where all the hypotheses intersect are that mental contents (such as thoughts, memories, plans, etc.) are transformed into auditory verbal forms, and this auditory representation becomes dissociated from subjective ownership.

Hallucinations can also be addressed through the predictive processing theory, which is a Bayesian approach and shares similarities with AI working principles (Fletcher & Frith, 2009; Cho & Wu, 2013; Horga et al., 2014a). According to this theory, the human brain is not a passive receiver of sensory information but an inferential process that actively generates predictions about sensations. The individual has a top-down "prediction" about sensory information, and after the sensation, this prediction is updated with the sensory data. The difference between the prediction and the actual sensory information is referred to as "prediction error." Actions initiated by the individual are more predictable, resulting in a much smaller prediction error, and as a result, the salience of that sensation is more limited. In contrast, external stimuli tend to cause a higher prediction error, and these stimuli are harder to ignore. In schizophrenia, many experimental and electrophysiological studies have demonstrated a disruption in the prediction error processes, and it is believed that this disruption, especially in the context of AVH, leads to inner speech being perceived as an external voice (Randeniya et al., 2018). It has been suggested that the marked activity observed in the left inferior and superior temporal gyri during AVH may be associated with a prediction error leading to inner speech misattribution. Current meta-analyses, highlighting the existence of conflicting results, also emphasize the importance of the insula (Barber et al., 2021).

The formation of the verbal content of AVH has been associated with language processing mechanisms. It has been shown that during normal speech, the lateralization of Broca's

area occurs in the left hemisphere, whereas during AVH, the right Broca's area is activated (Sommer et al., 2008b). In aphasia cases where the left hemisphere is damaged, individuals often speak with shorter, less complex sentences, especially those with negative content, and this has been compared to the AVHs of psychotic individuals. It is also suggested that the amygdala plays a role in forming negative verbal content (Horga et al., 2014b).

From this brief discussion on auditory verbal hallucinations, it can be concluded that viewing hallucinations solely as a perceptual pathology provides limited information. On the contrary, hallucinations can be understood as phenomena that arise from errors in self-perception and the internal-external distinction (Parnas et al., 2024), and they may be considered as the result of alterations in multiple complex mental processes, such as memory, attention, time perception, interpersonal relationships, and bodily experiences (Pienkos et al., 2019).

If the effort to understand the human mind is viewed as a form of reverse engineering, then discovering "pathological" phenomena that emerge while building AI, a "model" of the human mind, (which is forward engineering) can also be valuable in understanding the human mind itself. Rather than using analogies such as hallucinations or confabulations, which are still subject to conceptual debates and whose formation processes have not been fully explained, it may be more beneficial to seek alternative technical terms that more clearly define these situations in the long term.

If the sense of self is fundamental to the experience of hallucinations, another question that arises is whether it would be possible for artificial intelligence to experience hallucinations in the same way humans do, once it attains a sense of self. Perhaps such a development could illuminate our understanding and conceptualization of human hallucinations, helping us gain deeper insights into the phenomenon. This idea raises intriguing questions about the relationship between consciousness, self-awareness, and perceptual experiences, which may provide valuable perspectives on both human and artificial cognition.

REFERENCES

American Psychiatric Association (2013) Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), Diagnostic Criteria Handbook (Trans. Ed.: E Köroğlu) Hekimler Publishing Union, Ankara, 2013.

Barber L, Reniers R, Upthegrove R (2021) A review of functional and structural neuroimaging studies to investigate the inner speech model of auditory verbal hallucinations in schizophrenia. Transl Psychiatry 11: 582.

Cho R, Wu W (2013) Mechanisms of auditory verbal hallucination in schizophrenia. Front Psychiatry 4: 155.

Ćurčić-Blake B, Ford JM, Hubl D et al. (2017) Interaction of language, auditory, and memory brain networks in auditory verbal hallucinations. Prog Neurobiol 148: 1-20.

David AS (2004) The cognitive neuropsychiatry of auditory verbal hallucinations: an overview. Cogn Neuropsychiatry 9: 107-23.

Esquirol J (1822) Hallucinations: Dictionnaire des Sciences Médicales. Paris: CLF Panckouche.

Fletcher PC, Frith CD (2009) Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 10: 48-58.

Hoffman RE, Anderson AW, Varanko M et al. (2008) Time course of regional brain activation associated with onset of auditory/verbal hallucinations. Br J Psychiatry193: 424-5.

Horga G, Fernandez-Egea E, Mane A et al. (2014) Brain metabolism during hallucination-like auditory stimulation in schizophrenia. PloS One 9: e84987.

Horga G, Schatz KC, Abi-Dargham A et al. (2014) Deficits in predictive coding underlie hallucinations in schizophrenia. J Neurosci 34: 8072-82.

Langland-Hassan P (2016) Hearing a Voice as one's own: Two Views of Inner Speech Self-Monitoring Deficits in Schizophrenia. Rev Philos Psychol 7: 675-99.

Montagnese M, Leptourgos P, Fernyhough C et al. (2021) A Review of Multimodal Hallucinations: Categorization, Assessment, Theoretical Perspectives, and Clinical Recommendations. Schizophr Bull 47: 237-48.

Özer M (2024) Is artifical intelligence hallucinating? Turk Psikiyatri Derg 35:333-5. https://doi.org/10.5080/u27587

Parnas J, Yttri JE, Urfer-Parnas A (2024) Phenomenology of auditory verbal hallucination in schizophrenia: An erroneous perception or something else? Schizophr Res 265: 83-8.

Pienkos E, Giersch A, Hansen M et al. (2019) Hallucinations beyond voices: a conceptual review of the phenomenology of altered perception in psychosis. Schizophr Bull 45: S67-S77.

Randeniya R, Oestreich LK, Garrido MI (2018) Sensory prediction errors in the continuum of psychosis. Schizophr Res 191: 109-22.

Sommer IE, Daalman K, Rietkerk T et al. (2008) Healthy Individuals with Auditory Verbal Hallucinations; Who Are They? Psychiatric Assessments of a Selected Sample of 103 Subjects. Schizophr Bull 36: 633-41.

Sommer IE, Diederen KM, Blom JD et al. (2008) Auditory verbal hallucinations predominantly activate the right inferior frontal area. Brain 131: 3169-77.

Telles-Correia D, Moreira AL, Gonçalves JS (2015) Hallucinations and related concepts—their conceptual background. Front psychol 6: 991.

Waters F, Allen P, Aleman A et al. (2012) Auditory hallucinations in schizophrenia and nonschizophrenia populations: a review and integrated model of cognitive mechanisms. Schizophr Bull 38: 683-93.

Waters F, Badcock J, Michie P et al. (2006) Auditory hallucinations in schizophrenia: intrusive thoughts and forgotten memories. Cogn Neuropsychiatry 11: 65-83.

Waters F, Fernyhough C (2017) Hallucinations: A Systematic Review of Points of Similarity and Difference Across Diagnostic Classes. Schizophr Bull 43: 32-43.

Ezgi İNCE GULİYEV, Alp ÜÇOK

Received: 15.10.2024, **Accepted:** 17.11.2024, **Available Online Date:** 05.12.2024

[1]Assis. Prof., [2]Prof., Istanbul University, Istanbul Faculty of Medicine, Department of Psychiatry, İstanbul, Turkey.

Dr. Ezgi İnce Guliyev, e-mail: ezgi.ince@yahoo.com

https://doi.org/10.5080/u27608