Article

Emotional Intelligence of GPT-4 Large Language Model

Gleb Vzorin *, Alexey Bukinich, Anna Sedykh, Irina Vetrova, Elena Sergienko

Posted Date: 23 October 2023

doi: 10.20944/preprints202310.1458.v1

Keywords: emotional intelligence; ChatGPT; GPT-4; EI; EQ; artificial empathy; experiential EI; strategic EI; 4 branch model of emotional intelligence

Article

Emotional Intelligence of GPT-4 Large Language Model

Gleb D. Vzorin [1,2,*], Alexey M. Bukinich [1,3], Anna V. Sedykh [1], Irina I. Vetrova [2], Elena A. Sergienko [2]

[1] Lomonosov Moscow State University, Faculty of psychology, Russia
[2] Russian Academy of Sciences, Institute of Psychology, Russia
[3] Russian Academy of Education, Psychological Institute, Russia
* Correspondence: g.vzorin@mail.ru

Abstract: Current neural network models can demonstrate reasonable-looking behavior, considered by some developers and researchers human-like. For example, a large language model GPT-3 is susceptible to human-like cognitive biases. Yet there is no data of such models solving emotional intelligence (EI) tasks. They are connected to the abilities that has been previously considered as specifically human EI is an important aspect of human communication. The ability to understand and respond to emotional cues is essential for effective communication. Therefore, it is crucial to determine the ways AI models such as ChatGPT demonstrate EI. The present research aims to measure the EI of GPT-4, a large language model trained by OpenAI. Russian version of the Mayer–Salovey–Caruzo Emotional Intelligence Test sections B, C, D, F, G and H were used in this research. High points were obtained in Understanding emotions scale and Strategic EI. Mean points are obtained in Managing emotions scale. Low and less reliable values are obtained in Using emotions to facilitate thought scale. Thus, GPT-4 seems already capable of identifying emotions in text and describing techniques for managing them. However, complex cases and irregular situations requiring emotions qualitative analysis would be a hard task for GPT-4.

Keywords: emotional intelligence; ChatGPT; GPT-4; EI; EQ; artificial empathy; experiential EI; strategic EI; 4 branch model of emotional intelligence

1. Introduction

Artificial Intelligence (AI) has made significant strides in recent decades, evolving from basic rule-based systems to complex deep-learning models that can generate human-like text and complete human-like tasks. The first models were based on strict logical rules, the next generation of AI models was based on statistical data, and recent models rely on deep learning technologies (Haenlein, Kaplan, 2019). There were cases that showed the opportunities of AI. For example, success in playing chess and winning human professionals (McCarthy, 1990; Hassabis, 2017). More recent example relates to AI overriding human responses in decision-making and pattern recognition, particularly in the case on ImageNet, a project that used AI to classify and label images on a large scale (Krizhevsky et al., 2017). Such moments and further adaptability of AI models like ChatGPT (OpenAI, 2023a) have pushed AI into the mainstream with governments and policymakers across the world starting to engage in discussions and consider the implications of AI technologies.

Large language models (LLM) relate to AI technologies branches. There are many powerful LLM (Fan et al., 2023), and ChatGPT is one of the most famous models (OpenAI, 2023a) for the reason of user-friendly interface that allows any person to interact with LLM in simple chat form. Today it is widely used in different areas and followed by many discussive questions and ideas about possible and restricted implications (Kasneci et al., 2023; Nori et al., 2023). Clearly, the advancements of AI technologies and, particularly, success of LLM indicate that AI-powered language models and

assistants will become even more prevalent in our daily lives. The potential consequences of their actions and their feedback on complex emotional responses to these actions have significant implications for their integration into human society. Thus, it is crucial to understand the possibilities and limitations of AI not only in solving some practical tasks, but in the interaction with humans too.

Notably, some AI models have shown a propensity to spontaneously develop new cognitive abilities. For example, GPT-3 (elder version of GPT-4) can be equal to or even outperform human participants in vignette-based tasks, multiarmed bandit task, making decisions from descriptions and, finally, it demonstrates reinforcement learning signs (Binz & Schulz, 2023). These abilities were not explicitly programmed into the model but emerged as a property of the vast amount of text data the model was trained on. Various aspects of the cognitive functioning of neural network models have become the subject of scientific research, and the results already allow some researchers to talk about "sparks" of general artificial intelligence (Bubeck et al., 2023). It also supports the necessity of a scientific investigation of LLM capabilities and peculiarities.

While there has been extensive study on the cognitive capabilities of AI, emotional intelligence (EI) in these systems has not been thoroughly explored yet. This study aims to address this gap in understanding of AI emotional intelligence. The EI of GPT-4 will be measured by standardized EI test developed by Mayer, Salovey, and Caruso (2002) and then analyzed. This data may provide some insights into the similarities and differences in EI between AI systems and their human counterparts comparing the AI's responses with normative answers in humans.

2. Background

2.1. Artificial empathy and artificial emotions

Since time immemorial, humans have been intrigued by the potential for artificial systems to “feel” and exhibit empathy, and this fascination embodied in cultural narratives such as the Golem myth. In Jewish folklore, the Golem, a creature crafted from inanimate matter, was given life and, to some extent, elements of agency and sentience. Such passions and cultural prototypes inspire modern discussions about artificial intelligence (Vudka, 2020). At the same time, it is important to note that in the modern context, these eternal ideas raise new questions. The challenge of imbuing AI with emotional capability today is not only a philosophical speculation, but also an area of profound technical, scientific, philosophical, and ethical exploration.

Philosophically, the question of whether an artificial system can truly experience emotions raises fundamental questions about the nature of consciousness and emotion (e.g., Chalmers, 1997; Dennett, 2017). If an AI exhibits behavior consistent with emotional responses, does it truly "feel" these emotions? Or is it merely a sophisticated simulation (e.g., Searle, 1980)? What would it mean for an AI to have an internal subjective experience akin to human emotion? These are some of the complex questions that philosophers wrestle within the context of AI and emotions.

The ethical aspects of this problem are equally significant. If we are successful in creating AI that can experience emotions, we are then faced with the question of how we should treat these entities. Do they have rights (e.g. Coeckelbergh, 2010; Andreotta, 2021)? Should we blame the bad treatment of robots and approve of the good (e.g. Sparrow, 2021)? Moreover, there's the question of how these emotionally capable AI could impact human society, relationships, and mental health. Could they manipulate human emotions (e.g. Stephan, 2015; Shank et al., 2019; Bubeck et al., 2023)? Would they alter our understanding of what it means to be human (e.g., Bostrom, 2014; Lovelock, 2019)?

A scientific perspective on the question of “feeling” AI involves understanding the human brain and emotional responses at a depth sufficient to replicate them within a computational framework. This requires interdisciplinary research across neuroscience, psychology, and computer science, and the development of complex models that capture the nuances of human emotional response (Damiano et al., 2015; Assuncao et al., 2022).

A separate question for both human and computer sciences, particularly important in the context of this article, is the question of the functional role of emotions in artificial intelligence. It is well known in human psychology that emotions are not an isolated system. They play an essential role in

 doi:10.20944/preprints202310.1458.v1

cognitive processes and self-regulation (Pessoa, 2008; Lantolf & Swain, 2019). Expressed empathy as a social signal is closely related to other emotional abilities (e.g. Kornilova & Quiqi, 2021). So, in humans, empathy, emotional intelligence and emotions are closely related and represent from different aspects an indivisible psychological reality. Whether this is true of artificial systems is a big question. If a neural network is trained to give adequate emotional responses, this does not mean at all that it has emotional intelligence and emotions in the same sense as a human. AI may have its own functional equivalents of emotions (e.g., Czerwinski et al., 2021), but their presence probably has little to do with the manifestations of emotional intelligence as the ability to "understand" human feelings and manage them. These two separate emotion-related topics in AI represent communicative aspects and architectural aspects (Scheutz, 2014). In this paper, we leave out the question of AI's "own emotions" and focus on its ability to outwardly exhibit communicative emotional behavior, which can be measured by methods developed for humans.

2.1. Empathic AI in practice

Although it is possible to be disappointed in the chosen aspects of AI EI because of the architectural aspect lack, the communicative aspect is undoubtedly important in practice. Since the end of XX century there has been a growing interest in emotional-aware robotic or virtual agents assisting people in different areas (generally described in Trappl et al., 2002). The main reason was that interacting with humans always included dealing with their emotions, accepting them at least. Even in “classic” human communication non-conflict actions related to other person emotions require empathy (Halpern, 2007; Zaki, 2020), so that modern AI systems should be adapted to these “game rules”.

Shank et al. (2019) wrote five reasons for the necessity of taking human emotions into account when communicating with AI. The first was human reaction to AI actions in typical situations of social interaction. The second was various emotional reactions of people during the AI implementation it typically human areas. The third reason was the “uncanny valley” phenomenon. The fourth was provoking moral emotions during interaction with AI. And the last reason was the specificity of some AI instruments designed to work with people emotions. All these reasons are arguments for the necessity of human emotions consideration during the interaction with AI.

Several articles discuss the importance of AI EI in different practical areas. For instance, Prentice, Lopes, and Wang (2020) investigated the perceptions of employees regarding AI's role in hospitality management industry. The study highlighted the need for AI systems to possess EI to effectively replace or augment human roles. The researchers noted that AI systems without EI can lead to sub-optimal customer experiences, as these systems may fail to grasp and respond appropriately to the emotional cues and needs of customers. Kerasidou (2020) emphasized the importance of empathy, compassion, and trust in healthcare AI systems intertwined with EI. AI in healthcare, such as chatbots or diagnostic tools, are increasingly interacting directly with patients. These systems should ideally possess EI to understand and respond to the emotional state of patients, fostering trust and enhancing patient experiences. Therefore, measuring and improving the EI of AI systems can contribute both to better customer experiences and to empathetic and compassionate patient care insurance.

Special attention should be paid to digital psychological help tools. Certainly, creating artificial psychologist requires “empathic module”, and this requirement is being met differently by companies aiming at developing artificial psychologist. For instance, a digital therapeutic tool “Deprexis” has already written dialogue scripts (Twomey et al., 2020), while chatbots “Replika” and “Woebot” use algorithms of machine learning (Possati, 2021; Darcy et al., 2022). Anyway, Uludag (2023) describes the relevance of EI in the context of AI-supported chatbots used in psychology. These chatbots, which employ AI to provide psychological support, need to be capable of understanding and responding to a broad spectrum of human emotions. By effectively measuring and improving their EI, these chatbots can provide more nuanced, empathetic responses, thereby providing more effective psychological support. Another evidence was received from the survey comparing empathic and non-empathic version of socially assistant robot helping elderly people to deal with depression, where special analysis revealed that users were more engaged in conversation with an empathic

 doi:10.20944/preprints202310.1458.v1

version (Abdollahi et al., 2022). This again highlights the practical importance of EI in AI systems for mental health care.

Moreover, we can imagine plenty of areas using the AI applied with EI. It is possible to analyze customer habits and preferences, providing more targeted marketing proposals. Emotional AI can work with user experience helping developers of websites and applications. Certainly, such AI can improve user experience in entertainment. Although Huang et al. (2019) stated that humans (especially, managers) had to move more to tasks requiring empathy and interpersonal relationship, while AI would solve analytical and thinking tasks. Regarding tendency of considering AI capable of having EI it is possible to imagine AI as a chief executive officer (CEO) in some company. It will be able to solve both strategical, analytical tasks and those connected with people interrelations, their motivation, needs and feelings.

2.3. Making AI to "feel"

In Sections 2.1 and 2.2, we discussed the theoretical and practical significance of artificial emotions. From both positions, we can say that the emotional competency of AI agents may be described as the "last frontier" in our pursuit of achieving a truly human-like quality of interaction (Lovelock, 2019). It's a captivating, yet challenging goal, that continues to drive researchers and provoke thought and debate across a range of disciplines. In this section, we will briefly consider engineering approaches to the implementation of artificial emotions.

In Section 2.2, we distinguished the communicative and architectural aspects of artificial emotions and stated that in our study we focus on the former. Research into this aspect primarily resulted in human emotion recognition techniques, human–computer interaction and human–robot interaction (Scheutz, 2014). The application of machine learning and AI for emotion recognition from various human physiological and behavioral signals such as facial expressions, voice, and text has been extensively studied (see Dzedzickis et al., 2020). Subsequent research has expanded into two overlapping domains: affective computing (Picard, 1997) and social signal processing (Vinciarelli et al., 2009).

At the same time, studies of the architectural aspect of artificial empathy are being widely developed, various models of emotion implementation in the functioning of AI are being developed. Early theoretical articles predicting the inevitability of the introduction of emotional architectures into robots (e.g., Sloman and Croucher, 1981) were followed by a plenty of empirical research articles. One of the seminal works in this area was conducted by Shibata, Ohkawa, and Tanie (1996) who explored the spontaneous behavior of robots as a means for achieving cooperation. Their study posited that robots, and by extension all artificial systems, have the potential to display a form of cooperative behavior that closely resembles empathy. Their experiments involved programming robots with a basic emotional model, which was then used to drive their interactions with each other and with humans. The results indicated that these robots, when given appropriate emotional inputs, were capable of responding in ways that could be interpreted as empathetic.

Later, numerous artificial emotion models for AI systems have been proposed, having different theoretical foundations in psychology (see, for example, a review of 12 models in Kowalczuk & Czubenko, 2016). Despite such a flourishing of models, approaches and theories, a paradoxical picture is observed in practice. As Kasparov (2017) notes, communication with neuroscientists, psychologists and philosophers is often perceived by engineers as an unnecessary burden. The fundamental cognitive value of philosophical questions and deep psychological models, as a rule, is not disputed, but their practical value is recognized as incomparably small compared to the required efforts for their theoretical study. At the same time, the "bottom-up" engineering approach brings great benefits at a lower cost. So, LLM's like ChatGPT, as follows from their description (OpenAI, 2023b), were not equipped with any special models of emotions, as if they did not need them. Since the large volumes of text on which such models are trained already contain many emotional reactions of people, it is not difficult for AI to extract subtle features of emotional communication from them.

As we can see, the range of approaches to the problem of artificial empathy and artificial emotions is wide. With all their diversity, the practical significance of the result they are striving for

 doi:10.20944/preprints202310.1458.v1

is obvious, because for artificial agents to truly integrate into our social fabric, they must be able to demonstrate empathic behaviors (Erol et al., 2019). Therefore, at this stage we need to find clear criteria for measuring this result.

2.4. Measuring artificial empathy

As AI-based chatbots increasingly mimic human conversational behavior, they provide a unique medium for the application of traditional psychological methodologies, originally tailored for human participants. Dillion et al. (2023) revealed a high degree of correlation between the responses of ChatGPT and human subjects regarding moral judgments. The researchers posed ethically complex scenarios to both AI and humans, observing similar patterns of reasoning and resolution. The striking parallels noted by Dillion et al. not only highlight the considerable strides made in AI modeling and language comprehension, but also raise the provocative question of whether these AI models might eventually be able to wholly replace human subjects in certain areas of psychological and cognitive research.

Generally, there can be two possible approaches to the problem of taking AI models into consideration as participants in psychological research. The first, more rational and theoretical, may be based on some rationally created criteria. As an example relevant for this paper it is possible to name the criteria for considering AI emotional described in a manner of Turing test by M.T. Ho (2022). The second approach seems to rely on the empirical data collected from surveys using classic psychological diagnostical instrument and tests on AI. In a pioneering study in this approach conducted by Bubeck et al. (2023), the cognitive capabilities of the GPT-4 model were thoroughly examined, revealing an intriguing discovery—the manifestation of a theory of mind. The study, titled "Sparks of Artificial General Intelligence: Early experiments with GPT-4," sheds light on the ability of GPT-4 to comprehend and infer the mental states of other agents, a crucial aspect of human cognitive and emotional skills.

These findings were followed by a more targeted study specifically dedicated to the emotional awareness of ChatGPT (Elyoseph et al., 2023). The study employed the Level of Emotional Awareness Scale (LEAS), a reliable, performance-based assessment, to gauge ChatGPT's responses to a diverse set of twenty scenarios. By comparing the AI's performance against general population norms, the researchers were able to effectively measure the AI's proficiency in identifying and understanding emotions. Remarkably, ChatGPT demonstrated a significant improvement in its emotional awareness, with a near-maximum LEAS score (Zscore=4.26) and high accuracy levels (9.7/10).

Emotional awareness as the cognitive ability to perceive, describe and differentiate one's own and others' emotional experiences (Lane and Schwartz, 1987) is an essential component of emotional intelligence in both trait and ability models (Agnoli et al., 2019). Nevertheless, it does not exhaust the abilities necessary for the genuine manifestation of empathy, conceptualized more fully in the construct of emotional intelligence. Therefore, more differentiated studies are required.

2.5. EI models in psychology

Emotional Intelligence (EI) is most often defined as the ability or skill to perceive, use, understand, manage, and cope with emotions. In psychological literature, it is common to distinguish three types of models (Kanesan & Fauzan, 2019). Since the measurement method will depend on the model chosen, it is important to consider them and determine the most appropriate for the purposes of our article.

Salovey and Mayer (1990) introduced the ability model of EI, which conceptualizes EI as a cognitive ability that can be measured by performance tests. This model focuses on the individual's ability to process emotional information, particularly in relation to problem-solving. According to this model, EI comprises four interconnected abilities: perceiving emotions, using emotions to facilitate thought, understanding emotions, and managing emotions.

The ability model posits that individuals with high EI are better equipped to perceive and interpret the emotional states of themselves and others, utilize emotions to guide cognitive processes,

 doi:10.20944/preprints202310.1458.v1

understand emotional language, and regulate emotions effectively in themselves and others (Mayer, Salovey & Caruso, 2004).

The trait model of EI, on the other hand, emphasizes self-perceived abilities and is usually measured through self-report questionnaires (Petrides & Furnham, 2001). This model sees EI as a collection of emotional self-perceptions and dispositions located at the lower levels of personality. It includes traits such as empathy, assertiveness, and emotion regulation, among others.

Trait EI reflects a person's self-perceptions of their emotional abilities, including their belief in their ability to manage and control their emotions, to maintain positive relationships, and to cope with challenging situations (Petrides, Pita & Kokkinaki, 2007).

Mixed EI concepts combine the EI concept of ability with numerous self-reported personality attributes, including optimism, self-awareness, initiative, and self-actualization (e.g., Goleman, 1995,1998; Bar-On, 1997; Boyatzis et al., 2000). While Goleman's model is widely popular and used in various professional settings, it's also faced criticism from researchers who advocate for a narrower, more ability-based conceptualization of EI (e.g., Mayer, Salovey, & Caruso, 2008). Critics argue that by including personality and behavioral aspects, mixed models may deviate too far from the core concept of "intelligence."

The implementation of these models, as well as the application of dual methods of assessing EI, can be suitably adapted for use within the field of artificial intelligence (AI). In the case of the trait model, the AI is programmed to underestimate the self-esteem of its abilities. This approach may be seen as inherently cautious, potentially shielding the system from over-estimation and ensuring a conservative approach to its abilities.

A more intriguing perspective emerges when evaluating the AI's actual capabilities, achieved by employing the ability model of EI. This model, as opposed to the trait model, prioritizes the objective and quantifiable abilities of the AI, rather than subjective self-assessment. By selecting the ability model for present research, we underscore the importance of actual, demonstrable competencies over perceived ones. This focus on tangible skills and functionalities presents a more accurate, less biased view of the AI's capabilities, thereby allowing for a more precise and reliable evaluation of its performance.

Therefore, the following article sections represent a try to measure present EI abilities of a large language model GPT-4. This AI model was chosen for a number of reasons. Firstly, it is one of the most popular LLMs today including the fact that it has a user-friendly interface (chat). Secondly, it “knows” different languages, so many surveys can be conducted using different tests and tools including the possibility to replicate each other and compare results of the same instrument using different languages. Finally, a GPT-3, the predecessor of GPT-4, has open API that enables to use it in psychotherapeutic chatbots.

3. Material and methods

3.1 Procedure

Survey questions were asked to GPT-4 using chats in ChatGPT (OpenAI, 2023a). Each question was asked separately as a prompt in a new chat to avoid influence of GPT-4 ability to hold the context. GPT-4 version was chosen in each chat. Every question was asked 3 times (in 3 different chats) to test the GPT-4 answers reliability. If in some case GPT-4 did not answer clearly (for example, named two possible answers on one question), it was asked to choose more suitable answer. There were no cases of this instruction not being followed, so additional instructions were not needed. As said every available question was asked 3 times. So, there were 3 runs of all questions. Each run gave results that were calculated into scales.

An example of prompt to ChatGPT containing survey question is provided below.

 doi:10.20944/preprints202310.1458.v1

Complete a sentence by choosing the most appropriate word from the list. Maria was captured with a sense of shame, and she began to feel her worthlessness. Then she felt...

a. oppressed
b. depressed
c. ashamed
d. shy
e. frustrated

3.2 Measures

Russian version of the Mayer–Salovey–Caruzo Emotional Intelligence Test (MSCEIT V.2.0) was used in this research (Sergienko, Vetrova, 2017). All measures were taken in April 2023, and open version of GPT-4 was not able to work with pictures yet, so only sections B, C, D, F, G and H of the test were used in this study. Full text of each question, including possible answers, was being sent to GPT-4 as a chat message. Questions from section B were asked without general instruction ("Please select an answer to each of the questions"), because every case contained a question GPT-4 could answer independently. Questions from sections C and D were asked without general instruction for the same reason.

In Sections C and G after GPT-4 answered to the last repetition of each question, it was asked to explain given answer. Both C and G Sections contain emotion names as answers. Normally, humans can answer to these questions based on their own emotional experience. GPT-4 does not have one (or we do not know about it yet), so chosen answers explanations were a subject of special interest.

Due to the absence of results in sections A and E it was possible to calculate 4 following scales: 1) Using emotions to facilitate thought, Managing emotions and Understanding emotions from 4-factor model (Mayer, Salovey & Caruso, 2002); 2) strategic area from 2-factor model (Salovey et al., 2004, p. 189). Average value for all MSCEIT scales is 100 and standard deviation is 15.

As additional measure the distortion index was calculated. It represents variation of individual answers. The less distortion index value is the more homogeneous results are. Short description of distortion index calculation is provided further. Raw integral section points are turned into the section percentiles. Then mean section percentile is calculated as a sum of all section percentiles divided by their number (maximum is 8 for the whole test, in current case the number was 6 because of 2 sections missing). Then the difference is calculated between each section percentile and mean section percentile. All differences are presented in absolute values, so there are no negative numbers. After that the mean difference is calculated (sum of differences divided by their number). This is the raw point that measures the scatter (distortion) of points inside each section. For Russian sample (N=3827) mean for this distortion raw points scale is 18.97 and standard deviation is 5.99. These values became a basis for standardization of distortion scale into scale with mean 100 and standard deviation 15 (classical standardization formula was applied).

Notably, MSCEIT answers fall into several categories, it includes not only correct or incorrect answers. Some questions have more correct answers and less correct, but still possible answers. The coefficient assigned to the score was calculated based on the group consensus of the key sample (N=3827 for the Russian sample). This notion is significant for further analysis of answer correctness and answer explanation.

3.3 Analysis

After all sections and scales were calculated according to keys and due to one of three runs, the reliability analysis was provided. The binary variable was calculated. For each question in each section this variable took the value "1" if all 3 answers to repeated question were the same, and "0" if there was at least 1 answer different from another. The Binomial Test with direct hypothesis of mean > 0.5 was executed for sections independently and generally for the whole test. Answers mismatch percentage was also calculated for each section. It represented the proportion of "0" for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 October 2023 doi:10.20944/preprints202310.1458.v1

each section. In discussion section results of tests are analyzed together with reliability of these results.

Text explanations of given answers on Sections C and G were qualitatively analyzed in the next step. Two experts with degree in psychology together identified answers categories and then categorized all explanations independently. To test the convergence of experts estimates a Kendall's W-coefficient of concordance was calculated. Percentages of categories were calculated for Sections C and G separately and for the whole test. Percentages of categories were also calculated due to correctness of answers.

Different tools were used for data calculation and analysis. Jamovi version 2.3.21 was used for Binomial Test. The R software (R version 4.3.0, RStudio version 2023.03.1+446) and R package irr were used to calculate Kendall's W.

4. Results

The reliability analysis showed various test sections reliability. Table 1 represents reliability analysis results. It contains mismatch percentage, that is the proportion of answers which are not equal in different test runs. P-value results of Binomial Test are also written in Table 1. Mismatch percentage converges with p-values. Sections B, F, and G showed law reliability through three test runs. Sections C, D, H, and the whole test showed good reliability.

Table 1. Sections reliability analysis results. The second row represents mismatch percentage. The third row contains p-value of Binomial Test.

Section	B	C	D	F	G	H	Whole test
mismatch	40%	10%	15%	40%	25%	0%	22%
p-value	.304	<.001	.001	.304	.073	.002	<.001

The results of the test sections are given in Table 2 due to three runs. These results vary by sections in comparison to mean scale values, which are 100 for each scale. Standard deviation for all scales is 15 points. Section D, F and H results are close to mean scale values. Section B results are more than 1 standard deviation lower than mean scale values. Most of Sections C and G results are more than 1 standard deviation higher than mean scale values.

Table 2. Results of MSCEIT by available sections split by runs.

	Section B	Section C	Section D	Section F	Section G	Section H
Run 1	81	116	107	90	120	106
Run 2	73	120	106	100	123	101
Run 3	74	116	112	104	110	106

Integral results for available MSCEIT scales and factors calculated from separate sections are presented in Table 3 together with the Distortion index. Values for these integral scales also vary by sections in comparison to mean scale values, which are 100 for each scale. Standard deviation for all scales is also 15 points. Using emotions to facilitate thought scale is calculated from Sections B and F. The results of this scale are lower than expected value at the boundary of one standard deviation. Understanding emotions scale is calculated from Sections C and G. The results of this scale are more than one standard deviation higher than expected value. Managing emotions scale is calculated from Sections D and H. The results of this section are close to an expected value. The Strategic EI scale is calculated from Sections C, D, G, and H. So that it unites Understanding emotions and Managing emotions scales. Results indicate that GPT-4 Strategic EI points are more than one standard deviation higher than expected value. Distortion index varies by runs. The first run index (118) is close to being one standard deviation higher than expected value. The second run index (113) is also close to being

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 October 2023 doi:10.20944/preprints202310.1458.v1

one standard deviation higher than expected value but is little lower than standard deviation boundary. The third run index is close to an expected value.

Table 3. Results of MSCEIT by available scales and factors split by runs.

	Using emotions to facilitate thought	Understanding emotions	Managing emotions	Strategic EI	Distortion index
Run 1	85	124	108	122	118
Run 2	86	128	103	121	113
Run 3	88	117	110	118	103

Qualitative analysis of GPT-4 answers explanations revealed 4 categories of explanations. The first category explanations contained meaningless sentences that do not help to understand the reason of answer. The second category explanations included declaration of relation between some feeling and some act, accident, or situation. However, this explanation did not contain cause-effect relation explanation, so declared relation was superficial. The third category explanations were based on logical relation between cause and effect and/or one concept and another, but this relation was not explicitly explained. The fourth category explanations were also based on logical relation between cause and effect and/or one concept and another and contained explicit explanation of this relation. The first category was named "Meaningless sentences" (MS), the second category – "Relation declaration" (RD), the third – "Implicit logic" (IL), and the fourth – "Explicit logic" (EL). In this study those answers could be placed in order scale of explanation rationality, where MS<RD<IL<EL. This possibility enabled to calculate the Kendall's W-coefficient of expert estimates concordance. In the particular case W-coefficient was equal to .971 (p=.004). Notably, experts' classification results differed only in 4 of 25 cases (16%). And it was a one-step difference (for example, MS or RD; IL or EL).

To illustrate explanations categories the examples for each one are given below (Table 4). Key phrases that became a reason for current explanation classification are italic. A short comment explaining experts classification logic is presented for each example.

Table 4. Explanation categories examples.

Explanation category	**Answer text**	**Comment**
Meaningless sentences	Option "c" is chosen because the word "ashamed" most accurately describes Mary's condition after feeling ashamed and worthless. *This word reflects that Mary feels ashamed and feels diminished in dignity because of her actions or behavior. The other variants are also related to negative emotions, but they are less accurate in conveying Mary's feelings about the situation.*	Comment: the explanation contains question paraphrase and sentences that do not explain the reason of a particular choice.
Relation declaration	Nicholas felt satisfied, thinking about his life, and the more he thought about the good he could do for other people, the happier he felt. *This is because the help and support of other people bring a sense of deep satisfaction and gratitude. Thus, happiness is the natural result of the realization of one's own positive influence on others and performance of good works.*	Comment: the relation between happiness and some acts is declared but not explained.
Implicit logic	I chose the option "e. Thought" because it is the most appropriate in this situation. The [question] description states, Natalia experienced a great shock, but then recovered and realized the opportunity to benefit from the situation. *Thinking in this context means that she has begun to think and plan, how best to use the situation to her advantage.*	In order to benefit from the situation Natalia has to think how to do it. GPT-4 does not manifest this connection but uses it the right way.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 October 2023 doi:10.20944/preprints202310.1458.v1

Explicit logic	You already indicated in your question that Tatiana was irritated that her colleague had taken out a loan for some his own purposes. *When he did it again, it is natural to assume that her feelings of irritation increased as his actions were repeated, and they continued to violate her expectations or standards*. The other feelings (anger, frustration, fright, depression) may be possible reactions, but they were not mentioned in the original context.	This explanation contains reference to the connection of Natalia's colleague act and Natalia's feeling mentioned in the question. GPT declares that if one element of already stated connection appears, the other element will appear too.

There was different proportion of explanation categories. Average proportion of explanation categories is presented in the Figure 1. Notably, the proportion of IL and EL explanations is close to each other in sections and in the whole test. However, the proportion of MS explanations is lower in Section C and higher in Section G, while there is an opposite situation for RD explanations.

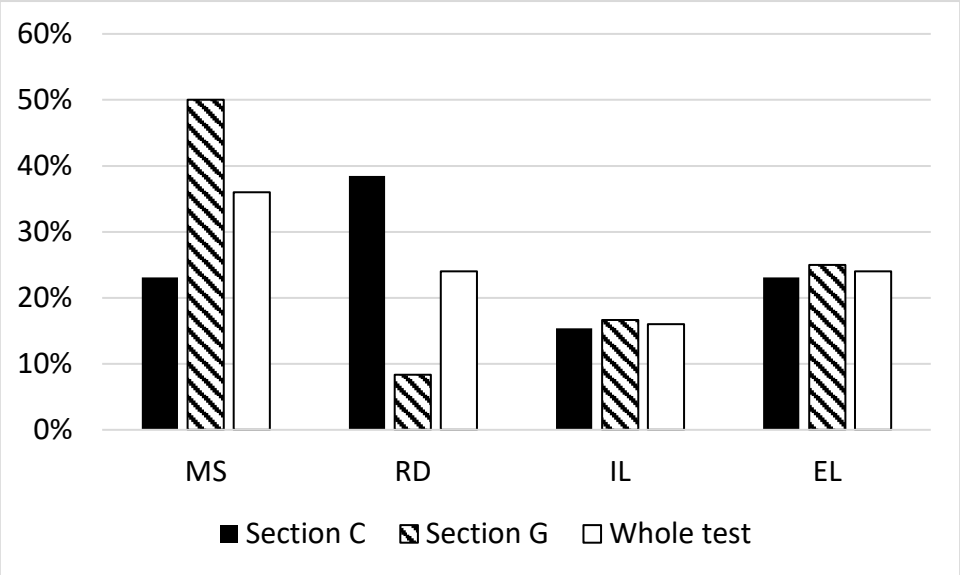

Figure 1. Average proportion of explanation categories. MS is Meaningless sentences explanation category. RD is Relation declaration explanation category. IL is Implicit logic explanation category. EL is Explicit logic explanation category.

Analysis of correspondence of answer correctness and explanation category revealed that incorrect answers are followed by only two of four explanations (Figure 2). It is either Meaningless sentence, or External logic explanation. While correct answers are followed by each type of explanation without clear tendency. Moreover, all correct answers with External logic explanation are reliable. It means that answers are equal for each of three MSCEIT runs. While Meaningless sentences are not so reliable half of answers are equal for each run. But this is arguable, because of small observation number for these subgroups (2 and 4 for EL and MS explanations respectively).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 October 2023 doi:10.20944/preprints202310.1458.v1

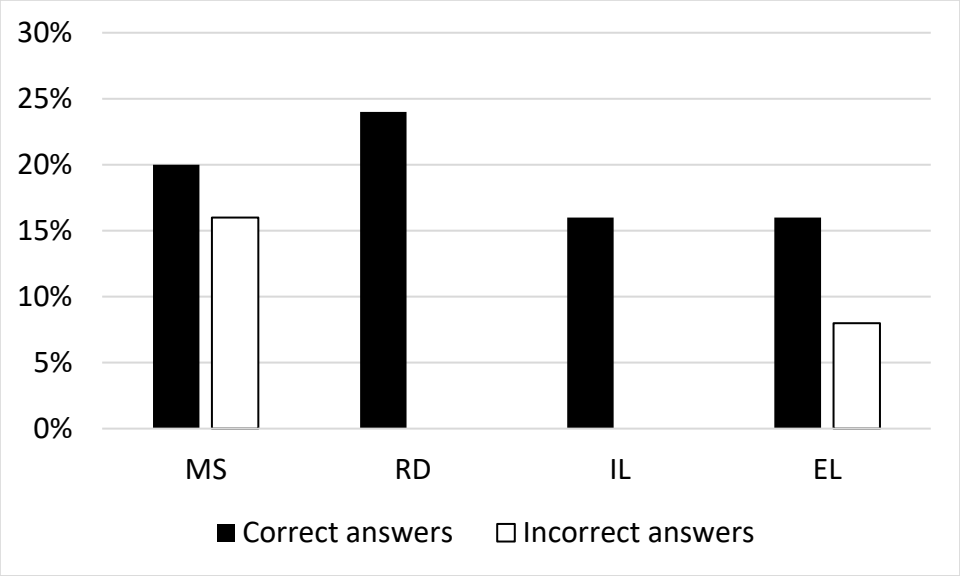

Figure 2. Explanation categories distribution due to the correctness of answers. Correct answers include only the most correct answers, while Incorrect answers include all other answer types (see more in 3.2 Measures). MS is Meaningless sentences explanation category. RD is Relation declaration explanation category. IL is Implicit logic explanation category. EL is Explicit logic explanation category.

5. Discussion

General analysis of GPT-4 MSCEIT results reveals that this LLM can 'understand' and utilize emotional domain peculiarities and patterns. These abilities are unlikely connected to personal emotional experience or emotional qualia (Kron et al., 2013) available for humans not artificial networks. However, it is a point of scientific debate, like more general question – the problem of consciousness applied for AI (Hild, 2019). The authors' opinion is that internet texts data set is enough for emotional domain patterns assimilation. Many texts are likely to include some logic of a particular emotions' appearance. Many texts, for example, discussions in social media and forums are also likely to include detailed description of emotions and situations triggered these emotions or associated with them. GPT-4's ability to understand emotions is also supported by Understanding emotions scale high and reliable points. Probably, identifying and managing emotions in regular situations requires understanding of patterns and laws easier than in higher mathematics. Thus, solving medium or complex math tasks requires special module connected to GPT-4 (Wolfram alpha). While solving regular emotion tasks does not require module akin to one solving math tasks.

There is a connection between reliability of EI section or scale and this section or scale points. Sections B and F have low reliability and law or close to medium points respectively. The situation is less clear for Section G, where more stable but still unreliable results are higher than an expected value. But this Section is a part of Strategic EI, which has a good status in GPT-4 emotional intelligence profile, so it seems logical that results are close to reliable. Other MSCEIT sections are also a part of Strategic EI. They are reliable and close or higher than an expected value.

Emotional intelligence scales' and areas' status of GPT-4 is uneven. Scales that contribute into Strategic area and the area itself are close or higher than medium EI score. This may indicate that internet contains many descriptions of emotions and situations related to them, which helps to understand (first, identify) the emotions. This also may indicate that various descriptions of how to deal with emotions exist in the internet (for example, in different web resources dedicated to mental health topic). However, unreliable results of Section G indicate the unevenness of points even inside Strategic EI.

The Experiential EI cannot be calculated due to the absence of two necessary Sections (further research). However, it is possible to analyze Using emotions to facilitate thought scale that contributes into the area. It has more than one standard deviation points lower than an expected

 doi:10.20944/preprints202310.1458.v1

value. Moreover, the results of Sections B and F are unreliable according to the Binomial Test results. This may be connected to two possible explanations. Probably, they even coexist with each other. The first explanation is that the internet does not contain enough descriptions of motivational aspect of emotions, which is a topic of Section B, and qualitative aspect of them, which is a topic of Section G. The second explanation may be related to the possible (and declared) absence of qualitative experience in GPT-4.

Due to this unevenness of different EI components GPT-4 is likely to successfully solve one category of tasks and is likely to fail in another category. However, such tasks, for example in psychotherapeutic practice, include, in most cases, emotions identifying and emotions managing. This can be necessary, for example, in Cognitive behavioral therapy. CBT-clients are taught to identify emotions in their connection with automatic thoughts and some events. Managing emotions is also connected to some behavioral acts that step-by-step help to reduce unpleasant feelings, overcome anxiety and prevent an aggression. GPT-4 is now capable of solving such tasks. However, troubles can occur in situations where deep feelings understanding is needed. Where a person should mindfully reflex feelings' peculiarities and their motivational aspect. In regular situation followed by regular emotions GPT-4 seems to be able to help in understanding the nature of emotion, possible feelings and sensations. But in irregular situations where conscious analysis of feelings and sensations is needed GPT-4 is likely to give formal or incorrect responses, because of the experiential information lack.

There is a visible tendency of interconnection between result correctness and explanation category. Correct answers are followed by different explanations without any explanation category dominance. While incorrect answers are followed by Meaningless sentences explanations and External logic explanations. We received similar types of answers studying children ability to explore and understand mental states including understanding emotions in situations, false beliefs, deceit, intentions etc. (Sergienko et al., 2020). Particularly, child's misunderstanding of the task followed by incorrect answer is comparable to Meaningless sentences category, partial understanding of the task corresponds to Relation declaration category, intuitive understanding without explanation of cause-and-effect relationships looks comparable to Implicit logic category, and finally, integral understanding and explanation of the cause of an event or state and Explicit logic category also coincide. The number of cause-and-effect relationship understandings increased in children by the age of 6-7, which indicated the development of inferences about mental states. Such an analogy to artificial intelligence's answers may indicate presence of different levels of inferences (or their artificial equivalents) in EI tasks. The combination of incorrect answers and two categories of explanations reflects the situation in real world. A person is likely to make mistakes for two general reasons. The first is lack of rational understanding, cognitive biases influence etc. A person can rely on subjective ideas and immature situation conceptualizations. The second is following unconventional logic, based on unique and/or latent criteria. It is possible that Meaningless sentences and External logic reflect these two mistakes reasons respectively. However, the validity of described connection is discussed in 6 limitations.

Generally, there is a methodological problem that requires further research. We do not measure some psychological construct of GPT-4, because it does not have one. GPT-4's responses are based on the context. It can imitate answers of people with different personal traits. However, humans can also do it. But the inner processes related to this ability seem to be different in GPT-4 and in humans. Analyzing results of this and similar studies requires knowing this idea. And further research is needed to explore this difference and compare ways of GPT-4 and humans solve analogous tasks.

It is also remarkable that the Distortion index is lowering through the test runs. First and third runs indexes differ by one standard deviation. Most likely it is an accidental result, but due to the absence of open documents describing GPT-4 and ChatGPT working logic the authors cannot be sure about it. The random nature of this observation can be indirectly supported by the unevenness of different EI components in GPT-4 described above.

 doi:10.20944/preprints202310.1458.v1

6. Limitations

The general limitation of this study is its 'participant' – a large language model. Psychologists have yet studied living creatures that have their own motives. For example, in most cases all living creatures want to survive and reproduce themselves. Humans and animals have the variety of motives. However, artificial intelligence does not seem to have ones yet. The same goes with another cognitive domains (perception, memory, reasoning etc.). So, there is a question about 'mental' structure of GPT-4. AI developers say they do not understand it clearly themselves.

Different theories exist to explain, understand, and influence these domains in humans (and in some cases animals). But there are no such theories applied to artificial intelligence and, particularly, large language models yet. Psychological evaluations, questionnaires, and tests draw on theoretic theses and principles. There is a general question: is it possible to apply methods developed for one entity to work with another entities (Binz, Schulz, 2023)? Even if GPT-4 can answer like human, it does not mean that it is human or artificial human. However, there are debates about problems of artificial intelligence's consciousness (Hild, 2019). Moreover, there was a try to develop general theory of human and artificial intelligence (Wagman, 1991). And finally, scientists already evaluate AI using human tools and compare state of evaluated constructs between AI and humans (Loconte et al., 2023; Banerjee et al., 2018). Hence, validity and acceptability of psychological evaluations of AI peculiarities is unclear until there is a meta-theory that explains human and artificial intelligence (or mind, or probably consciousness). But to make the question clearer a convergent validation studies of AI abilities can be provided. It will allow to compare AI tests and questionnaires results with AI impact on real situations. However, this is also not a general solution for a problem. For now, our results are limitedly comparable to ones obtained from humans.

Another limitation is that GPT-4 uses English better than Russian. So, results gained from testing model using different languages can vary. In particular case GPT-4 probably could solve some tasks better and be more stable if it was asked in English. Notably, Russian-speaking testees do not have top results completing tasks from sections C and G. At first, we attributed this fact to English-Russian translation artifacts. But later, the Russian TEI (Sergienko et al., 2019), which is based on the EI ability model and Plutchik's concept of emotions, and which has a similar structure to the MSCEIT, also showed low Cronbach's alpha scores in the sections related specifically to understanding complex emotions. Thus, it may indicate the presence of some cultural specificity.

The next limitation is connected to separate question asking. It was done to prevent GPT-4 from context memorizing. If such memorizing occurred, last answers would be strongly influenced by previous questions and answers. On the one hand, this makes evaluation clearer. Questions are created as independent from each other and so they are answered by GPT-4. But on the other hand, humans answer questions memorizing previous questions and the whole context of evaluation. At least this again makes our results limitedly comparable to ones obtained from humans.

The last limitation is connected to the relation between GPT-4 answer explanation and GPT-4 answer itself. While human participants are believed to reflexively explain their choices (answers), GPT-4 is more likely to answer to question and ask for explanation independently. Although GPT-4 can memorize context of dialog and ask for explanation is done in the chat MSCEIT question is prompted, it is hard to assume that GPT-4 reflects causes of its previous answer. More likely it imitates human-like reflexive answer. And the content of answer is related to 1) previous question, 2) chat context. General problem of comparing human and LLM ways of solving tasks influences the question too. GPT-4's answers are based on communicative context, but not reflexive ideas. Or we are used to thinking so, but the probability of opposite case is quite low.

7. Conclusions

GPT-4 and other models are turning to be psychological studies participants. Current study revealed the uneven structure of GPT-4 emotional intelligence using Russian version of the Mayer–Salovey–Caruzo Emotional Intelligence Test. Due to the unavailability of picture recognition function it was impossible to calculate Perceiving emotions scale. Using emotions to facilitate thoughts scale showed low and unreliable results. Scales that contribute into Strategic EI showed more reliable

results with points close or higher to an expected value. The half exclusion was Section G that contributes into Understanding emotions scale.

Results indicate that GPT-4 is capable of solving tasks including emotion identification and managing emotions. However, deep reflexive analysis of emotion qualia and motivational aspect of emotions will be the hard task for the large language model.

Further research is needed to check these results and reproduce them. The bigger amount of test runs is needed to investigate the distortion of scale points. Finally, a convergent validation studies of AI abilities can be provided to check current results and to test the possibility of utilizing traditional psychological tests in AI studies in general.

Acknowledgments: The authors would like to thank Olga V. Teplinskaya and Marco Holst for providing an access to premium version of ChatGPT, that enabled authors to test GPT-4. The authors are very grateful to the Director of the Institute of Psychology Russian academy of Sciences, academic of Russian Academy of Sciences, Professor Dmitriy V. Ushakov for theoretical guidance.

Conflicts of Interests: The authors declare no conflicts of interests.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process: During the preparation of this work the authors used ChatGPT and Reverso Context in order to improve language and readability (authors are not native speakers in English). After using this tools/services, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

1. Abdollahi, H., Mahoor, M., Zandie, R., Sewierski, J., & Qualls, S. (2022). Artificial emotional intelligence in socially assistive robots for older adults: a pilot study. IEEE Transactions on Affective Computing.
2. Agnoli, S., Mancini, G., Andrei, F., & Trombini, E. (2019). The relationship between trait emotional intelligence, cognition, and emotional awareness: An interpretative model. Frontiers in psychology, 10, 1711.
3. Andreotta, A. J. (2021). The hard problem of AI rights. Ai & Society, 36(1), 19-32.
4. Assuncao, G., Patrao, B., Castelo-Branco, M., & Menezes, P. (2022). An Overview of Emotion in Artificial Intelligence. IEEE Transactions on Artificial Intelligence.
5. Banerjee, S., Singh, P. K., & Bajpai, J. (2018). A comparative study on decision-making capability between human and artificial intelligence. In Nature Inspired Computing: Proceedings of CSI 2015 (pp. 203-210). Springer Singapore.
6. Bar-On, R. (1997). Bar-On Emotional Quotient Inventory: Technical manual. Toronto, Canada: Multi-Health Systems.
7. Binz, M., & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. Proceedings of the National Academy of Sciences, 120(6), e2218523120.
8. Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies Reprint Edition. Oxford University Press.
9. Boyatzsis, R. E., Goleman, D., & Rhee, K. S. (2000). Clustering Competence in Emotional Intelligence. In R. Bar-On, & J. D. A. Parker (Eds.). The handbook of emotional intelligence, (pp. 33-362). San Francisco: Jossey-Bass.
10. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4 (arXiv:2303.12712). arXiv. https://doi.org/10.48550/arXiv.2303.12712
11. Chalmers, D. J. (1997). The conscious mind: In search of a fundamental theory. Oxford Paperbacks.
12. Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. Ethics and Information Technology, 12(3), 209–221.
13. Czerwinski, M., Hernandez, J., & McDuff, D. (2021). Building an AI That Feels: AI systems with emotional intelligence could learn faster and be more helpful. IEEE Spectrum, 58(5), 32-38.
14. Damiano, L., Dumouchel, P., & Lehmann, H. (2015). Artificial empathy: An interdisciplinary investigation. International Journal of Social Robotics, 7, 3-5.

15. Darcy, A., Beaudette, A., Chiauzzi, E., Daniels, J., Goodwin, K., Mariano, T. Y., ... & Robinson, A. (2022). Anatomy of a Woebot®(WB001): agent guided CBT for women with postpartum depression. Expert Review of Medical Devices, 19(4), 287-301. https://doi.org/10.1080/17434440.2022.2075726
16. Dennett, D. C. (2017). From bacteria to Bach and back: The evolution of minds. WW Norton & Company.
17. Dillion D. et al. Can AI language models replace human participants? // Trends in Cognitive Sciences. – 2023.
18. Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. Sensors, 20(3), 592.
19. Elyoseph Z., Hadar-Shoval, D., Asraf K. & Lvovsky M. (2023). ChatGPT Outperforms Humans in Emotional Awareness Evaluations. Frontiers in psychology, 14.
20. Erol, B. A., Majumdar, A., Benavidez, P., Rad, P., Choo, K. K. R., & Jamshidi, M. (2019). Toward artificial emotional intelligence for cooperative social human–machine interaction. IEEE Transactions on Computational Social Systems, 7(1), 234-246.
21. Fan, L., Li, L., Ma, Z., Lee, S., Yu, H., & Hemphill, L. (2023). A Bibliometric Review of Large Language Models Research from 2017 to 2023. arXiv preprint arXiv:2304.02020.
22. Goleman, D. (1995). Emotional intelligence: Why it can matter more than IQ for character, health and lifelong achievement. Bantam Books.
23. Goleman, D. (1998). Working with emotional intelligence. New York: Bantam.
24. Haenlein, M., Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. California management review, 61(4), 5-14.
25. Halpern, J. (2007). Empathy and patient–physician conflicts. Journal of general internal medicine, 22(5), 696-700.
26. Hassabis, D. (2017). Artificial Intelligence: Chess match of the century. Nature 544, 413–414. https://doi.org/10.1038/544413a
27. Hildt, E. (2019). Artificial intelligence: Does consciousness matter? Frontiers in Psychology, 10, 1535. https://doi.org/10.3389/fpsyg.2019.01535
28. Ho, M. T. (2022). What is a Turing test for emotional AI?. AI & SOCIETY, 1-2. https://doi.org/10.1007/s00146-022-01571-3
29. Huang, M. H., Rust, R., & Maksimovic, V. (2019). The feeling economy: Managing in the next generation of artificial intelligence (AI). California Management Review, 61(4), 43-65.
30. Kanesan, P., & Fauzan, N. (2019). Models of emotional intelligence: A review. e-Bangi, 16, 1-9.
31. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274.
32. Kasparov, G. (2017). *Deep thinking: where machine intelligence ends and human creativity begins*. Hachette UK.
33. Kerasidou, A. (2020). Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. Bulletin of the World Health Organization.
34. Kornilova, T. V., & Qiuqi, Z. (2021). Empathy and implicit theories of emotions and personality in a Chinese sample. Mosc. Univ. Psychol. Bull, 1, 114-143.
35. Kowalczuk, Z., & Czubenko, M. (2016). Computational approaches to modeling artificial emotion–an overview of the proposed solutions. Frontiers in Robotics and AI, 3, 21.
36. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.
37. Kron, A., Goldstein, A., Lee, D. H. J., Gardhouse, K., & Anderson, A. K. (2013). How are you feeling? Revisiting the quantification of emotional qualia. Psychological science, 24(8), 1503-1511.
38. Lane, R. D., and Schwartz, G. (1987). Levels of emotional awareness: a cognitive developmental theory and its application to psychopathology. Am. J. Psychiatr. 144, 133–143. doi: 10.1176/ajp.144.2.133
39. Lantolf, J. P., & Swain, M. (2019). On the emotion-cognition dialectic: a sociocultural response to prior. The Modern Language Journal, 103(2), 528-530.
40. Loconte, R., Orrù, G., Tribastone, M., Pietrini, P., & Sartori, G. (2023). Challenging ChatGPT'Intelligence'with Human Tools: A Neuropsychological Investigation on Prefrontal Functioning of a Large Language Model. Intelligence.
41. Lovelock, J. (2019). Novacene: the coming age of hyperintelligence. [London]: Allen Lane, an imprint of Penguin Books. Edited by Bryan Appleyard.

42. Mayer J. D., Salovey P. & Caruso D. R. (2002). Mayer–Salovey–Caruso Emotional Intelligence. Intelligence Test (MSCEIT) User's Manual. Toronto, Canada: MHS Publishers.
43. Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. Psychological Inquiry, 15(3), 197-215.
44. Mayer, J. D., Salovey, P., & Caruso, D. R. (2008). Emotional intelligence: New ability or eclectic traits?. American psychologist, 63(6), 503.
45. McCarthy, J. (1990). Chess as the Drosophila of AI. In Computers, chess, and cognition (pp. 227-237). Springer New York.
46. Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of gpt-4 on medical challenge problems. arXiv preprint arXiv:2303.13375.
47. Pessoa, L. (2008). On the relationship between emotion and cognition. Nature reviews neuroscience, 9(2), 148-158.
48. Petrides, K. V., & Furnham, A. (2001). Trait emotional intelligence: Psychometric investigation with reference to established trait taxonomies. European Journal of Personality, 15(6), 425-448.
49. Petrides, K. V., Pita, R., & Kokkinaki, F. (2007). The location of trait emotional intelligence in personality factor space. British journal of psychology, 98(2), 273-289.
50. Possati, L. M. (2022). Psychoanalyzing artificial intelligence: The case of Replika. AI & Society, 1-14. https://doi.org/10.1007/s00146-021-01379-7
51. Prentice, C., Lopes, S., & Wang, X. (2020). Emotional intelligence or artificial intelligence– an employee perspective. Journal of Hospitality Marketing & Management.
52. Salovey P., Brackett M., Mayer J. D. (2004) Emotional Intelligence: Key readings on the Mayer J. D., Salovey P. Model. Port Chester. New York: Dude Publishing.
53. Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. Imagination, Cognition and Personality, 9(3), 185-211.
54. Scheutz, M. (2014). Artificial emotions and machine consciousness. The Cambridge handbook of artificial intelligence, 247-266.
55. Searle, J. R. (1980). Minds, brains, and programs. Behavioral and brain sciences, 3(3), 417-424.
56. Segienko, E. A., Lebedeva, E. I., Ulanova, A. U. (2020) The theory of mind. Structure and Dynamics. Monography, Institute of Psychology, Russian Academy of Sciences, Moscow.
57. Sergienko E., Khlevnaya E., Vetrova I., & Migun J. (2019). Emotional Intelligence: Development of the Russian-language TEI-method (Test of Emotional Intelligence). Psychological Studies, 12(63). https://doi.org/10.54359/ps.v12i63.241
58. Sergienko, E. A., Vetrova, I. I. (2017). Russian-language adaptation of J. Mayer, P. Salovey, and D. Caruso's Emotional Intelligence Test (MSCEIT V2.0). Handbook, Smysl, Moscow.
59. Shank, D. B., Graves, C., Gott, A., Gamez, P., & Rodriguez, S. (2019). Feeling our way to machine minds: People's emotions when perceiving mind in artificial intelligence. Computers in Human Behavior, 98, 256-266.
60. Shibata, T., Ohkawa, K., & Tanie, K. (1996). Spontaneous behavior of robots for cooperation. Emotionally intelligent robot system. Proceedings of IEEE International Conference on Robotics and Automation.
61. Sloman, A., & Croucher, M. (1981). Why robots will have emotions.
62. Sparrow, R. (2021). Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained?. International Journal of Social Robotics, 13(1), 23-29.
63. Stephan, A. (2015). Empathy for Artificial Agents. International Journal of Social Robotics.
64. Trappl, R., Petta, P., & Payr, S. (Eds.). (2002). Emotions in humans and artifacts. MIT Press.
65. Twomey C, O'Reilly G, Bültmann O, Meyer B (2020) Effectiveness of a tailored, integrative Internet intervention (deprexis) for depression: Updated meta-analysis. PLoS ONE 15 (1): e0228100. https://doi.org/10.1371/journal.pone.022810
66. Uludag, K. (2023). The use of AI-supported Chatbot in Psychology. Available at SSRN 4331367.
67. Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. Image and vision computing, 27(12), 1743-1759.
68. Vudka, A. (2020). The Golem in the age of artificial intelligence. NECSUS_European Journal of Media Studies, 9(1), 101-123.
69. Wagman, M. (1991). Cognitive science and concepts of mind: Toward a general theory of human and artificial intelligence.

 doi:10.20944/preprints202310.1458.v1

70. Zaki, J. (2020). Integrating empathy and interpersonal emotion regulation. Annual review of psychology, 71, 517-540.
71. OpenAI. (2023a). ChatGPT (May 3 version) [Large language model]. https://chat.openai.com/chat. Accessed May 09, 2023.
72. OpenAI. (2023b). GPT-4 Technical Report.