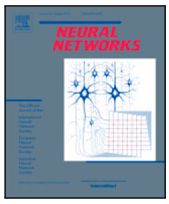
Neural Networks 187 (2025) 107362

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet



Full Length Article

A neural approach to the Turing Test: The role of emotions

Rita Pizzi [a], Hao Quan [b,*], Matteo Matteucci [b], Simone Mentasti [b], Roberto Sassi [a]

[a] Biomedical Image and Signal Processing Lab, Department of Computer Science, Università degli Studi di Milano, Via Celoria 18, Milan, 20133, Italy
[b] Artificial Intelligence and Robotics Lab, Department of Electronics, Information and Bioengineering, Politecnico di Milano, Via Ponzio 34/5, Milan, 20133, Italy

ARTICLE INFO

Keywords:
Turing Test
Emotion recognition
EEG analysis
Artificial Intelligence
Natural language processing
Qualia

ABSTRACT

As is well known, the Turing Test proposes the possibility of distinguishing the behavior of a machine from that of a human being through an experimental session. The Turing Test assesses whether a person asking questions to two different entities, can tell from their answers which of them is the human being and which is the machine. With the progress of Artificial Intelligence, the number of contexts in which the capacities of response of a machine will be indistinguishable from those of a human being is expected to increase rapidly. In order to configure a Turing Test in which it is possible to distinguish human behavior from machine behavior independently from the advances of Artificial Intelligence, at least in the short-medium term, it would be important to base it not on the differences between man and machine in terms of performance and dialogue capacity, but on some specific characteristic of the human mind that cannot be reproduced by the machine even in principle. We studied a new kind of test based on the hypothesis that such characteristic of the human mind exists and can be made experimentally evident. This peculiar characteristic is the emotional content of human cognition and, more specifically, its link with memory enhancement. To validate this hypothesis we recorded the EEG signals of 39 subjects that underwent a specific test and analyzed their signals with a neural network able to label similar signal patterns with similar binary codes. The results showed that, with a statistically significant difference, the test participants more easily recognized images associated in the past with an emotional reaction than those not associated with such a reaction. This distinction in our view is not accessible to a software system, even AI-based, and a Turing Test based on this feature of the mind may make distinguishable human versus machine responses.

1. Introduction

1.1. The Turing Test

Alan Turing proposed the Turing Test in his 1950 work "Computing Machinery and Intelligence" (Hernández-Orallo, 2020; Müller & Ayesh, 2012; Oppy & Dowe, 2003; Proudfoot, 2020; Sterrett, 2000; Turing, 1950). Turing describes the so-called "imitation game", played by three people: a man(A), a woman(B) and an interrogator (C). The interrogator stays in a different room apart from the other two. The main goal of the interrogator is to ask questions to a man (A) and a woman (B) room in order to figure out what the gender of the two players is. The interrogator has the right to ask questions to A and B using a teleprinter. Then Alan Turing poses a new question: "What would happen if a machine takes part in place of man (A)?" would a machine be able to trick the interrogator as often as man (A) in this game? Following Turing, the answer would address the question "Can machines think?". Turing's test has proved to be both very influential and much criticized since its introduction, and it has become a key notion in AI theory.

1.2. Language Turing Test from large language model (LLM)

Natural Language Processing (NLP) is a field that combines computer science and linguistics to help machines understand and respond to human language in a way that is both meaningful and valuable (Chowdhary & Chowdhary, 2020; Nadkarni, Ohno-Machado, & Chapman, 2011). This field includes tasks such as translating languages, figuring out the feeling behind the text, and summarizing articles (Liddy, 2001). A big step in NLP has been the creation of large language models (LLMs), which can use a great number of data to create texts that sounds like a human wrote them. These models seem to understand the rules, meanings, and context of language. These models have led as an application to the development of advanced chatbots, which can have meaningful conversations with people, answer questions, create creative content, and do other language-related tasks. LLMs and chatbots have greatly improved communication between humans and machines, changing many industries and parts of our daily lives (Baby, Khan, & Swathi, 2017; Cahn, 2017; Nagarhalli, Vaze, & Rana, 2020).

* Corresponding author.
E-mail address: hao.quan@polimi.it (H. Quan).

https://doi.org/10.1016/j.neunet.2025.107362
Received 5 March 2024; Received in revised form 28 February 2025; Accepted 4 March 2025
Available online 14 March 2025

Going back in time, Joseph Weizenbaum created the first chatbot, Eliza, a mock Rogerian psychotherapist, in 1966 (Weizenbaum, 1966). In 1972, Kenneth Colby developed a PARRY, which aimed at simulating a patient with paranoid schizophrenia by incorporating a strategy for conversation (Colby, 2013). Since then, many products like Microsoft Bot Framework (Biswas & Biswas, 2018), IBM Watson Assistant (IBM, 2023), Amazon Lex (Amazon, 2023), and Google LaMDA (Thoppilan et al., 2022) have been created. ChatGPT and GPT-4, both by OpenAI (OpenAI, 2023), are two of the most well-known products today. ChatGPT, which was released in November 2022, was based on GPT-3 models (at the time of this writing, the release is GPT-3.5), *i.e.* third-generation Generative Pre-trained Transformers, neural network machine learning models trained using supervised and reinforcement learning on internet data to generate any text. The way ChatGPT answered questions across many different topics was impressive (Wu et al., 2023).

On March 14, 2023, OpenAI released GPT-4. GPT-4 can handle more text, is more accurate, can create songs and other creative text in different styles, and can even recognize and generate images (Kocoń, Cichecki, Kaszyca, Kochanek, Szydło, Baran, Bielaniewicz, Gruza, Janz, Kanclerz, et al., 2023), giving accurate text responses. Tests have shown that GPT-4 can perform as well as humans on a variety of professional and academic tasks, like writing code (Liu et al., 2023). For example, GPT-4 passed a simulated law exam with a score in the top 10%, while GPT-3.5 scored in the bottom 10%. GPT-4 is also good at scientific writing and structuring information, which is hard for other NLP models (Elkins & Chun, 2020; Turing, 2009). Additionally, GPT-4 can understand images and can describe them in detail.

Both ChatGPT and GPT 4 were developed by transformer architectures, initially trained on large-scale datasets to grasp language patterns, and then fine-tuned with human feedback for conversational skills. They generate responses by sequentially predicting words, maintaining context from the conversation, and adhering to ethical guidelines to ensure appropriateness and safety in their interactions (Egli, 2023).

Researchers have tested GPT-4's ability to understand and create text in many of the world's major languages. They looked at how well the model could create sentences that made sense together, followed normal storytelling rules, understood complex questions, and recognized emotions, irony, and basic worldviews to make logical conclusions. In short, both ChatGPT and GPT-4 have reached a high level of skill in understanding written text and images and often surpass many people's writing abilities.

Even though ChatGPT and GPT-4 reached an impressive level, they still have some trouble in passing the Turing Test: basically, while they are really good at many tasks, they have a hard time with tough math problems like logic, algebra, and calculus. Lyu et al. (2023).

Moreover, the measures in place to prevent unethical responses are not foolproof Ray (2023). While the model might not respond to explicit queries about illegal activities, such as bank robbery, it can sometimes provide a fictional narrative on the same topic when prompted differently. Another limitation is their inability to generate new scientific experimental ideas for the research community (Farrokhnia, Banihashem, Noroozi, & Wals, 2023), a realm where human researchers still reign supreme.

The accessibility of these tools also varies significantly. While the free version of ChatGPT can experience latency and slower response times during peak periods, the "PLUS" version of GPT-4, which comes with image understanding features, is not freely accessible (Salvagno, Taccone, Gerli, et al., 2023).

Furthermore, the current version of GPT-4 can process image data but lacks the capability to handle video inputs, something that humans can do effortlessly. The absence of real-time interactive visual and auditory communication is another significant hurdle, making ChatGPT seem less human-like in its interactions.

Moreover, ChatGPT lacks knowledge of users' personal lives, including details about their parents, relatives, and friends. The legal aspects surrounding the use of training data by OpenAI and the potential risk of personal data leakage during user interactions are additional challenges that need to be addressed. Until these legal and ethical questions are adequately resolved, the capacity of ChatGPT to successfully face the Turing Test remains in question (Lecler, Duron, & Soyer, 2023).

In light of these shortcomings, it is posited that while the latest LLMs exhibit proficiency in generating contextually relevant and coherent responses, they fall short of emulating the intricacies and complexities inherent in human language and behavior, including emotion-driven expressions, both textual and facial, and bodily gestures (Sallam, Salim, Barakat, & Al-Tammemi, 2023).

1.3. Visual Turing Test

Another area of ML with very rapid progress is image or video recognition, another area for which a Turing test may soon prove challenging. A Visual Turing Test (Floridi & Chiriatti, 2020; Geman, Geman, Hallonquist, & Younes, 2015) has been used to assess how computer vision systems interpret images in comparison to humans. Computer vision research is a very rapidly developing sector, with the objective to create systems that are as near to the human visual system as possible. Nevertheless, some recent results show that computers still have difficulties passing the visual Turing test (Floridi & Chiriatti, 2020).

1.4. Reverse Turing Test – CAPTCHA

CAPTCHA stands for "Completely Automated Public Turing Test to tell Computers and Humans Apart". The concept was invented by Luis von Ahn, Manuel Blum, Nicholas J. Hopper, and John Langford in 2003 (Von Ahn, Blum, Hopper, & Langford, 2003). Its primary function is to assess whether the user is human or a machine, safeguarding against automated bot attacks and spam activities. The first version of CAPTCHA was suggested in 1997. It requires users to enter a sequence of numbers or letters provided in a distorted picture. Unlike the traditional Turing Test, which is administrated by people, this test is administrated by computers, so CAPTCHAs are usually called Reverse Turing tests (Gossweiler, Kamvar, & Baluja, 2009).

In general, there are three types of CAPTCHA (Guerar, Verderame, Migliardi, Palmieri, & Merlo, 2021; Kumar, Jindal, & Kumar, 2022; Madathil, Greenstein, & Horan, 2019; Singh & Pal, 2014; Xu, Liu, & Li, 2020), *i.e.*, text, image, and audio. In the first releases, humans were authenticated using text-based CAPTCHAs, which could either employ familiar words or phrases or arbitrary sequences of letters and numbers. Some text-based CAPTCHAs also incorporated alterations in capitalization (Bursztein, Martin, & Mitchell, 2011; Wang, Gao et al., 2023; Zhang, Ebrahimi, Li, & Chen, 2022). To supplant text-based CAPTCHAs, image-based ones were created. These CAPTCHAs employ identifiable graphical components, such as animal photographs, shapes, or landscapes. Generally, image-based CAPTCHAs mandate users to choose images that correspond to a certain theme or to detect images that do not fit (Alqahtani & Alsulaiman, 2020; Madathil et al., 2019; Sukhani, Sawant, Maniar, & Pawar, 2021). As a substitute that facilitates visually impaired users, audio CAPTCHAs were created. These CAPTCHAs are frequently used in conjunction with text or image-based ones. Audio CAPTCHAs offer an audio rendition of a sequence of letters or numbers that the user must enter (Alnfiai, 2020; Choudhary, Saroha, Dahiya, & Choudhary, 2013; Gao, Liu, Yao, Liu, & Aickelin, 2010).

Any software able to pass the CAPTCHAs tests may be, in principle, utilized to solve difficult unsolved AI issues. In the case of picture and text-based CAPTCHAs, if an AI could complete the challenge successfully without exploiting faults in a specific CAPTCHA design, it would have solved the difficulty of constructing an AI capable of sophisticated object identification in scenes. It is expected that the Reverse Turing Test in maybe 10–15 years could fail in the network security, and a machine will be able to break the security (Chow, Susilo, & Thorncharoensri, 2019; Dinh & Hoang, 2023; Guerar et al., 2021).

1.5. Emotional Turing Test

1.5.1. Current emotion recognition approaches

Current approaches to emotion recognition can be broadly categorized into two main fields: sentiment analysis and affective computing. Sentiment analysis focuses specifically on computational text analysis to determine emotional orientation and opinions (Bordoloi & Biswas, 2023; Cambria, Schuller, Xia, & Havasi, 2013; Hamborg, Donnay, Merlo, et al., 2021; Sharma, Ali, & Kabir, 2024; Zhang, Deng, Liu, Pan, & Bing, 2023).

In contrast, affective computing encompasses a broader multimodal approach, processing multiple data streams simultaneously, coming from facial expressions, body language, vocal nuances, and physiological signals. Recent advances in deep learning have significantly improved both fields, but significant challenges remain.

For example, with regard to video processing in affective computing, currently the search for features using moving images in general still originates from static techniques, while it is important to be able to follow the dynamics of microexpressions that contribute to expressiveness (Li, Zhan, Xu, & Wu, 2019; Wang, Li et al., 2023).

The same problem is found in speech emotion recognition, for which work is being done on handling time synchronization of stacked networks. Here, too, further progress is needed (Al-Dujaili & Ebrahimi-Moghadam, 2023; Bhangale & Kothandaraman, 2023; de Lope & Graña, 2023; Li & Deng, 2020; Mellouk & Handouzi, 2020; Shen, Liu, Wang, Wang, & Zhou, 2024).

Multimodal analysis technology combines the search for emotional features in the subject's expressions and speech with that obtained through the analysis of physiological signals such as EEG (Electroencephalography), fMRI (functional Magnetic Resonance Imaging), ECG (Electrocardiogram), HRV (Heart Rate Variability), EMG (Electromyography), GSR (Galvanic Skin Response) as well as gesture analysis and other data sources.

Several methods have been studied for the extraction of emotional features from EEG (Geng, Shi, & Hao, 2024; Liu, 2024). A more advanced method is the "cross-perception" that captures and integrates the complementary information provided by EEG and fMRI data (Carmichael et al., 2024; Huster, Debener, Eichele, & Herrmann, 2012; Qin, Zong, & Liu, 2024).

Recently, multimodal approaches that integrate diverse non-homogeneous sources of information – including video, voice, and physiological signals – have been developed, proving particularly effective in emotion recognition tasks (Ahmed, Al Aghbari, & Girija, 2023; Bhatlawande, Shilaskar, Pramanik, & Sole, 2024; Fu et al., 2021; Gao, 2024; Gao, Li, Chen, & Zhang, 2020; Gohumpu, Xue, & Bao, 2023; Haque et al., 2024; Hatipoglu Yilmaz, Kose, & Yilmaz, 2024; Ilyas, Nunes, Nasrollahi, Rehm, & Moeslund, 2021; Kraack, 2024; Lian et al., 2023; Liu et al., 2023; Luo et al., 2022; Medjden, Ahmed, & Lataifeh, 2020; Saxena, Khanna, & Gupta, 2020; Wu & Li, 2023).

An advanced multimodal dataset is DEAP (Gong, Jia, Wang, Zhou, & Zhang, 2023; Koelstra et al., 2011) that collects signals from EEG, respiration amplitude, skin temperature, blood volume pressure by plethysmograph, ECG, HRV, EMG, EOC (electrooculogram), functional Near Infrared Spectroscopy (fNIRS), ECG, GSR, EMG. Frontal face videos are also recorded.

1.5.2. Applications

Despite current limitations, this field of research has already resulted in wide-ranging applications. In healthcare, these technologies enable better patient care and pain detection (Chubarov & Azarnov, 2018; Fei et al., 2020; Kusal, Patil, Kotecha, Aluvalu, & Varadarajan, 2021). Educational applications focus on student engagement assessment (Chen & Wang, 2011; Gehl, 2013; Wang, Xu, Niu, & Miao, 2020), while marketing applications analyze consumer reactions (Natale, 2021; Stark, 2019). The latest developments also enhance human–computer interaction (Jacquet, Jamet, & Baratgin, 2021; Neufeld & Finnestad, 2020; Wheeler, 2020).

Models of emotion analysis are already available that can be used to examine the personality of subjects. The paper (Liu et al., 2022) proposes a cross fertilization between deep learning and emotional psychology. It acquires expressive features through video, and maps the outputs through several well-known emotion classification systems, showing about 80% agreement with current personality tests.

Machine Learning methods are used in composition to process voice traces, images, text, and physiological signals (BCI based emotion recognition) (Erat et al., 2024; Floreani, Orlandi, & Chau, 2022; Nagels-Coune, Riecke, Benitez-Andonegui, Klinkhammer, Goebel, De Weerd, Lührs, & Sorger, 2021) and by integrating we arrive at not only classification but also quantification, up to individual trait identification, which is useful in computational psychology and psychiatry for identification and quantification of traits such as depression, anxiety, stress, facilitates monitoring, diagnosis and treatment (Baig & Kavakli, 2019; Huang, 2021; Li et al., 2019; Liu, 2024; Pan et al., 2022; Xiao & Wu, 2023).

1.5.3. Taxonomy of emotions

While sentiment analysis leverages text-based models like Plutchik's wheel of emotions, which identifies 8 primary and 24 complex emotions, affective computing integrates various sensory inputs through machine learning techniques to capture a more comprehensive emotional state. The field has evolved from early systems focused on Ekman's six basic emotions (Ekman & Friesen, 1971; Keltner, Sauter, Tracy, & Cowen, 2019; Russell, 1991) to more sophisticated approaches using dimensional models (Reisenzein et al., 2013; Russell, 1980).

But current systems particularly struggle with complex emotional states such as fatigue, anxiety, satisfaction, confusion, and frustration (Canal et al., 2022; Lee, Kim, Kim, Park, & Sohn, 2019; Misra & Gaj, 2006; Zhang et al., 2022).

These limitations are especially evident in real-world scenarios (Achlioptas, Ovsjanikov, Guibas, & Tulyakov, 2023), where contextual understanding becomes crucial. While deep learning models can effectively identify basic emotions, they often fail to comprehend the underlying complexities or contextual implications of emotional expressions (Canal et al., 2022; Ge, Zhu, Dai, Wang, & Wu, 2022; Sarvakar et al., 2023; Wang, Song et al., 2022). This gap between algorithmic recognition and human-like emotional understanding remains one of the key challenges in developing truly empathetic human-technology interaction systems (Maruf et al., 2024; Al-Saadawi, Das, & Das, 2024; Kalateh, Estrada-Jimenez, Hojjati, & Barata, 2024; Xu, Lin, Zhou, & Shan, 2024).

However, much work still needs to be done on the classification of emotions, which is still extremely coarse compared to the palette of emotional nuances subjectively felt by human beings.

1.5.4. Comparison with our method

Only a few studies have explored the emotional Turing Test, aiming to distinguish humans from machines based on emotional responses (Elkins & Chun, 2020; Floridi & Chiriatti, 2020; Ho, 2022). These studies concentrate on textual contents and reveal that machines often err in interpreting human emotions. A multimodal emotional AI model capable of attempting a Turing Test has not yet been developed. This difficulty underscores the need for a new framework in AI development — one that capitalizes on the distinct characteristics of the human mind, encompassing a reimagined approach to understanding and integrating the emotional framework (Ho, 2022; Olague, Olague, Jacobo-Lopez, & Ibarra-Vazquez, 2021; Wheeler, 2020).

Our research highlights fundamental differences between computational emotion recognition and human emotional processing. While recent advances in multimodal Artificial Intelligence systems have shown promising progress, as described above, significant limitations remain in processing complex emotional intersections.

The challenge extends beyond simple emotion recognition to understanding the complex interplay between different emotional stimuli. While comprehensive emotion datasets exist (kapodi, 2024), they

typically focus on isolated emotional expressions rather than the dynamic integration of multiple emotional cues that characterizes human emotional processing. Current systems can extract emotional information from facial expressions or analyze emotional content separately, but struggle with integrating multiple emotional cues - particularly when dealing with neutral facial expressions modified by contextual emotional factors.

It is crucial to note that while current AI approaches focus on detecting emotions directly expressed in the input data, our proposed method addresses a fundamentally different aspect of emotional processing. Our approach examines how emotions that are not explicitly present in the target stimuli, but rather emerge from contextual associations, influence memory and recognition.

This fundamental difference in approach makes direct quantitative comparisons between our method and current AI emotion recognition systems methodologically inappropriate, as they address distinctly different aspects of emotional processing. While current emotional Turing tests focus on the machine's ability to recognize and respond to explicit emotions, our method examines the deeper cognitive-emotional mechanisms that characterize human intelligence. Current limitations underscore the need for new approaches that can address the multifaceted nature of emotional cognition, particularly in understanding the subtle interplay between emotions, memory, and social context.

1.6. Cognition and emotions in humans

1.6.1. Neurophysiology of emotions

The idea underlying this paper and the novel kind of Turing Test we are going to propose has its foundation in the neurophysiology of emotions.

Emotions are complex processes triggered by sensorial or interoceptive[1] stimulations (Adolphs, 2002; Etkin, Büchel, & Gross, 2015; Kober et al., 2008; LeDoux, 2000; Pessoa, 2018) and involve multiple brain regions working in concert. These structures interact with each other and with sensory inputs to generate emotional experiences and guide behavioral responses. The specific activation and connectivity of these brain regions can vary depending on the type of emotion being experienced and the individual's unique responses to different stimuli.

This complex network of interconnected brain structures includes specifically Amygdala, Hypothalamus, Hippocampus, Cingulate Cortex and ventromedial Prefrontal Cortex (Adolphs, 2002; Etkin et al., 2015; Pessoa, 2008, 2018), although many other areas are involved.

Amygdala (Anderson & Phelps, 2001; Phelps & LeDoux, 2005) is a small, almond-shaped structure located deep within the temporal lobe plays a central role in the processing and regulation of emotions, especially fear and aggression. The amygdala is involved in evaluating the emotional significance of stimuli and triggering appropriate emotional responses.

Hypothalamus (Masserman, 1941; Zimmerman, 2016) is a region located below the thalamus. It is involved in various autonomic functions and also plays a role in emotional regulation. The hypothalamus can influence emotional responses by modulating physiological processes like heart rate, blood pressure, and hormonal release.

Hippocampus (Phelps, 2004) is primarily associated with memory and learning, but it also has connections with the amygdala and plays a role in associating emotions with memories. This can impact the emotional significance and recall of past experiences.

Cingulate Cortex (Stevens, Hurley, & Taber, 2011) is connected to various other brain regions, including the amygdala and Prefrontal Cortex and is implicated in emotional empathy, playing a role in regulating emotional responses.

Amygdala, Hypothalamus, Hippocampus, and Cingolate Cortex identify the so-called Limbic system. The Limbic system is closely connected with other brain regions involved in the control of higher cognitive functions, such as the frontal lobe.

In particular, the ventromedial Prefrontal Cortex (vmPFC) (Suzuki & Tanaka, 2021; Winecoff et al., 2013) is involved in emotional regulation, decision-making, and social behavior. It helps to control and modulate emotional responses, allowing for more adaptive behavior and appropriate social interactions.

However, brain areas deputed to seeing faces and listening to music are also involved in the experiment we are going to present.

Music listening involves various brain areas that work together to process different aspects of the music (Alluri et al., 2012; Baumgartner, Lutz, Schmidt, & Jäncke, 2006; Janata, 2009; Koelsch, 2014; Peretz & Zatorre, 2005; Salimpoor et al., 2013; Watanabe, Yagishita, & Kikyo, 2008). It is a highly complex process that engages multiple brain areas simultaneously, and these regions often interact with each other to create our overall musical experience. Additionally, individual differences in musical training, cultural background, and personal preferences can influence how these brain areas respond to and process music.

Auditory Cortex (Alluri et al., 2015; Kanwisher, McDermott, & Chun, 1997; Liégeois-Chauvel, Bénar, Krieg, Delbé, Chauvel, Giusiano, & Bigand, 2014) is the primary region responsible for processing auditory information. It receives and analyzes sound signals from the ears, helping to decode basic elements of music such as pitch, rhythm, and timbre.

But still, the Prefrontal Cortex, Hippocampus, and Amygdala (Baumgartner et al., 2006; Koelsch, 2014; Watanabe et al., 2008) are involved: Prefrontal Cortex plays a role in processing and analyzing complex musical structures and patterns. The hippocampus helps in recognizing familiar melodies and tunes, allowing us to recall and connect emotionally to specific musical pieces. The amygdala plays a crucial role in the emotional response to music. The Amygdala's involvement contributes to why certain music can evoke strong emotional reactions and can influence mood.

The specific brain areas involved in watching faces are the Occipital Face Area (OFA) (Baumgartner et al., 2006; Pitcher, Walsh, & Duchaine, 2011), situated in the occipital lobe, which is responsible for early processing of facial features and their configurations.

Moreover, these already mentioned areas play a role even in face recognition: Amygdala plays a role in attaching emotional significance to facial expressions, particularly threat-related expressions. The prefrontal Cortex is involved in interpreting facial expressions in social contexts and making judgments about others' mental states. The hippocampus plays a role in processing known and familiar faces and connecting them with past experiences and memories. Finally, Insula helps in experiencing and understanding the emotions conveyed through facial expressions (Brooks et al., 2012; Krolak-Salmon, Hénaff, Vighetto, Bertrand, & Mauguière, 2004; Liégeois-Chauvel et al., 2014).

1.6.2. Neurophysiological roots of the connection between emotions and memory

The key point of the Turing Test we wish to propose is to search for the difference between human and machine memory recall capabilities by exploiting a specificity of the human mind, namely the brain's ability to process cognitive functions such as recognition and memory integrating them with emotional ones.

The relationship between emotions and memory is intricate and profound. Emotions can significantly enhance attention and focus, influencing the encoding, consolidation, and retrieval of memories, and shaping our ability to remember and recall past events and experiences.

One of the key aspects of the relationship between emotions and memory is that emotionally charged events tend to be remembered more vividly and accurately compared to neutral events. The emotional

[1] Interoception refers to the perception and awareness of bodily sensations, such as heartbeat, breathing, or hunger. Emotional responses can also be triggered by interoceptive stimulations without conscious awareness of the specific physiological changes.

arousal associated with an event can enhance the encoding process, making the memory more robust and easier to recall (Brooks et al., 2012; Buchanan, 2007; Mather & Sutherland, 2011; McGaugh, 2000; Rimmele, Davachi, Petrov, Dougal, & Phelps, 2011; Sharot & Phelps, 2004; Talarico & Rubin, 2003).

Amygdala probably plays a crucial role in the influence of emotions on memory (Phelps, 2004). As above mentioned, it is involved in processing emotional significance and is particularly important for the consolidation of emotionally charged memories.

Moreover, emotions, particularly stress and arousal, can trigger the release of stress hormones, such as cortisol and adrenaline (Buchanan & Lovallo, 2001; Cahill & McGaugh, 1995; Critchley & Harrison, 2013; Kensinger & Schacter, 2006; Mather & Sutherland, 2011; Phelps & Sharot, 2008; Strange, Hurlemann, & Dolan, 2003; Talarico & Rubin, 2003). These hormones can modulate memory processes, affecting both short-term and long-term memory formation and allowing emotionally charged events to be remembered more vividly and persistently than neutral event. It must be mentioned that in specific cases, high levels of stress hormones may, vice versa, impair memory.

In summary, emotions play a crucial role in shaping how memories are encoded, consolidated, and retrieved, thanks to the involvement of specific brain regions and the influence of stress hormones in a quite complex way.

1.6.3. Emotions and awareness

Emotions can be elicited by conscious awareness of external events or stimuli. But they can also be triggered by stimuli that operate outside of conscious awareness. Subliminal or subtle sensory cues, such as facial expressions, body language, or brief auditory or visual signals, can influence emotional responses without the person consciously perceiving the triggering stimulus (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006). Research in psychology and neuroscience has demonstrated that unconscious and interoceptive processes can play significant roles in shaping emotional experiences (Marg, 1995; Phelps, 2006). Subconscious processes in the brain, often involving the amygdala and other limbic system structures, can process emotional information, thus behavior, before it reaches conscious awareness.

As we will see in the next paragraphs, the role of emotions, the influence of emotions on memorization and recall, and the possibility that emotional information is processed by the brain even outside of awareness identify a specificity of the human brain that we believe is not easily reproducible by a software. On this basis, we set up the new Turing Test we intend to propose.

2. Materials and methods

2.1. Proposal for a new Turing Test

The working hypothesis underlying this research is based on the observation that, in order to be recorded in memory and then recalled at later times and used for analogies, discriminations, deductions, inductions, or abductions, emotions in human beings do not necessarily have to pass through conscious attention, as is the case with perceptions. Conversely, the machine, by definition, can acquire and register information only if its sensors detect it within their specific focus or if it reads it as a digital datum, and to use it later, it must necessarily recall it entirely from memory: the machine can handle emotions in the only way that is inherently possible for it, that is, by acquiring or recalling them from memory as is the case with perceptions.

It is therefore interesting to identify a test that asks questions involving the substrate of emotions that remains outside consciousness and are accessible to the human being, but not accessible to the machine.

The brain, in fact, can be considered having two cognitive levels, one rational and the other emotional, which are inextricably intertwined by neural feedbacks that interconnect several brain areas, as detailed in paragraph 1.6.

We hypothesize that each perception engages a part of the emotional substrate, which does not need the conscious attention focus to be activated. Consequently, a test involving the association of perceptions that have never been directly associated but can be associated through unconscious or preconscious emotions should give different results when comparing machine and human beings.

A test based on this criterion was designed and administered to 39 subjects of different genders, ages, and professions through the online platform EMOTIVLABS made available by EMOTIV Inc, that produces EEG wearable hardware and software systems (Badcock et al., 2015; Emotiv, 2023; He et al., 2023; Khiani et al., 2022; Torse, Khanai, Pai, & Iyer, 2022; Williams et al., 2023; Williams, McArthur, de Wit, Ibrahim, & Badcock, 2020; Xolmurotova & Adilova, 2023).

2.2. Description of the experiment

This experiment was approved by the Ethics Committee of the University of Milan by opinion MP_88_21 dated 9/15/2021.

Following the guidelines of the Ethics Committee of the University of Milan, test participants signed a release form after reading and understanding an information sheet specifying the test procedures. After reviewing the test, the Ethics Committee established that the information sheet would include the statement 'There are no known risks either in viewing the videos and images or in acquiring EEG signals from the EMOTIV headset.'

It should also be noted that participants could only be EMOTIV EEG headset owners, therefore by choice accustomed to wearing the headset, and due to the nature of research and experiments carried out by owners of these headsets, used to receiving sensory stimuli of different nature for the purpose of developing or using software to analyze and interpret the resulting signals.

Furthermore, research has consistently shown that experiments involving emotional stimuli and EEG recordings typically do not cause significant participant distress. Badcock et al. (2015) in their validation study of the EMOTIV EPOC EEG system specifically assessed participant comfort and found no adverse effects. Similar findings were reported by Williams et al. (2023) in their study of crowdsourced EEG experiments using EMOTIVLABS, confirming the safety and low-stress nature of such protocols. Studies involving emotional stimuli presentation, such as those reviewed by Olague et al. (2021) have established that brief exposure to emotionally charged content in controlled experimental settings does not typically induce lasting psychological effects or significant distress in participants.

The experiment consists of three phases, which together take approximately 10 minutes:

1. A first training phase, in which the participant watches a video absorbing a series of visual and auditory sensations. The video consists of an animated red background accompanied by music apt to be possibly remembered with intense emotion (fear) (Fig. 1(a)).
2. A second training phase, where the participant is administered a series of images consisting of 18 women's faces on a white background, accompanied by three pieces of music, each accompanying 6 faces. One of the music pieces is the background music of the previously administered video (Fig. 1(b)).
3. A third phase (testing), in which the degree of memorization of what the participant has learned in the training phase is assessed.

Participants are administered the same 18 faces in random order, this time represented with the same red background as in the video.

Then, they are asked to rate the extent to which each of these images reminds them of an image already seen during training, indicating a value on a scale of 1 to 10 (1 — no memory, 10 — very sharp memory) (Fig. 1(c)).

(a) Snapshot of the video. An animated sequence is accompanied by music (first training phase).

(b) One of the faces administered to the participants accompanied by a music, that can be similar of different from the one in the video (second training phase).

(c) The same face is administered with a background similar to the one in the video (testing phase).

Fig. 1. The experiment.

It is observed that there is no previous association between faces and red background. However, the purpose of the experiment is to assess whether faces that during training were accompanied by the same music in the video are more easily recognized.

Our previous research hypothesized (Pizzi, 2020) that human beings possess a substrate (not dependent on conscious attention) that associates perceptions and emotions in an integrated binding that persists in the unconscious. The moment conscious attention rests on an image, if it is composed of elements that have in the past triggered an emotion, that emotion is evoked along with the other emotions in the binding, each linked to its own perceptions.

In this case, the red background evokes the emotion triggered by the video, which is connected by binding to the same music that was also present in some of the images of women's faces.

For this reason, the participant, seeing a face on a red background, should more easily recognize the faces that were accompanied by the music of the video in the training phase.

According to the working hypothesis, this would not be possible for a computer because it did not store any association between women's faces and the red background.

2.3. Rationale of the new Turing Test

The proposed Turing Test consists in comparing the performance of human and machine in recognizing various women's faces already previously read/seen, as mentioned above. The experimental workflow of the Turing Test can be followed through the flowchart in Fig. 2.

Both the computer and the research participants are made aware of the following:

- a video that associates a music with the color red (RED Music1);
- images that associate the same Music1 with a group of women's faces;
- images that associate other music (Music2, Music3) with two other groups of women's faces.

After reading the various files, with an appropriate Machine Learning system, the computer would then be able to show that it knows:

- the video (RED Music1);
- all images with women's faces;
- the association between the women's faces and the related music.

However, our hypothesis is that this information does not allow it to predominantly recognize the women's faces that had been presented in the training together with the same music as the video.

In fact, the software has no way to predominantly connect such images with the video RED Music1 and with the images of women's faces that had been presented to it together with Music1, because the software does not have any information that connects them in a unique framework.

The situation for the human participant is different.

In fact, when she or he watches the video, the drama of the music triggers in the participant an emotion related to the video as a whole ("RED Music1"). So, at the end of watching the video, an emotional layer surrounds both the music and the red color as a whole: a specific emotional connection between "Music1" and "RED" occurs. It must be noted that the color red alone does not (in general) arouse any particular emotion, being part of common perceptions.

Next, the participant is presented with images of women's faces with musical backgrounds (Music1, Music2, Music3). Now, the images previously paired with Music1 have the possibility of triggering the same emotion felt while watching the video "RED Music1".

At the time of the test, when the participant views images with women's faces, all of which have red backgrounds, he or she will be able to remember more easily those that were presented with the same background music (Music1) as the video, *i.e.* the Image1 group, because those faces were stored with the emotional substrate "RED Music1". The emotion connecting red color and video music is not conscious, but it exists. As seen in paragraph 1.6.3, awareness is not necessary for emotions to integrate with the cognitive system to enhance memorization and recall.

This emotional substrate that associates the video, the training faces with the same music as the video, and the same images in testing, thus allowing the memory of those images to be enhanced, is not present in ML software (see Fig. 3).

Conversely, human beings add to each perception that is submitted to them an emotional substrate that possesses intersections with

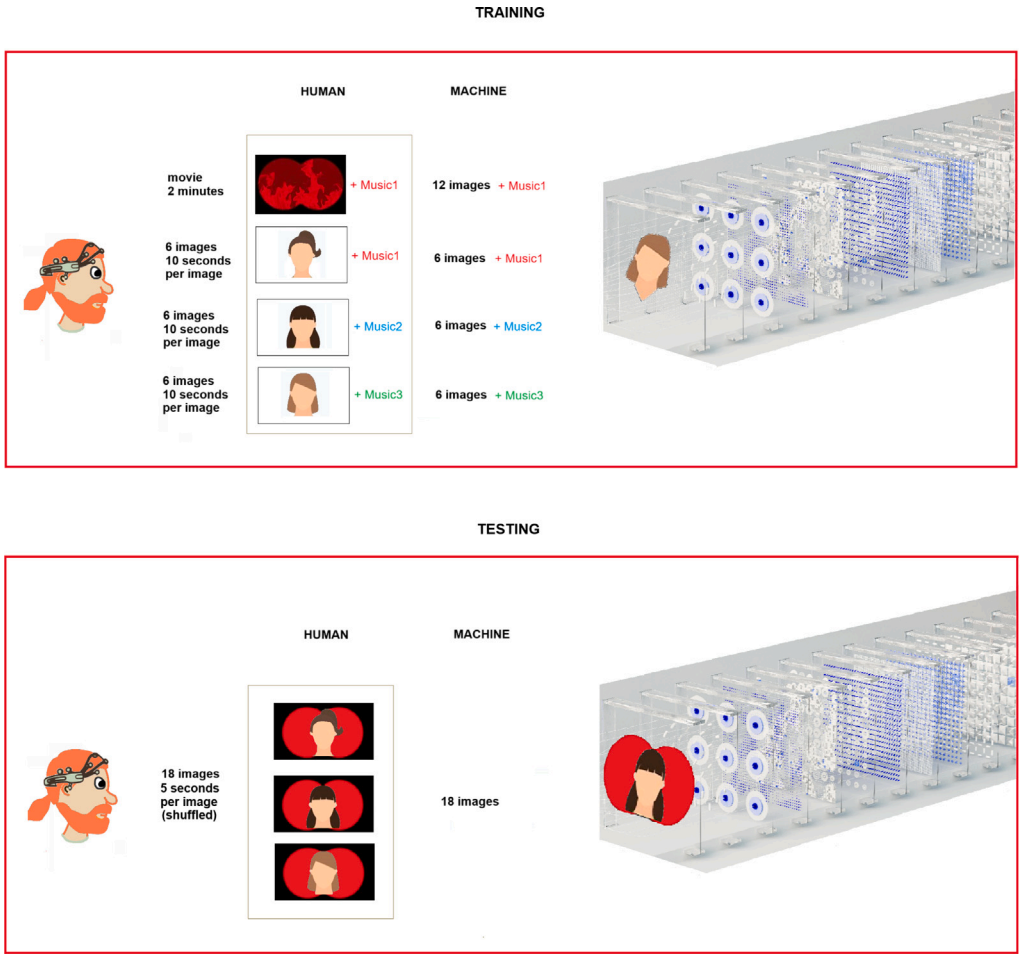


Fig. 2. Training: humans see 1 video with RED background and Music 1, 1 video with 6 faces and Music1, 1 video with 6 faces and Music2, 1 video with 6 images and Music3. The machine is subministered 12 images with RED background AND Music1, 6 images with Music1, 6 images with Music2, 6 images with Music3. Testing: humans are subministered the same 18 faces with RED background. Machine is subministered the same 18 faces with RED background.

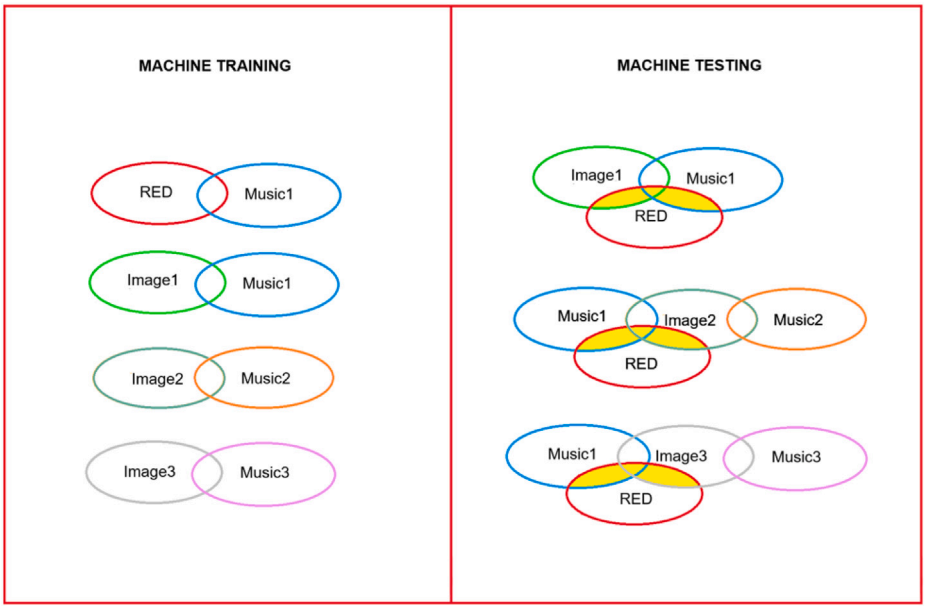


Fig. 3. MACHINE TRAINING: During training, the software simultaneously memorizes the perception of images and music. MACHINE TESTING: During testing, the software remembers the RED image presented along with the faces, but this is never associated with any face. Consequently, for the software there is no difference in the recall of various face images, because it was never able to associate any face with the RED image.

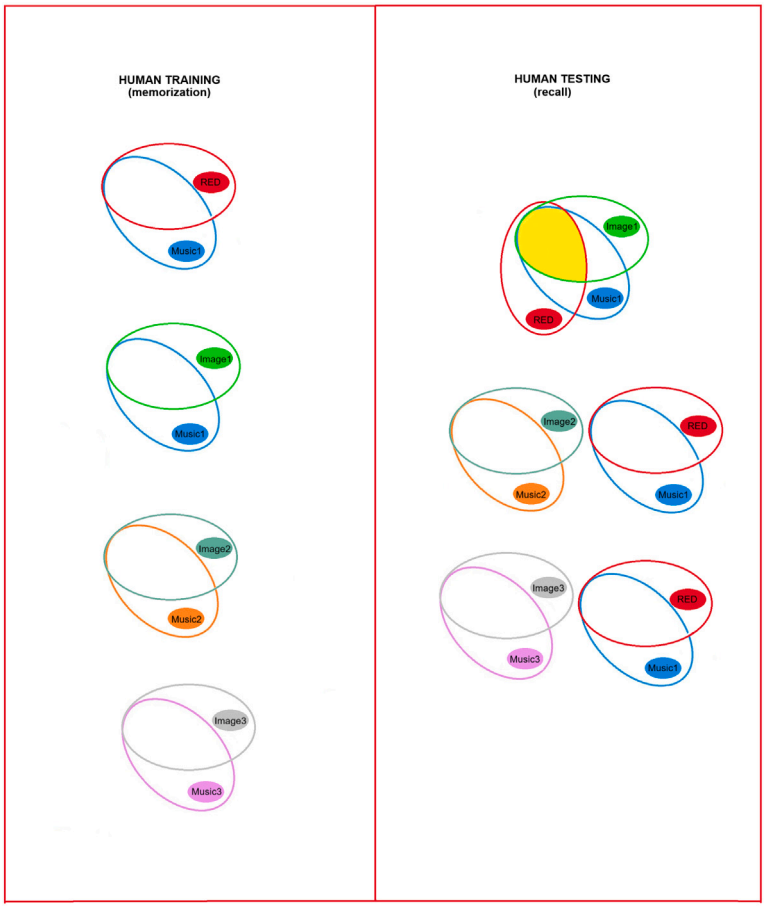


Fig. 4. HUMAN TRAINING: The participant is presented with four series of images and sounds. The first series consists of scary music (Music1) accompanied by a 2-minute video showing a series of images with a red background (RED). Subsequently, 18 face images are presented, divided into 3 series (Image1, Image2, Image3), each accompanied by music. The Image1 face series is accompanied by the same Music1 present in the first series with RED images. It is hypothesized that participants develop an "emotional shell", individually more or less conscious, corresponding to each of these four perceptual/emotional events. The perceptual/emotional event Image1+Music1 will develop an emotional shell that intersects with the one generated by viewing the first series RED+Music1. HUMAN TESTING: The testing consists of presenting 18 face images, identical to those shown in training but with the same RED background used in the first video, RED+Music1. The key point is that during training there was no perceptual association between face images and RED background. As shown in the figure, the "emotional shells" present in the participants' minds overlap. However, the intersection between these shells is not equivalent. For test images corresponding to Image1 from training, there will be an intersection with the RED+Music1 emotional shell, involving not only the RED color but also Music1. Consequently, the recall of Image1 images will be easier in (Test1) compared to Image2 and Image3 images, *i.e.*, (Test2) and (Test3).

other substrates belonging to other perceptions, which allows them to enhance the memory related to those perceptions. As shown in Fig. 4, the intersection between perceptions and emotions is more distinct and larger than in the case of an ML system and can lead to easier recall.

As seen in paragraph 1.5, currently, ML is able to recognize only a few primary emotions from texts or images, but it is very difficult to imagine how it can derive even in the medium-term information fine enough to enable it to emulate the human neurophysiological system from the point of view of the interaction among emotions and between perceptions and emotions.

In the following, we will show that the test administered to the research participants appears to validate this theory.

2.4. Results of the human test

All 39 participants in the experiment were members of the EMOTIVLABS community using EPOC EEG wireless headsets with 14 electrodes or EMOTIV more advanced models.

Most participants were from Australia (11), USA (7) and Italy (7). 29 participants were males, 10 were women, and the average age was 31. Most of them (33) listen to music assiduously or play an instrument. 33 of them associate memories to music at least occasionally. 28 believe that the first video was important for the scores of the testing phase.

When asked about a specific impression that motivated their answers, 15 mentioned the facial features, but 18 mentioned music, and 6 mentioned the red color or the fear evocated by the music in the video (elements not present in the testing phase).

The responses to these general questions are associated randomly with greater or lesser recognition of the group of faces presented with Music1, supporting the hypothesis that the ability to connect perceptions with an emotional substrate is often unconscious and is a property of all minds.

Nonetheless, the maximum score difference in recognition between groups of faces, favoring faces previously presented with Music1, belongs to people who expressed as a reason for better recognition: 29 participants: music 16 participants: red color 13 participants: expressions and emotions in the faces 3 participants specifically mentioned the video as the reason for the best recognition.

Indeed, the difference between recognizing the faces previously presented with the same music of the video versus recognizing the other faces was statistically significant. In fact the analysis performed with a 2-tailed paired Student's t-test (Welch), significance level $\alpha = 0.05$, gave P-value = 0.01445 (as shown in Table 1). That is, the participants recognized the faces previously presented with the same music of the video more easily than the others.

Table 1
Comparison between participants' recognition of different groups of images — Student's t-test results.

Parameter	
Sample size	39
Average of differences	−32.8205
SD of differences	79.9654
t	−2.5632
P-value	0.01445

2.5. EEG signal processing

Another interesting analysis was obtained from the EEG signals recorded from the participants. The neurophysiological reactions to different videos and images were flagged with markers, allowing the reactions to individual perceptions to be analyzed and compared to each other. For this purpose, a self-organizing ANN developed by our group (Pizzi, 2020; Pizzi, Cino, Gelain, Rossetti, & Vescovi, 2007; Pizzi, de Curtis, & Dickson, 2003; Pizzi, Musumeci, et al., 2017a, 2017b; Pizzi et al., 2009) (ITSOM, Inductive Tracing Self-Organizing Map) was used. ITSOM makes it possible to identify mental states from EEG signals, allowing each portion of the signal referring to the individual reaction to be identified with a specific binary code.

The Self-Organizing Map (SOM) (Kohonen, 1990) features are well known. Also well known are its limits in classifying topologically entangled input structures. To overcome these limits we developed the ITSOM architecture based on the following observation: even though the SOM winning weights vary at any given presentation epoch, their temporal sequence tends to repeat itself. The dynamical properties of the SOM have been investigated (Ermentrout, 1992; Ritter & Schulten, 1986, 1988), and show periodic oscillations and limit cycles. In particular we observed that the sequence of winning weights constitutes chaotic attractors that univocally characterize the input element that has determined them. Thus these sequences make it possible to finely classify the corresponding input value. In Pizzi et al. (2017a) the analysis of the underlying dynamical system is performed.

The ITSOM architecture implies that the winning weight represents an approximation of the input. At every epoch the new winning weight, along with the weight that won in the previous epoch, constitutes a second order approximation of the input value, and so on, creating in time a specific configuration.

After a limited number of epochs (400 in this proposed application), the ITSOM is stopped, and the quasi-periodical time sequence of winning neurons is processed, giving rise to numeric configurations that characterize univocally the input signals that produced them, as shown in Fig. 6.

In this way it is possible to derive the input value by comparing the specific configurations of each input with a set of reference configurations, whose value is known. Thus a real process of induction is realized, because once a vector quantization many-to-few from the input layer on the weight layer is carried out, a few-to-many step is operated from reference configurations to the whole input, as shown in Fig. 5.

This form of induction is much finer than the one obtainable from the only final winning neurons of the SOM network, because the choice among a set of competitive layer neurons is too limited to provide a meaningful classification. Instead the possible ITSOM outputs are 2^n, where n is the number of neurons of the competitive layer, that make it possible to finely discriminate the input features.

The best suited algorithm to recognize the configurations created by the network has proved to be a z-score-based one. The cumulative scores for each input are normalized according to the distribution of the standardized variable z given by

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

where μ is the average of the scores on the neurons of the competitive layer and σ is the standard deviation. Once set a threshold τ, which therefore constitutes one of the hyperparameters of this type of network (2),

$$0 < \tau \leq 1 \tag{2}$$

each configuration of winning neurons is represented by a binary number formed by as many ones and zeros as many the output layer neurons (3). Then it is immediate to use these binary numbers as templates of the input patterns.

$$z = \begin{cases} 1, & \text{for } z > \tau \\ 0, & \text{for } z \leq \tau \end{cases} \tag{3}$$

In our case, each input pattern (in this case, a portion of the signal corresponding to a specific video or image and collected from a specific electrode, filtered with Gamma or Beta frequency) is represented by a binary code.

These codes are easily comparable, so the ANN allows a fine classification of the signal on the basis of its dynamical self-organization in time. The flexibility of the ANN allows to attribute the same codes to similar sensory or cognitive events, well differentiated from the codes corresponding to different stimuli, as shown in Fig. 6.

A unique hyperparameter configuration was used for all analyses so that the resulting binary codes, set to 10 bits, could be directly compared.

On the basis of the results of their responses highlighted in the previous paragraph, we chose to compare the signals of the three participants who best recognized the images coupled with the music in the video (group Y) and the three subjects who to the least extent saw distinctions between the three groups of images (group N).

2.6. Signal preprocessing

The EMOTIV headsets used by experimenters could be of different types, but the minimum required model was EMOTIV EPOC+ v.1.1 with 14 electrodes. This type of headset has a sampling rate of 2048 Hz, bandwidth 0.1–45 Hz with notch filters at 50 and 60 Hz, and a built-in digital 5th order Sinc filter, which is a low-pass filter useful for reducing high-frequency noise, limiting signal bandwidth preventing aliasing, overall improving signal quality. After applying the low-pass filter at 64 Hz, the signal no longer contains frequencies above 64 Hz.

Subsequently, the data are downsampled to 128 Hz or 256 Hz for transmission without loss of significant information, thanks to the previous filtering. The high internal sampling frequency allows for more precise filtering before downsampling, then downsampling reduces the amount of data to be transmitted, optimizing battery usage and bandwidth.

Furthermore, the EMOTIV system contains a procedure called EEG Quality, a set of algorithms that automatically determine the signal quality based on multiple metrics, assessing whether the recording data accurately capture the underlying brain signal. These metrics include:

-ML Signal Quality (SQ): A machine learning algorithm trained on high-quality EEG recordings that were assessed and collected by the EMOTIV Research team.

-Contact Quality (CQ) : an impedance measurement that indicates the quality of the electrical signal passing through the sensors and the reference

-Signal Magnitude Quality (SMQ): a measure of the signal amplitude. Sometimes, the above metrics *i.e.* CQ, and SQ are good but the signal amplitude is very small, therefore small power fluctuations in the FFT would be undetectable. Often, this is due to poor sensor hydration or poor scalp contact.

Each of these metrics is important in determining the signal quality and allows the user to acquire the signal only when it exceeds a sufficient EEG Quality score.

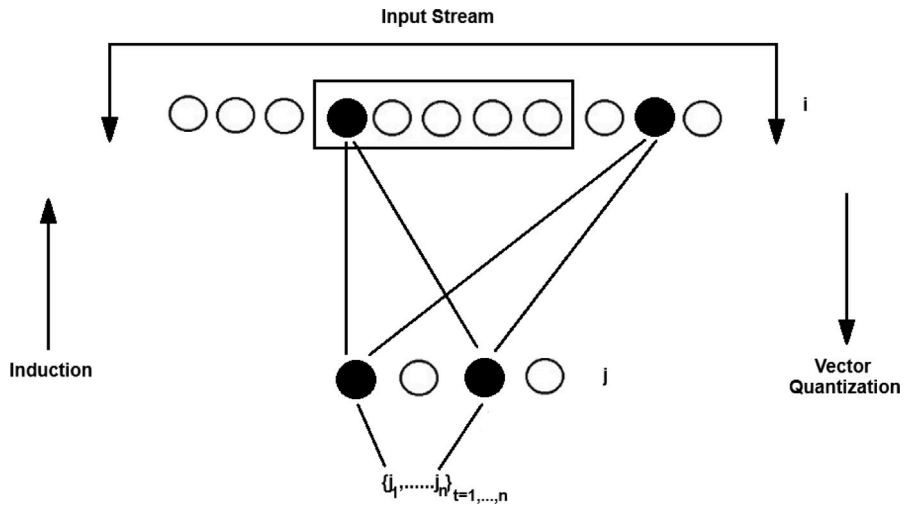


Fig. 5. The Inductive Tracing Self Organizing Map (ITSOM) architecture. Vector quantization is realized over time and identifies a time series of winning neurons that uniquely characterize the input. This enables a many-to-one induction process from different input sequences to a unique template.

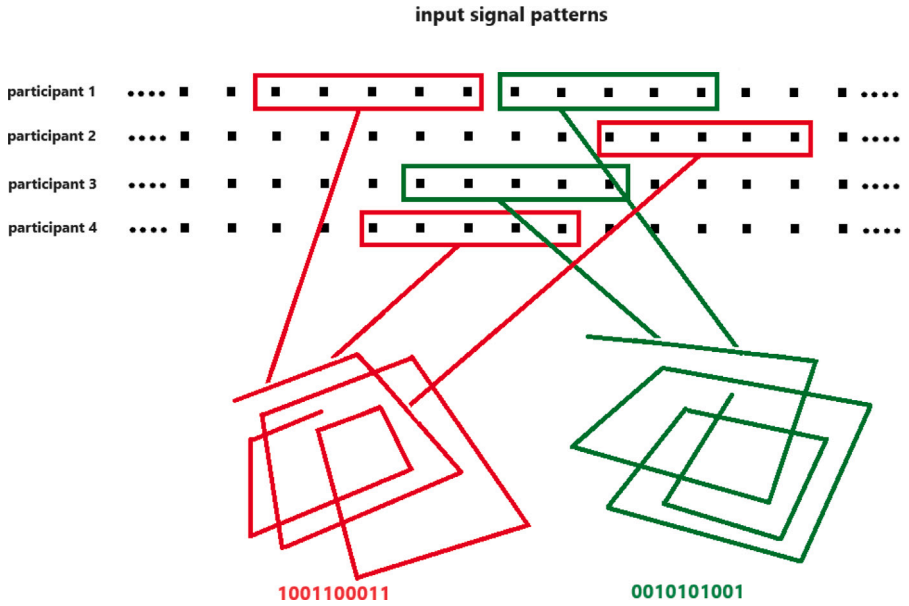


Fig. 6. The ITSOM ANN: each signal segments corresponding to a specific perception gives rise to a quasi-periodic sequence of winning neurons, each one transformed into a binary code.

2.7. Artifact management

Before starting, the experimenter is advised to remain still and relaxed and is trained to correctly position the electrodes and check EEG quality.

Despite these precautions, some muscle artifacts are inevitable. However, to manage muscle artifacts in EEG, low-pass filters are mainly used because most of these are characterized by high frequencies. As seen, the adopted low-pass filter limits the bandwidth to 45 Hz.

To prepare the signals for subsequent analyses, the following procedures are also implemented:

-Signal re-referencing, aimed at reducing noise common to all electrodes and improving the signal-to-noise ratio. We adopted the interquartile mean, robust against outliers and artifacts. This process helps isolate the specific brain activity of each area, removing signals that are common to all electrodes.

-High-pass filtering (typically with cutoff frequency of 3 dB at 0.5 Hz) is used to remove unwanted low-frequency components from the signal. This filtering eliminates slow signal drift caused by sweating, slow movements, and variations in electrode-skin impedance, also preserves EEG components of interest (typically above 1 Hz) and improves baseline signal stability.

-Slew rate limitation (typical value: 30 μV/sample). This technique limits the maximum rate of signal variation between consecutive samples. The system imposes a maximum limit of 30 μV on the amplitude difference between two successive samples: amplitude changes that exceed this threshold are automatically limited to the maximum allowed value. This technique is effective for attenuating movement artifacts, reducing the impact of electrostatic disturbances, and filtering sudden peaks and transients that do not belong to the EEG signal.

As in our mentioned previous research, signals from electrodes P7, F7, O1, T8 were processed because they were identified as most significant for our study. In more detail:

- T8 (Temporal 8), is located on the right side of the scalp, over the temporal lobe. The temporal lobes are involved in auditory processing, memory, and language comprehension. T8 and is also connected with the Limbic system and is often associated with

the right hemisphere of the brain, which is home to creativity, emotional processing and plays a significant role in recognizing faces and interpreting emotional expressions.
- P7 (Parietal 7) is positioned on the left side of the scalp, over the parietal lobe. The parietal lobes are involved in sensory integration, spatial awareness, and attention.
- F7 (Frontal 7) is placed on the left side of the scalp, over the frontal lobe. The frontal lobes are associated with various higher cognitive functions, including problem-solving and memory. As we have seen, the prefrontal cortex, which is part of the frontal lobes, interacts with limbic structures and plays a role in emotional regulation and decision-making.
- O1 (Occipital 1) is located on the left side of the scalp, over the occipital lobe. The occipital lobes are primarily responsible for visual processing.

The EEG bands considered were: Theta: 4–8 Hz, Alpha: 8–12 Hz, Beta: 12–25 Hz, and Gamma: 25–45 Hz.

The decision to analyze only Beta and Gamma frequencies stems from two motivations. The most obvious one is to facilitate comparison and highlight signal differentiation among the 39 participants. However, the choice is primarily motivated by their greater specificity for the cognitive processes of interest. Indeed, our study requires sustained attention and places maximum relevance on information integration processes.

Numerous studies (Kisley & Cornwell, 2006; Steriade, Gloor, Llinas, Da Silva, & Mesulam, 1990; Tallon-Baudry, Kreiter, & Bertrand, 1999) have demonstrated that Beta and Gamma oscillations are particularly relevant for processing complex cognitive information, coordinating neural activity between different brain regions, and in attentional and working memory processes. Specifically:

- Beta waves are associated with states of active attention and concentration, and are dominant during conscious thinking, cognitive processing, and problem-solving.

- Gamma waves are indicative of complex cognitive processing and correlated with high-level cognitive processes. They are especially specific to the problem we are addressing because they are involved in the integration of information between different brain areas.

Furthermore, our selection was influenced by the need for consistency with our previous research, which had similarly focused on these two frequency bands for the aforementioned reasons, producing results that we believe validate and justify the continued use of this methodological approach.

2.8. Signal analysis

The ANN analyzed both the individual signals and their composition, considering all four electrodes separately and simultaneously. Since the probability of finding identical binary codes is very low (1/1024 for a 10 bit code), the most interesting results, namely with the highest number of matches, were found in the analysis of the electrodes taken together.

Among the three subjects with better recognition scores on the images paired with the red background video (Y), it is noted that the codes related to the RED Music1 video have numerous matches with the codes of other Y participants related to both the video and the Music 1 image. Codes related to these images have numerous correspondences with other codes of the same participant and of the video codes of other participants. These correspondences are widely prevalent among Y subjects compared to the *N* group of participants (depending on the participant, 5–7 correspondences for the Y group versus 0–2 for the *N* group). It is interesting to note that two out of three participants in group Y have identical codes relative to four Music1 images. This is not the case in group N; on the contrary, there is a virtually complete lack of matches relative to these images. By calculating the total number of identical codes relative to the video and the Music 1 images for group Y and group N, it is possible to assess whether the difference between the two groups is statistically significant. Because the sample size is very low a t-test was used. In this case, we used a two sample t-test (Welch) right-tailed, significance level $\alpha = 0.05$. The p-value equals 0.03212, and since p-value $< \alpha$, the null hypothesis is rejected, and a statistical significance of the result is reached, as shown in Table 2.

Table 2
Comparison between best and worst participants' emotive recognition abilities — Student's t-test results.

Group name:	**Y**	**N**
Sample average:	775.33	356.33
Sample SD:	165.50	222.994
t	2.6134	
P-value	0.03212	
Effect size	2.13	

The sample size is small, with low priori power (0.2016); thus, the statistical significance obtained should be further investigated with larger samples. However, we should note that the observed effect size (d) is quite large, 2.13. This indicates that the magnitude of the difference between the two groups is definitely significant.

Effect size is a correct parameter in this context because it is independent of the sample size, which, in this case, is very small. But it must be said that the choice of the groups' subjects was obtained from the processing of hundreds of files related to signals from four electrodes on two frequencies, taken while viewing images and video.

2.9. Comparison with other methods

The comparative performance evaluation of self-organizing methods, such as the ITSOM used in this study, presents inherent challenges due to the unsupervised nature of these approaches. While a comparison with other methods, such as traditional Self-Organizing Maps (SOM) or clustering algorithms like k-means, might provide additional perspectives, it is important to emphasize that such direct comparisons are inherently limited in this specific context.

The main reason lies in the absence of an objective 'ground truth' for classifying EEG signals in relation to mental states induced by visual and auditory stimuli in our experiment. Each self-organizing method might converge toward unique data representations, making a universal evaluation metric problematic.

Furthermore, these methods' sensitivity to initial conditions and configuration parameters further complicates objective comparison.

In our specific case, ITSOM was chosen for its demonstrated capability to accurately identify the organized structures of electrophysiological signals in our previous studies (the already mentioned (Pizzi, 2020; Pizzi et al., 2007, 2003, 2017a, 2017b) , effectively capturing the temporal dynamics of EEG signals, a crucial aspect for analyzing cognitive responses to presented stimuli.

In particular, the theoretical framework of ITSOM was validated by demonstrating its inductive capabilities in the research work described in Pizzi et al. (2009), where electric signals from human neurons processed by ITSOM learned to reply selectively to different stimulation patterns, and their response signals could effectively be decoded by the ITSOM to operate a minirobot.

Rather than relying on direct comparison with other methods, such as those mentioned in Section 1.5.1, which might not provide a definitive evaluation of one approach's superiority over another, we opted to validate our results through:

- the internal coherence of obtained results, as evidenced by significant correspondences between binary codes for similar stimuli.
- the correspondence between ITSOM analysis results and participants' behavioral responses.
- statistical analysis that demonstrated significant differences between participant groups, consistent with their test responses.

This multi-criteria evaluation approach, while not providing a direct comparison with other methods, offers robust internal validation of the results obtained with ITSOM in the specific context of this study.

2.10. Description of the machine learning system

In order to complete the Turing Test, a computer would have to perform the exact same test administered to human participants. For this purpose, we implemented a Machine Learning system and subjected it to the same test. Let us look in detail at how we developed the software and performed the test.

The computer test was designed to be exactly comparable to the one administered to human participants. Specifically: the ML system, that will be described in detail below, receives as input:

- RED images and sounds Music1
- Images of 18 faces in three groups Image1, Image2, Image3, each paired with a music piece
- Music pieces for each of the three groups: Music1, Music2, Music3.

The testing phase consists of 18 face images with RED background, identical to those administered in the test for human participants. The RED images were calculated to be 12 in number to match the exposure time of participants to the face images: one minute for each group of 6 faces and their respective music, 2 min for the RED+Music1 video. The three music pieces were divided into 22 .wav segments each, then further processed into chunks of .96 s each, to conform to the VGGish method.

A mapping system was necessary to correctly dimension the structures within the ML code, since training involves both images and music, while testing occurs solely through images. Therefore, correspondence maps between images and music were inserted in the training phase, and a map between training image codes and test image codes was implemented in the testing phase.

The initially developed multimodal convolutional neural network immediately showed a strong tendency toward overfitting caused by the minimal training dataset, resulting in extremely poor results. This necessitated a series of significant code modifications. At the end of this process, the network configuration described in the following was chosen.

This network configuration was executed 39 times to obtain statistics similar to those obtained for human experimenters.

2.10.1. Transfer learning

Our transfer learning choice focused on VGGFace (Wan, Liu, Huo, & Fang, 2017) for images and VGGish (El-Latif, El-Sayad, Mohammed, Darwish, & Hassanien, 2024) for music.

Despite several state-of-the-art pretrained networks being available for face recognition (Bhavani & Karthikeyan, 2024; Huu et al., 2022; Parkhi, Vedaldi, & Zisserman, 2015; Schroff, Kalenichenko, & Philbin, 2015; Wang & Deng, 2021; Wang, Peng et al., 2022), we chose VGGFace, considering the specific characteristics of the images to be analyzed. Indeed, VGGFace proves to be particularly effective in controlled conditions such as those in question:

- The faces it seeks to recognize are all in the same position and of equal size. This aligns well with the type of data on which VGGFace was trained. - Absence of expressions and background: VGGFace is particularly effective in recognizing invariant facial features, so the absence of expressions in its images should be an advantage. Specifically, its evolution VGGFace2 is trained to handle variabilities that are not present in our case, such as multiple poses, lighting variations, different facial expressions, complex backgrounds. Therefore, VGGFace still has better performance in controlled conditions, with less risk of overfitting on unnecessary variations.

Regarding music, our choice fell on VGGish, which is particularly suitable for analyzing music without speech. VGGish is effective in extracting low-level audio features, which can be very useful for analyzing and classifying different types of music. Moreover, compared to some more recent and complex audio models (Huu et al., 2022; Singh & Biswas, 2022; Zaman, Sah, Direkoglu, & Unoki, 2023), VGGish offers a good balance between performance and computational requirements.

2.11. Neural network structure

To implement an ML system able to compete with the human experimenters in the described Turing test, we thought to implement a neural network structured as follows.

General network structure: The network is a multimodal model that combines audio and image data processing. It consists of two main paths: one for audio input and one for image input.

1. Audio Branch:
- Input: An audio signal of length 15600 samples.
- Processing:
• VGGish Layer: Uses a pre-trained VGGish model to extract audio features.
• Dense Layer (512 units) with ReLU activation and L2 regularization.
• Dropout (30%)
• Dense Layer (256 units) with ReLU activation and L2 regularization.
• Dropout (30%)
- Output: Dense Layer with 3 units (number of audio classes) and softmax activation.

2. Image Branch:
- Input: RGB image of dimensions 224 × 224 pixels.
- Processing:
VGGFace (based on VGG16) pre-trained.
- The typical VGG16 structure includes:
• Two 3 × 3 convolutional layers with 64 filters, followed by max pooling.
• Two 3 × 3 convolutional layers with 128 filters, followed by max pooling.
• Three 3 × 3 convolutional layers with 256 filters, followed by max pooling.
• Three 3 × 3 convolutional layers with 512 filters, followed by max pooling.
• Three 3 × 3 convolutional layers with 512 filters, followed by max pooling.
• After these convolutional layers, the model adds customized fully connected layers to adapt to the specific task.
- Output: Dense Layer with 30 units (image classes) and softmax activation.

Additionally, we adopted Batch Normalization, applied after dense layers in both branches to normalize inputs and improve training stability.

For the image label mapping, the mapping structure consisted of 30 total classes (18 faces + 12 RED images), divided into three main groups a, b, c.[2] For the image-to-audio mapping, image classes were grouped into three audio macro-categories: a, b, c: All “a” image classes are matched to audio category “a”, all “b” to (Music1) “b”, and all “c” to “c”.

This structure aimed to balance the ability to extract complex features (through pre-trained networks) with adaptability to the specific task (through customized layers and fine-tuning). However, given the small dataset size, there remained a high risk of overfitting, as evidenced by our benchmark results.

Thus we attempted to improve the multimodal neural network structure by adopting the following techniques:

- The last 4 layers of VGGFace were unlocked for fine-tuning, meaning most of the pre-trained knowledge was preserved (frozen layers) while allowing some fine-tuning in the last layers to adapt to the specific task.

[2] In the results presented below, we have for brevity’s sake indicated the 18 faces with letters. From here on, by ‘b’ we mean the group of Image1 accompanied by the same Music1 music from the RED video.

Table 3
Summary of the used deep network.

Layer (type)	Output shape	Param #
image_input (InputLayer)	(None, 224, 224, 3)	0
conv1_1 (Conv2D)	(None, 224, 224, 64)	1792
conv1_2 (Conv2D)	(None, 224, 224, 64)	36 928
pool1 (MaxPooling2D)	(None, 112, 112, 64)	0
conv2_1 (Conv2D)	(None, 112, 112, 128)	73 856
conv2_2 (Conv2D)	(None, 112, 112, 128)	147 584
pool2 (MaxPooling2D)	(None, 56, 56, 128)	0
conv3_1 (Conv2D)	(None, 56, 56, 256)	295 168
conv3_2 (Conv2D)	(None, 56, 56, 256)	590 080
conv3_3 (Conv2D)	(None, 56, 56, 256)	590 080
pool3 (MaxPooling2D)	(None, 28, 28, 256)	0
conv4_1 (Conv2D)	(None, 28, 28, 512)	1180160
conv4_2 (Conv2D)	(None, 28, 28, 512)	2359808
conv4_3 (Conv2D)	(None, 28, 28, 512)	2359808
pool4 (MaxPooling2D)	(None, 14, 14, 512)	0
conv5_1 (Conv2D)	(None, 14, 14, 512)	2359808
conv5_2 (Conv2D)	(None, 14, 14, 512)	2359808
conv5_3 (Conv2D)	(None, 14, 14, 512)	2359808
audio_input (InputLayer)	(None, 15600)	0
pool5 (MaxPooling2D)	(None, 7, 7, 512)	0
vg_gish_layer (VGGishLayer)	(None, 128)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 512)	0
dense (Dense)	(None, 32)	4128
dense_2 (Dense)	(None, 32)	16 416
batch_normalization (BatchNormalization)	(None, 32)	128
batch_normalization_2 (BatchNormalization)	(None, 32)	128
dropout (Dropout)	(None, 32)	0
dropout_2 (Dropout)	(None, 32)	0
dense_1 (Dense)	(None, 32)	1056
dense_3 (Dense)	(None, 32)	1056
batch_normalization_1 (BatchNormalization)	(None, 32)	128
batch_normalization_3 (BatchNormalization)	(None, 32)	128
dropout_1 (Dropout)	(None, 32)	0
dropout_3 (Dropout)	(None, 32)	0
audio_output (Dense)	(None, 3)	99
image_output (Dense)	(None, 33)	1089
Total params:		14,739,044
Trainable params:		7,103,524
Non-trainable params:		7,635,520

- Data augmentation techniques were implemented for images to artificially increase the dataset size.

- Aggressive dropout (0.5) was adopted in the audio branch, which eliminates 50% of connections during training. Dropout helps prevent overfitting by randomly turning off a percentage of neurons during training, forcing the network to learn more robust and generalizable features.

- A more moderate dropout (0.3) was maintained in the image branch, since features from VGG16 are already robust.

- Batch size reduction to 4: small batch size promotes regularization.

- Global Average Pooling in the image branch after the last convolutional block, to reduce the number of parameters compared to a fully connected layer, again to limit the risk of overfitting.

- For complexity reduction, each branch (image and audio) was set with two Dense layers, with their size reduced to 32 units each, significantly smaller than the original configuration (which used 512 and 256 units). This reduction in layer size helps maintain a limited number of parameters and prevent overfitting.

- We added L2 regularization (0.001) to dense layers, which penalized large weights for additional overfitting control.

In essence, in both audio and image branches, we adopted a progressive dimensionality reduction, avoiding the memorization of irrelevant details. The network was significantly streamlined by drastically reducing the number of trainable parameters, ensuring more stable convergence and a better data/parameters ratio, while regularization techniques were adopted to improve model generalization. We show the summary of the model in Table 3.

2.11.1. Training optimization

As above mentioned, we adopted Transfer Learning using pre-trained VGG16 and VGGish weights.

To achieve more robust training, we chose Adam as an adaptive optimizer, automatically adjusting learning rates for each parameter, facilitating convergence.

We then applied a conservative learning rate (1e−5) with a minimum set to 1e−6 through ReduceLROnPlateau, technique that automatically reduces learning rate if validation loss stabilizes.

Early stopping was also introduced to halt training when validation set performance stops improving for a certain number of epochs (patience = 10), again to prevent overfitting.

Finally, we adopted Step decay, reducing the learning rate by a factor of 0.1 every 10 epochs, for very precise weight fine-tuning.

However, while these measures improved loss and accuracy trends during training, they did not improve testing performance, which remained polarized to one or two images, although not the same in each run.

2.11.2. Explanation of low performances

It should not be surprising that the network's performance was so poor, even though professional systems exist that can identify a face within databases of millions of images, such as Google Images (Shepley, 2019) or law enforcement software for identifying suspects. However, there is a fundamental distinction between the problem at hand and the operation of large-scale facial search systems. In our specific case, there is a limited set of target faces (the training dataset), and the goal is to classify new images within this limited set of categories.

Transfer learning is used to extract features, but essentially we are trying to perform final classification on a closed set of classes. Large-scale facial search systems, on the other hand, do not try to classify into a predefined set of categories but rather compare a single face against a vast database: the problem is more similar to a similarity search or a "retrieval" problem rather than classification.

The approach to transfer learning is also different: in our case, transfer learning is used to initialize feature extraction, but then a specific classifier is trained for the classes. Large-scale systems instead use transfer learning to create universal representations of faces, without a specific final classifier. Therefore, while in our case overfitting is a real risk because we are training on a very specific and limited set of faces, in large-scale systems, overfitting in the traditional sense is less problematic because they are not trying to classify into predefined categories but to create robust and general representations of faces, by means of image retreival and image matching techniques, which are effectively different from the classification problem we are addressing.

2.12. Alternative approaches and the hybrid network

Alternative approaches were then attempted. The first was a few-shot learning code, specifically designed to learn from few examples.

The implemented architecture is a prototypical network for few-shot learning that maintains a multimodal approach, combining audio and image data processing. The system uses two parallel branches: one dedicated to audio, leveraging the pre-trained VGGish model for sound feature extraction, and one dedicated to images, based on VGGFace for visual feature analysis. The features extracted from both branches are then combined, maintaining the same mapping scheme used in the original neural network.

The core of this architecture is the prototypical mechanism, which operates by creating representative prototypes for each class from a small set of examples (support set). The system then classifies new examples by calculating their Euclidean distance from the prototypes and assigning them to the class of the most similar prototype.

To optimize performance and prevent overfitting, various regularization techniques were implemented as mentioned above for the original neural network, like dense layers of reduced dimension (32 neurons), dropout, and batch normalization. Transfer learning was applied by keeping most of VGGFace's layers frozen, allowing fine-tuning only of the last four layers.

The training phase was initially configured with moderate parameters: a learning rate of 0.0001, with batch size of 32. The use of TimeDistributed allowed efficient processing of the support set, maintaining temporal coherence in processing.

However, despite many attempts at parameter tuning, during testing the network always polarized to a single image. Therefore, this approach was abandoned.

It was then decided to turn to a more traditional technique like Support Vector Machine (SVM). SVMs are known to work well with small datasets and can be more resistant to overfitting compared to deep neural networks, using fewer parameters to optimize.

SVMs, especially with RBF (Radial Basis Function) kernel as in this case, can also be more robust to outliers compared to some deep learning models, and this kernel provides a good compromise between linearity and non-linearity.

While the neural network can capture complex and hierarchical patterns, the SVM is effective in finding the best hyperplane separation in feature space.

Nonetheless, this algorithm tended to polarize to recognizing the same single image as well.

A hybrid code approach was then attempted, which, leveraging the previously explained mappings and features recognized by the neural network during training, applies an SVM in cascade.

Essentially, the neural network is primarily used as a feature extractor. The features extracted by the neural network are then used as input for an SVM. By combining the two approaches, we hoped to leverage the strengths of both methods.

The SVM was configured with probability=True, allowing probability estimates for classes, which are then combined with those from the neural network.

In fact, the ANN and SVM models operate in parallel and their predictions are combined with weights (*e.g.*, 70% NN, 30% SVM), operating as a form of ensemble learning, specifically of the "weighted averaging" type. Indeed, these are two different models operating in parallel on the same task, through a weighted combination of predictions.

This made it possible to adjust the relative contribution of the models and compare individual performances, setting weights complementarily with sum 100%.

In this way, we tried to leverage the strengths of two very different models with complementary characteristics, seeking a balance between the two in search of better robustness compared to using a single model. We empirically tested the performance of different weight combinations between the two models, evaluating both the case of ANN = 0% and SVM = 100% and the reciprocal ANN = 100% and SVM = 0%, as well as various intermediate cases.

After multiple trials, the (relatively) best performance was achieved with the combination ANN = 70% and SVM = 30%, which was adopted as definitive.

Overall, this Machine Learning modality can be defined as a hybrid multimodal ensemble learning approach.

Despite the introduction of transfer learning and all the described techniques, the final result, as detailed below, did not deviate from substantial randomness.

It is indeed extremely difficult for any model, especially one as complex as a deep neural network, to generalize effectively. With so few examples, the model is essentially "memorizing" the training data. Techniques such as dropout, L2 regularization, data augmentation, etc., can help mitigate overfitting, but there is a limit to how effective they can be with such a small dataset.

On the other hand, the introduction of a more traditional algorithm like SVM only brought minimal improvements and, when used alone, did not show any useful result. In essence, none of the approaches tried – whether the original neural network, the prototypical network for few-shot learning, the SVM alone, or the hybrid ensemble approach – managed to overcome the fundamental limitation posed by the extremely limited size of the training dataset.

2.13. ML system output

After applying all the described improvement techniques, the (relatively) best results were obtained from the hybrid system by setting hyperparameters consistently with the need to reduce the tendency to overfit.

As above mentioned, we set the initial learning rate to a conservative value of 1e−5, with a minimum set to 1e−6, and the ratio ANN = 70%, SVM = 30%.

After various attempts, epochs were set to 100, but due to the Early Stopping used, the average across all iterations was approximately 50 (49.871795).

The loss and accuracy graphs of the neural network Figs. 7 and 8, though improved compared to the initially implemented models, highlight the inconsistency of performances.

In this configuration that we considered optimal, over 39 iterations, the hybrid system recognized 0 images 9 times, 2 images 4 times, 3 images 2 times, and 1 image 24 times, with an average of 0.054 recognitions per iteration, a value extremely close to chance (1/18 = 0.0556).

This value corresponds to the mean accuracy of the combined model (ANN+SVM) considering all iterations.

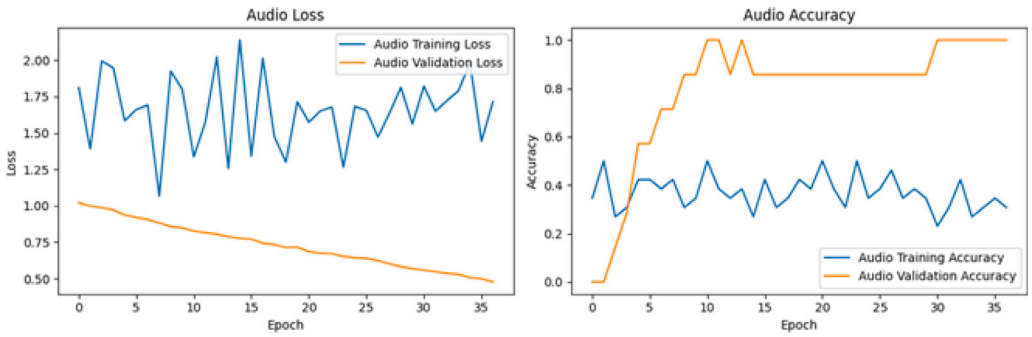


Fig. 7. Audio loss and accuracy.

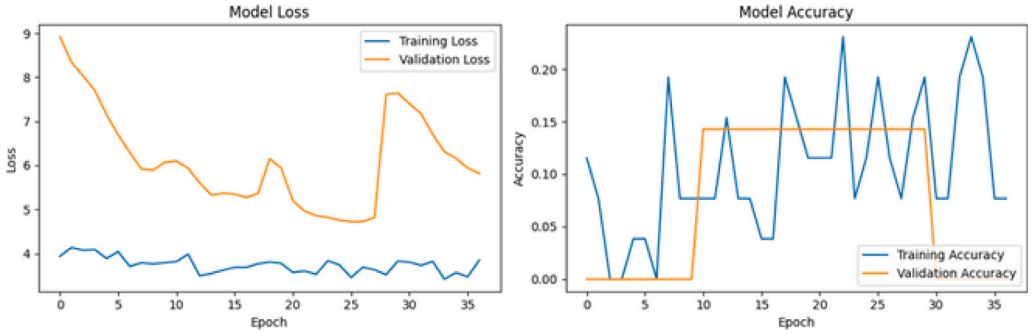


Fig. 8. Model loss and accuracy.

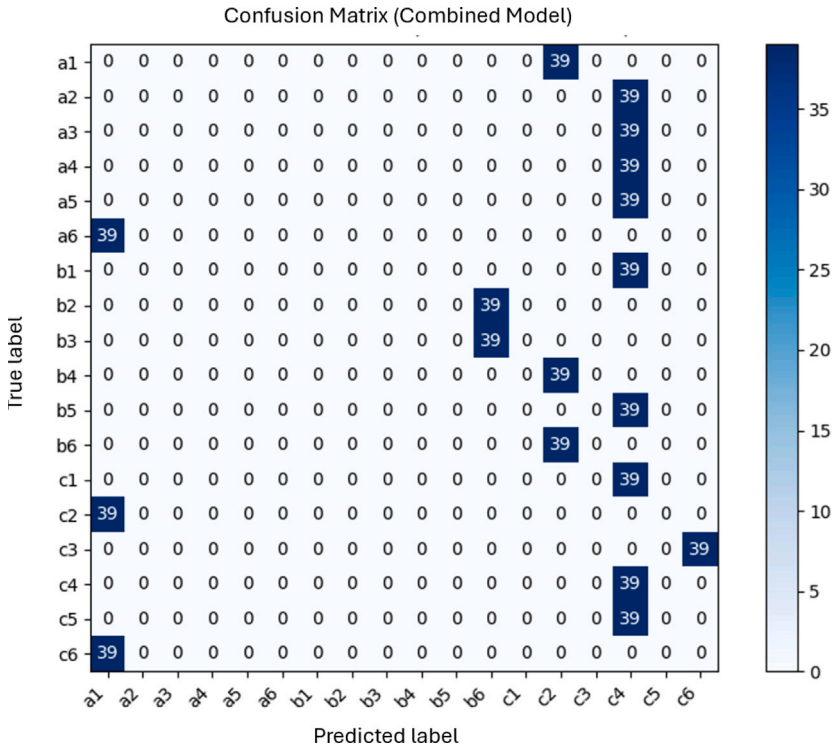


Fig. 9. Confusion matrix.

The confusion matrix generated remained identical both for each iteration and for the overall case of all iterations (Fig. 9)

The overall mean recognition probability for each class across the 18 images is indicated in Fig. 10 with small variations at each iteration, as shown in Fig. 11. Specifically, the overall mean recognition probability for each class across the 18 images is presented in Table 4 and the image with the highest overall mean probability was: c4 (0.1016).

The classes with the highest recognition probability per iteration are listed in Table 5.

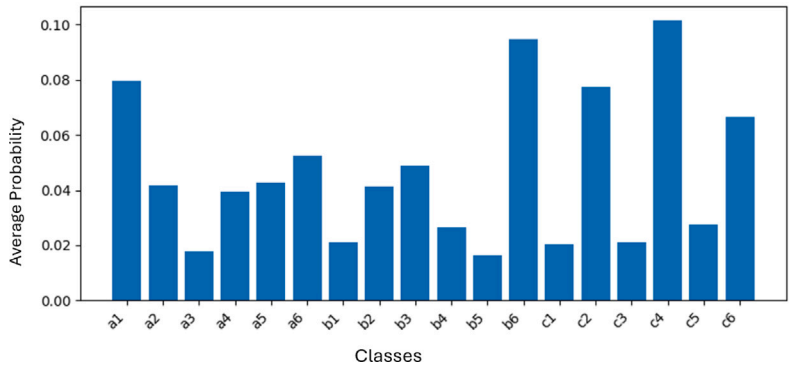


Fig. 10. Overall average recognition probabilities per class.

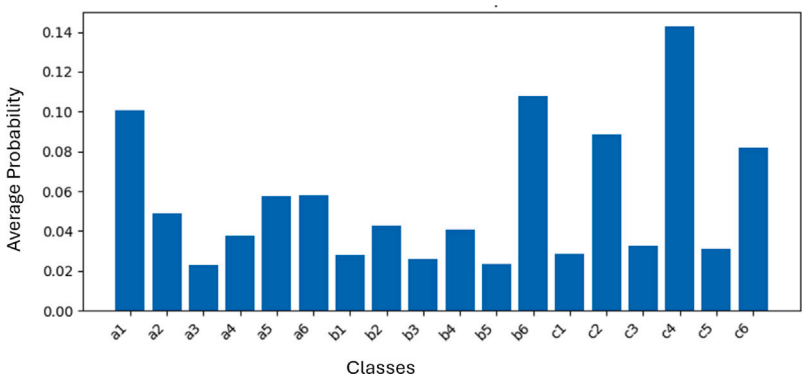


Fig. 11. Average recognition probabilities per class — iteration 1.

Table 4
Values for image classes a, b, and c.

Parameter	Value
a1	0.0794
a2	0.0416
a3	0.0178
a4	0.0395
a5	0.0428
a6	0.0524
b1	0.0210
b2	0.0413
b3	0.0490
b4	0.0264
b5	0.0165
b6	0.0948
c1	0.0205
c2	0.0773
c3	0.0210
c4	0.1016
c5	0.0276
c6	0.0667

Figs. 12 and 13 show the boxplots of respectively the distribution of average probabilities by group and the distribution of average probabilities by class.

2.13.1. Statistical analysis

To statistically evaluate the results and compare them with results obtained on human experimenters, we performed the same paired t-test using Welch's method.

In the human case the variability of results was very high from individual to individual, so the small sample size suggested Student's t-test as suitable for analyzing the data in question.

But in the machine case, although 39 iterations were performed, we saw that all produced identical confusion matrices, leading to a virtually null standard deviation (0.0132). This violates one of the fundamental assumptions of the t-test, which requires natural variability in the data. When the standard deviation is zero or nearly zero, the denominator in the t-test formula (which includes the standard deviation) approaches zero, making the test mathematically unstable and statistically inappropriate, as shown in Table 6. The small sample size is not the main issue in this case, but rather the lack of variability in the results.

Specifically, as seen, the mean probabilities observed over the 39 iterations are all very similar and close to chance level.

For this reason, it is more appropriate to use tests that are based directly on the observed proportions (such as chi-square) or a z-test of proportions rather than on variability between observations.

In this near-random performance scenario, not only is the variability between iterations null, but the differences between groups are minimal and lack practical significance. The chi-square test confirms this by showing in Table 7: - A very low Chi2 value (0.0000)

- A very high p-value ($1 \gg 0.05$)

This clearly indicates there is no significant difference between the two groups, and the small differences observed between groups are compatible with random fluctuations.

The z-test of proportions confirms this with Table 6:

- A low z-statistic (-0.2558)
- A high p-value ($0.7981 \gg 0.05$)

Table 5
Highest recognition probability per iteration.

Iteration	Parameter	Value	Iteration	Parameter	Value
1	c6	0.1265	21	a1	0.1764
2	c2	0.1671	22	b3	0.1369
3	a1	0.1179	23	b3	0.1140
4	a4	0.1018	24	b3	0.1429
5	a6	0.1197	25	a6	0.0877
6	b6	0.1561	26	b3	0.2010
7	b6	0.1406	27	b6	0.0979
8	b6	0.1643	28	b3	0.2167
9	c4	0.1101	29	b3	0.1464
10	c4	0.0864	30	b3	0.1018
11	a6	0.1339	31	b6	0.1088
12	b3	0.1287	32	b6	0.1237
13	c2	0.0697	33	c4	0.0921
14	a1	0.0983	34	b3	0.1844
15	c6	0.0911	35	b6	0.2800
16	b3	0.2172	36	b6	0.0674
17	c2	0.1930	37	a6	0.0862
18	b3	0.1296	38	a6	0.0815
19	c2	0.1173	39	c4	0.1347
20	b3	0.1381			

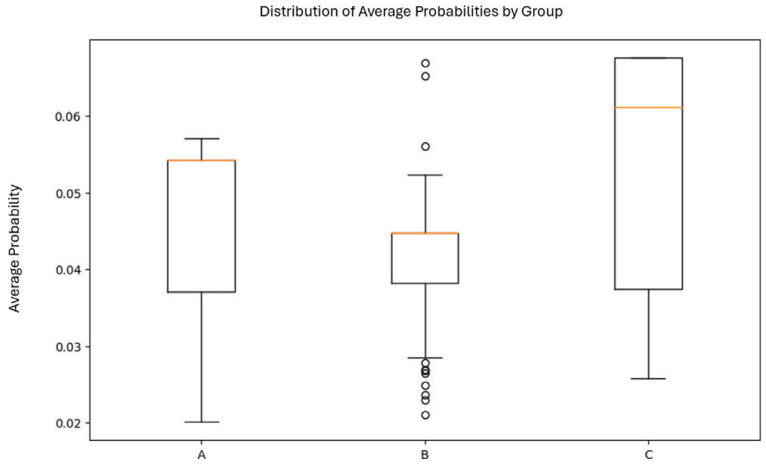


Fig. 12. Distribution of average probabilities by group.

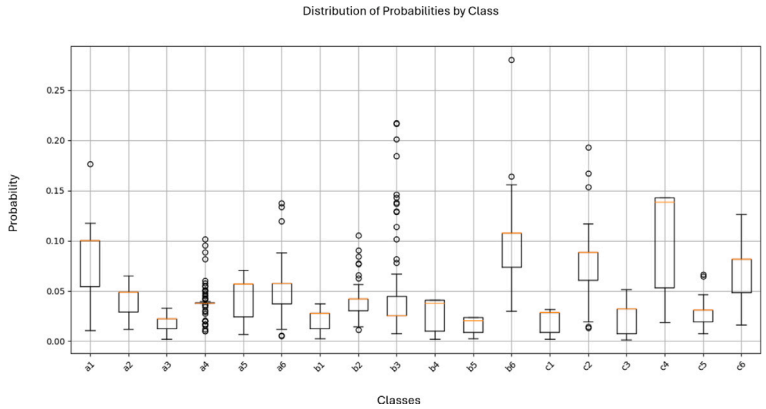


Fig. 13. Distribution of average probabilities by class.

supporting the conclusion that the difference is not statistically significant.

Both tests agree in showing there is no significant difference between group b and groups a/c combined, with very high p-values suggesting that the observed differences are compatible with random fluctuations. In other words, the model is essentially generating random predictions, making any apparent difference between groups non-significant. This links the near-random accuracy of the model with the impossibility of finding significant differences between groups, regardless of the statistical test used.

3. Conclusions and future developments

The ability to associate emotions unconsciously with images is a specific characteristic of human beings, as attested by the well-known "Takete-Maluma" test, due to Koehler (Köhler, 1970) and later taken

Table 6
Comparison between machine's recognition of different groups of images — Student's t-test results.

Mean of differences (b - ac)	−0.0075
Standard deviation of differences	0.0132
Overall mean probability for class 'a':	0.0456
Overall mean probability for class 'b':	0.0415
Overall mean probability for class 'c':	0.0525
Overall mean probability for classes 'a'/'c':	0.0490
Results of paired t-test (b vs combined a/c):	
t-statistic:	−5.02918
P-value:	3.13472e−06
Mean of differences (b - ac): −0.0075	
Standard deviation of differences	0.0132

Table 7
Chi-square Analysis and z-test for b vs Combined a/c.

Chi-square test b vs combined a/c:	
Mean probabilities:	
Group b:	0.0415 ± 0.0084
Groups a/c:	0.0490 ± 0.0034
Chi2:	0.0000
p-value:	1.0000
z-test of proportions (b vs a/c):	
z-statistic:	−0.2558
p-value:	0.7981
Difference (b - a/c):	−0.0075

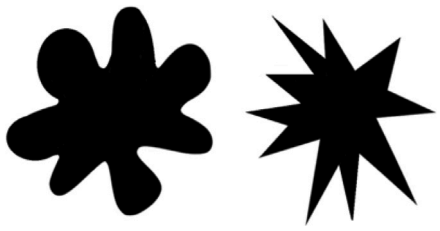


Fig. 14. On the left: a Bouba shape; On the right: a Kiki shape.

up by various authors, including in 2001 W.S. Ramachandran and E. Hubbard (Ramachandran, Marcus, & Chunharas, 2020). In this latest test, conducted on American university students and Tamil-speaking Indians, 95 percent associated the word "Bouba" with the figure on the left and "Kiki" with the one on the right (as shown in Fig. 14), despite the fact that the participants were from two completely different cultures (Bremner et al., 2013) and had never seen those figures or heard those two words, which are meaningless.

Because this association most likely comes from complex mental associations between sounds and images rooted in the unconscious of human beings, and it cannot come from logical evidence or prior knowledge, a software would not have the necessary data to produce the same answers.

The Turing Test proposed here is based on essentially the same considerations: the emotional world of human mind complexly involves any perception and influences cognitive responses involving those perceptions. This emotional substrate is not reproducible by software because it is not explicable in the form of data.

In the above-described experiment, administered to 39 participants, a video with music on a red background was displayed, and then they were asked to recognize 18 women's faces on a red background after seeing the same faces accompanied by different pieces of music, one of which was the music in the video.

Analyses performed on the experiment yielded two orders of results:

- Participants' responses revealed that images of faces that in training had been accompanied by a music (Music1) associated with the RED background color were recognized more often (to a statistically significant extent) than images that had been associated with different music, and for which therefore the red background had no emotional meaning.
- The group of participants who manifested more recognition of faces associated with Music1 than other faces and the group of participants who manifested the fewest such recognition were selected. For each participant in these groups, the signals corresponding to viewing the video and the faces used for training were analyzed with an Artificial Neural Network. To a statistically significant extent, the group with the best recognitions presented more identities (*i.e.*, identical binary codes) between signals corresponding to the RED Music1 video and Music1 images, than the other group.
- In order to complete the Turing Test and to fully validate our hypothesis that a software is unable to exhibit and thus exploit the cognitive property described above, the same experimental phases, with identical content and timing, has been administered to a suitably configured ML software.
- All attempts we made to implement software capable of emulating the ability shown by human experimenters to more easily distinguish certain images based on specific emotional intersections have failed. Statistical results show that the ML software arrives at completely random results.

These results appear to indicate a human memory enhancement by an unconscious emotional substrate. In this case, the unconscious association is the one that connects the red color paired with the music in the video, the faces whose image was accompanied by the same music, and the recall formed by the red color present as the background of the testing images.

The emotion associated with listening to music turns out to be an effective memorization tool: it appears that, at an unconscious level, it is possible to associate sensory experiences that have never been perceived simultaneously, as long as these evoke previously associated emotions.

The cognitive role of emotions in our view has to do with the qualitative aspects of our conscious experience, *i.e.*, what is called qualia (Chalmers, 1995; Dennett, 1988; Goguen, 2004; Northoff, 2003; Tye, 1994, 2017). We are therefore tempted to suggest that the difference that may emerge between humans and software in the proposed Turing Test is due to the fact that a computer does not have access to qualia.

This property of the human mind cannot be emulated by a computer, because the software must have read and stored in memory the data in order to be able to perform their recall. Our intent is to take advantage of this feature of the human mind to develop a novel Turing Test.

3.1. Future developments

3.1.1. Study expansion planning

Our initial study has validated the hypothesis that human beings possess a unique mental property allowing them to associate emotions with perceptions, thereby enhancing memory capabilities. While our results on 39 subjects achieved statistical significance, we recognize the importance of broader validation through an expanded research program that would strengthen the generalizability of our findings.

To this end, we envision expanding our study to include over 200 participants across multiple research centers, ensuring representation across age groups, gender, cultural backgrounds, and educational levels. Power analysis indicates this expanded sample size would provide greater than 95% confidence level in our findings. The methodological framework will be enhanced to include varied emotional stimuli, diverse testing intervals, and multiple types of musical and visual stimuli, with testing extended across different cultural contexts.

The technical aspects of our study should see significant advancement. We aim to expand our EEG analysis to encompass all available EMOTIV EPOC electrodes and frequency bands, while implementing standardized environmental conditions across testing sites. Furthermore, we plan to adopt clinical-grade EEG systems with full electrode arrays. Other non-invasive techniques could be combined with EEG to provide a broader spectrum of quantitative evaluations of emotional reactions, like all those used in the multimodal analyses mentioned in Section 1.5.1.

From a computational perspective, it would be valuable to further investigate the limitations of our implementation, which appears oversized in some aspects while not significantly benefiting from advanced or standard AI methods suitable for small samples. We therefore intend to conduct targeted research on technologies more suitable for the problem at hand, studying both the most advanced available techniques for solving the problem under examination, and implementing a battery of computational tests comprising diverse implementations.

These expansions would involve extending both the scope and depth of our analysis methodologies, focusing on the integration of multiple data streams, professional-grade equipment, specialized laboratory facilities for optimal data acquisition and processing, and advanced computational expertise for developing and implementing sophisticated AI solutions.

3.1.2. Possible future applications

Our approach is fundamentally grounded in a critical observation: software's capability to process the emotional components of sensory events remains limited compared to human abilities. This fundamental disparity presents a unique opportunity, particularly in the realm of security controls, where it could prove invaluable as existing CAPTCHAs become increasingly vulnerable to sophisticated software solutions. We anticipate the development of emotion-based CAPTCHAs, advanced biometric security systems incorporating emotional responses, and multi-factor authentication methods with emotion-based components.

Our approach also holds promise for advancing human–machine interaction through the development of emotionally-aware interfaces, adaptive learning systems, and enhanced virtual/augmented reality systems.

But the implications of this approach extend across multiple domains. In the medical field, our findings could enable early screening for neurological conditions affecting emotional memory, assessment of medication impacts, monitoring of recovery in patients with brain injuries, and differentiation between physiological and pathological cognitive decline.

In education, these findings could lead to optimized learning materials that leverage emotional-cognitive connections, personalized learning methods, and novel approaches to evaluating learning effectiveness through emotional markers.

Additionally, our methodology could find applications in marketing for testing emotional impact and branding effectiveness, in professional assessment for evaluating emotional intelligence and its role in decision-making, in forensics for analyzing testimony reliability and statement veracity.

3.2. Conclusion

While we have discussed the potential applications of the proposed method, we emphasize that this work primarily aims to contribute to basic science: in the field of neuroscience, by investigating the complex relationships between attention, memory, and emotions, leading to hypothesize a connection with the nature of qualia. In the computational field, by studying software's capability to emulate these functionalities.

We have attempted to highlight and quantify the limitations of computational capability in handling specific problems where emotions can indirectly influence memorization.

While it is certainly possible to improve or diversify the software we implemented, we believe that the obtained result is due to a fundamental deficiency in the current functionalities of Machine Learning compared to human cognitive and emotional capabilities.

Based on this deficiency, we consider it possible to implement a Turing Test robust enough to distinguish human behavior from computer behavior, at least in the short and medium term.

CRediT authorship contribution statement

Rita Pizzi: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Hao Quan:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Data curation. **Matteo Matteucci:** Writing – review & editing, Visualization, Validation. **Simone Mentasti:** Writing – review & editing, Visualization, Validation. **Roberto Sassi:** Writing – review & editing, Visualization, Validation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rita Pizzi reports equipment, drugs, or supplies was provided by Department of Computer Science University of Milan. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are indebted to Dr. Amidala Vishal Goud and Dr. Tobia Finzi for their in-depth insight that significantly contributed to the quality of this work.

Data availability

Data will be made available on request.

References

Achlioptas, P., Ovsjanikov, M., Guibas, L., & Tulyakov, S. (2023). Affection: Learning affective explanations for real-world visual data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6641–6651).

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, *1*(1), 21–62.

Ahmed, N., Al Aghbari, Z., & Girija, S. (2023). A systematic survey on multimodal emotion recognition using learning algorithms. *Intelligent Systems with Applications*, *17*, Article 200171.

Al-Dujaili, M. J., & Ebrahimi-Moghadam, A. (2023). Speech emotion recognition: a comprehensive survey. *Wireless Personal Communications*, *129*(4), 2525–2561.

Al Maruf, A., Khanam, F., Haque, M. M., Jiyad, Z. M., Mridha, F., & Aung, Z. (2024). Challenges and opportunities of text-based emotion detection: A survey. *IEEE Access*.

Al-Saadawi, H. F. T., Das, B., & Das, R. (2024). A systematic review of trimodal affective computing approaches: Text, audio, and visual integration in emotion recognition and sentiment analysis. *Expert Systems with Applications*, Article 124852.

Alluri, V., Brattico, E., Toiviainen, P., Burunat, I., Bogert, B., Numminen, J., et al. (2015). Musical expertise modulates functional connectivity of limbic regions during continuous music listening.. *Psychomusicology: Music, Mind, and Brain*, *25*(4), 443.

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., & Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *Neuroimage*, *59*(4), 3677–3689.

Alnfiai, M. (2020). A novel design of audio CAPTCHA for visually impaired users. *International Journal of Communication Networks and Information Security*, *12*(2), 168–179.

Alqahtani, F. H., & Alsulaiman, F. A. (2020). Is image-based CAPTCHA secure against attacks based on machine learning? An experimental study. *Computers & Security*, *88*, Article 101635.

Amazon (2023). Amazon lex. URL: https://aws.amazon.com/lex/.

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, *411*(6835), 305–309.

Baby, C. J., Khan, F. A., & Swathi, J. (2017). Home automation using IoT and a chatbot using natural language processing. In *2017 innovations in power and advanced computing technologies (i-PACT)* (pp. 1–6). IEEE.

Badcock, N. A., Preece, K. A., de Wit, B., Glenn, K., Fieder, N., Thie, J., et al. (2015). Validation of the Emotiv EPOC EEG system for research quality auditory event-related potentials in children. *PeerJ*, *3*, Article e907.

Baig, M. Z., & Kavakli, M. (2019). A survey on psycho-physiological analysis & measurement methods in multimodal systems. *Multimodal Technologies and Interaction*, *3*(2), 37.

Baumgartner, T., Lutz, K., Schmidt, C. F., & Jäncke, L. (2006). The emotional power of music: how music enhances the feeling of affective pictures. *Brain Research*, *1075*(1), 151–164.

Bhangale, K., & Kothandaraman, M. (2023). Speech emotion recognition based on multiple acoustic features and deep convolutional neural network. *Electronics*, *12*(4), 839.

Bhatlawande, S., Shilaskar, S., Pramanik, S., & Sole, S. (2024). Multimodal emotion recognition based on the fusion of vision, EEG, ECG, and EMG signals. *International Journal of Electrical and Computer Engineering Systems*, *15*(1), 41–58.

Bhavani, S. A., & Karthikeyan, C. (2024). Robust 3D face recognition in unconstrained environment using distance based ternary search siamese network. *Multimedia Tools and Applications*, *83*(17), 51925–51953.

Biswas, M., & Biswas, M. (2018). Microsoft bot framework. *Beginning AI Bot Frameworks: Getting Started with Bot Development*, 25–66.

Bordoloi, M., & Biswas, S. K. (2023). Sentiment analysis: A survey on design framework, applications and future scopes. *Artificial Intelligence Review*, *56*(11), 12505–12560.

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition*, *126*(2), 165–172.

Brooks, S. J., Savov, V., Allzén, E., Benedict, C., Fredriksson, R., & Schiöth, H. B. (2012). Exposure to subliminal arousing stimuli induces robust activation in the amygdala, hippocampus, anterior cingulate, insular cortex and primary visual cortex: a systematic meta-analysis of fMRI studies. *Neuroimage*, *59*(3), 2962–2973.

Buchanan, T. W. (2007). Retrieval of emotional memories. *Psychological Bulletin*, *133*(5), 761.

Buchanan, T. W., & Lovallo, W. R. (2001). Enhanced memory for emotional material following stress-level cortisol treatment in humans. *Psychoneuroendocrinology*, *26*(3), 307–317.

Bursztein, E., Martin, M., & Mitchell, J. (2011). Text-based CAPTCHA strengths and weaknesses. In *Proceedings of the 18th ACM conference on computer and communications security* (pp. 125–138).

Cahill, L., & McGaugh, J. L. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition*, *4*(4), 410–421.

Cahn, J. (2017). CHATBOT: Architecture, design, & development. *University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science*.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, *28*(2), 15–21.

Canal, F. Z., Müller, T. R., Matias, J. C., Scotton, G. G., de Sa Junior, A. R., Pozzebon, E., et al. (2022). A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Information Sciences*, *582*, 593–617.

Carmichael, D. W., Vulliemoz, S., Murta, T., Chaudhary, U., Perani, S., Rodionov, R., et al. (2024). Measurement of the mapping between intracranial EEG and fMRI recordings in the human brain. *Bioengineering*, *11*(3), 224.

Chalmers, D. J. (1995). Absent qualia, fading qualia, dancing qualia. *Conscious Experience*, 309–328.

Chen, C.-M., & Wang, H.-P. (2011). Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance. *Library & Information Science Research*, *33*(3), 244–255.

Choudhary, S., Saroha, R., Dahiya, Y., & Choudhary, S. (2013). Understanding CAPTCHA: text and audio based CAPTCHA with its applications. *International Journal of Advanced Research in Computer Science and Software Engineering*, *3*(6).

Chow, Y.-W., Susilo, W., & Thorncharoensri, P. (2019). CAPTCHA design and security issues. *Advances in Cyber Security: Principles, Techniques, and Applications*, 69–92.

Chowdhary, K., & Chowdhary, K. (2020). Natural language processing. *Fundamentals of Artificial Intelligence*, 603–649.

Chubarov, A., & Azarnov, D. (2018). Modeling behavior of virtual actors: a limited Turing test for social-emotional intelligence. In *Biologically inspired cognitive architectures (BICA) for Young scientists: Proceedings of the first international early research career enhancement school on BICA and cybersecurity* (pp. 34–40). Springer.

Colby, K. M. (2013). *vol. 49, Artificial Paranoia: A Computer Simulation of Paranoid Processes*. Elsevier.

Critchley, H. D., & Harrison, N. A. (2013). Visceral influences on brain and behavior. *Neuron*, *77*(4), 624–638.

de Lope, J., & Graña, M. (2023). An ongoing review of speech emotion recognition. *Neurocomputing*, *528*, 1–11.

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211.

Dennett, D. C. (1988). Quining qualia. In *The first version of this paper was presented at university college, London, England, in Nov 1978. A second version was presented at the Universities of Adelaide and Sydney in 1984, and in 1985 to Psychology Department Colloquia at Harvard University and Brown University*. Clarendon Press/Oxford University Press.

Dinh, N. T., & Hoang, V. T. (2023). Recent advances of captcha security analysis: a short literature review. *Procedia Computer Science*, *218*, 2550–2562.

Egli, A. (2023). ChatGPT, GPT-4, and other large language models: The next revolution for clinical microbiology? *Clinical Infectious Diseases*, *77*(9), 1322–1328.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124.

El-Latif, E. I. A., El-Sayad, N. E., Mohammed, K. K., Darwish, A., & Hassanien, A. E. (2024). Vggish transfer learning model for the efficient detection of payload weight of drones using mel-spectrogram analysis. *Neural Computing and Applications*, 1–17.

Elkins, K., & Chun, J. (2020). Can GPT-3 pass a writer's Turing test? *Journal of Cultural Analytics*, *5*(2).

Emotiv (2023). EMotiv. URL: https://www.emotiv.com.

Erat, K., Şahin, E. B., Doğan, F., Merdanoğlu, N., Akcakaya, A., & Durdu, P. O. (2024). Emotion recognition with EEG-based brain-computer interfaces: a systematic literature review. *Multimedia Tools and Applications*, *83*(33), 79647–79694.

Ermentrout, B. (1992). Complex dynamics in WTA neural networks with slow inhibition. *Neural Networks*, *5*, 403–409.

Etkin, A., Büchel, C., & Gross, J. J. (2015). The neural bases of emotion regulation. *Nature Reviews. Neuroscience*, *16*(11), 693–700.

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 1–15.

Fei, Z., Yang, E., Li, D. D.-U., Butler, S., Ijomah, W., Li, X., et al. (2020). Deep convolution network based emotion analysis towards mental health care. *Neurocomputing*, *388*, 212–227.

Floreani, E. D., Orlandi, S., & Chau, T. (2022). A pediatric near-infrared spectroscopy brain-computer interface based on the detection of emotional valence. *Frontiers in Human Neuroscience*, *16*, Article 938708.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*(4), 681–694.

Fu, Z., Liu, F., Zhang, J., Wang, H., Yang, C., Xu, Q., et al. (2021). SAGN: semantic adaptive graph network for skeleton-based human action recognition. In *Proceedings of the 2021 international conference on multimedia retrieval* (pp. 110–117).

Gao, J. (2024). Exploring key technologies for multimodal emotion recognition: Research and application analysis. In *AIP conference proceedings (Vol. 3194, No. 1)*. AIP Publishing.

Gao, J., Li, P., Chen, Z., & Zhang, J. (2020). A survey on deep learning for multimodal data fusion. *Neural Computation*, *32*(5), 829–864.

Gao, H., Liu, H., Yao, D., Liu, X., & Aickelin, U. (2010). An audio CAPTCHA to distinguish humans from computers. In *2010 third international symposium on electronic commerce and security* (pp. 265–269). IEEE.

Ge, H., Zhu, Z., Dai, Y., Wang, B., & Wu, X. (2022). Facial expression recognition based on deep learning. *Computer Methods and Programs in Biomedicine*, *215*, Article 106621.

Gehl, R. W. (2013). Teaching to the Turing test with Cleverbot. *Transformations: The Journal of Inclusive Scholarship and Pedagogy*, *24*(1–2), 56–66.

Geman, D., Geman, S., Hallonquist, N., & Younes, L. (2015). Visual Turing test for computer vision systems. *Proceedings of the National Academy of Sciences*, *112*(12), 3618–3623.

Geng, Y., Shi, S., & Hao, X. (2024). Deep learning-based EEG emotion recognition: a comprehensive review. *Neural Computing and Applications*, 1–32.

Goguen, J. (2004). Musical qualia, context, time and emotion. *Journal of Consciousness Studies*, *11*(3–4), 117–147.

Gohumpu, J., Xue, M., & Bao, Y. (2023). Emotion recognition with multi-modal peripheral physiological signals. *Frontiers in Computer Science*, *5*, Article 1264713.

Gong, P., Jia, Z., Wang, P., Zhou, Y., & Zhang, D. (2023). ASTDF-net: attention-based spatial-temporal dual-stream fusion network for EEG-based emotion recognition. In *Proceedings of the 31st ACM international conference on multimedia* (pp. 883–892).

Gossweiler, R., Kamvar, M., & Baluja, S. (2009). What's up CAPTCHA? A CAPTCHA based on image orientation. In *Proceedings of the 18th international conference on world wide web* (pp. 841–850).

Guerar, M., Verderame, L., Migliardi, M., Palmieri, F., & Merlo, A. (2021). Gotta CAPTCHA'Em all: a survey of 20 Years of the human-or-computer Dilemma. *ACM Computing Surveys*, *54*(9), 1–33.

Hamborg, F., Donnay, K., Merlo, P., et al. (2021). *NewsMTSC: a dataset for (multi-) target-dependent sentiment classification in political news articles*. Association for Computational Linguistics (ACL).

Haque, Y., Zawad, R. S., Rony, C. S. A., Al Banna, H., Ghosh, T., Kaiser, M. S., et al. (2024). State-of-the-art of stress prediction from heart rate variability using artificial intelligence. *Cognitive Computation*, *16*(2), 455–481.

Hatipoglu Yilmaz, B., Kose, C., & Yilmaz, C. M. (2024). A novel multimodal EEG-image fusion approach for emotion recognition: introducing a multimodal KMED dataset. *Neural Computing and Applications*, 1–16.

He, C., Chen, Y.-Y., Phang, C.-R., Stevenson, C., Chen, I.-P., Jung, T.-P., et al. (2023). Diversity and suitability of the state-of-the-art wearable and wireless EEG systems review. *IEEE Journal of Biomedical and Health Informatics*.

Hernández-Orallo, J. (2020). Twenty years beyond the Turing test: moving beyond the human judges too. *Minds and Machines*, *30*(4), 533–562.

Ho, M.-T. (2022). What is a Turing test for emotional AI? *AI & Society*, 1–2.

Huang, W. (2021). Elderly depression recognition based on facial micro-expression extraction. *Traitement Du Signal*, *38*(4).

Huster, R. J., Debener, S., Eichele, T., & Herrmann, C. S. (2012). Methods for simultaneous EEG-fMRI: an introductory review. *Journal of Neuroscience*, *32*(18), 6053–6060.

Huu, P. N., Thi, A. P., Danh, Q. M., Quynh, N. C., Hoang, T. M., & Minh, Q. T. (2022). Proposing algorithm to localize and extract facial information using FaceNet and MTCNN. In *2022 international conference on data analytics for business and industry* (pp. 587–592). IEEE.

IBM (2023). IBM watson assistant. URL: https://www.ibm.com/it-it/products/watsonx-assistant.

Ilyas, C. M. A., Nunes, R., Nasrollahi, K., Rehm, M., & Moeslund, T. B. (2021). Deep emotion recognition through upper body movements and facial expression.. In *VISIGRAPP (5: VISAPP)* (pp. 669–679).

Jacquet, B., Jamet, F., & Baratgin, J. (2021). On the pragmatics of the Turing test. In *2021 international conference on information and digital technologies* (pp. 123–130). IEEE.

Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, *19*(11), 2579–2594.

Kalateh, S., Estrada-Jimenez, L. A., Hojjati, S. N., & Barata, J. (2024). A systematic review on multimodal emotion recognition: Building blocks, current state, applications, and challenges. *IEEE Access*.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.

kapodi (2024). The searchable database of free emotional stimuli sets. URL: https://airtable.com/applhKyjKmrljh2wu/shrnVoUZrwu6riP9b/tblqcLYW1VCOif5FU/viw9aY5JrvOMmtbQh?blocks=hide.

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of Nonverbal Behavior*, *43*, 133–160.

Kensinger, E. A., & Schacter, D. L. (2006). Amygdala activity is associated with the successful encoding of item, but not source, information for positive and negative stimuli. *Journal of Neuroscience*, *26*(9), 2564–2570.

Khiani, S., Iqbal, M. M., Dhakne, A., Thrinath, B. S., Gayathri, P., & Thiagarajan, R. (2022). An effectual IOT coupled EEG analysing model for continuous patient monitoring. *Measurement: Sensors*, *24*, Article 100597.

Kisley, M. A., & Cornwell, Z. M. (2006). Gamma and beta neural activity evoked during a sensory gating paradigm: effects of auditory, somatosensory and cross-modal stimulation. *Clinical Neurophysiology*, *117*(11), 2549–2563.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *Neuroimage*, *42*(2), 998–1031.

Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., et al. (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion*, Article 101861.

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews. Neuroscience*, *15*(3), 170–180.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, *3*(1), 18–31.

Köhler, W. (1970). *Gestalt psychology: An introduction to new concepts in modern psychology (Vol. 18)*. WW Norton & Company.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464–1480.

Kraack, K. (2024). A multimodal emotion recognition system: Integrating facial expressions, body movement, speech, and spoken language. arXiv preprint arXiv:2412.17907.

Krolak-Salmon, P., Hénaff, M.-A., Vighetto, A., Bertrand, O., & Mauguière, F. (2004). Early amygdala reaction to fear spreading in occipital, temporal, and frontal cortex: a depth electrode ERP study in human. *Neuron*, *42*(4), 665–676.

Kumar, M., Jindal, M., & Kumar, M. (2022). A systematic survey on CAPTCHA recognition: types, creation and breaking techniques. *Archives of Computational Methods in Engineering*, *29*(2), 1107–1136.

Kusal, S., Patil, S., Kotecha, K., Aluvalu, R., & Varadarajan, V. (2021). AI based emotion detection for textual big data: techniques and contribution. *Big Data and Cognitive Computing*, *5*(3), 43.

Lecler, A., Duron, L., & Soyer, P. (2023). Revolutionizing radiology with GPT-based models: Current applications, future possibilities and limitations of ChatGPT. *Diagnostic and Interventional Imaging*, *104*(6), 269–274.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*(1), 155–184.

Lee, J., Kim, S., Kim, S., Park, J., & Sohn, K. (2019). Context-aware emotion recognition networks. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10143–10152).

Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, *13*(3), 1195–1215.

Li, X., La, R., Wang, Y., Niu, J., Zeng, S., Sun, S., et al. (2019). EEG-based mild depression recognition using convolutional neural network. *Medical & Biological Engineering & Computing*, *57*, 1341–1352.

Li, Q., Zhan, S., Xu, L., & Wu, C. (2019). Facial micro-expression recognition based on the fusion of deep learning and enhanced optical flow. *Multimedia Tools and Applications*, *78*, 29307–29322.

Lian, H., Lu, C., Li, S., Zhao, Y., Tang, C., & Zong, Y. (2023). A survey of deep learning-based multimodal emotion recognition: Speech, text, and face. *Entropy*, *25*(10), 1440.

Liddy, E. D. (2001). Natural language processing.

Liégeois-Chauvel, C., Bénar, C., Krieg, J., Delbé, C., Chauvel, P., Giusiano, B., et al. (2014). How functional coupling between the auditory cortex and the amygdala induces musical emotion: a single case study. *Cortex*, *60*, 82–93.

Liu, F. (2024). Artificial intelligence in emotion quantification: A prospective overview. *CAAI Artificial Intelligence Research*, *3*.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., et al. (2023). Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-Radiology*, Article 100017.

Liu, F., Wang, H.-Y., Shen, S.-Y., Jia, X., Hu, J.-Y., Zhang, J.-H., et al. (2022). OPO-FCM: a computational affection based OCC-PAD-OCEAN federation cognitive modeling approach. *IEEE Transactions on Computational Social Systems*, *10*(4), 1813–1825.

Luo, J., Zhang, G., Su, Y., Lu, Y., Pang, Y., Wang, Y., et al. (2022). Quantitative analysis of heart rate variability parameter and mental stress index. *Frontiers in Cardiovascular Medicine*, *9*, Article 930745.

Lyu, Q., Tan, J., Zapadka, M. E., Ponnatapura, J., Niu, C., Myers, K. J., et al. (2023). Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential. *Visual Computing for Industry, Biomedicine, and Art*, *6*(1), 9.

Madathil, K. C., Greenstein, J. S., & Horan, K. (2019). Empirical studies to investigate the usability of text-and image-based CAPTCHAs. *International Journal of Industrial Ergonomics*, *69*, 200–208.

Marg, E. (1995). Descartes'error: emotion, reason, and the human brain. *Optometry and Vision Science*, *72*(11), 847–848.

Masserman, J. H. (1941). Is the hypothalamus a center of emotion? *Psychosomatic Medicine*, *3*(1), 3–25.

Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science*, *6*(2), 114–133.

McGaugh, J. L. (2000). Memory–a century of consolidation. *Science*, *287*(5451), 248–251.

Medjden, S., Ahmed, N., & Lataifeh, M. (2020). Adaptive user interface design and analysis using emotion recognition through facial expressions and body posture from an RGB-D sensor. *PLoS One*, *15*(7), Article e0235908.

Mellouk, W., & Handouzi, W. (2020). Facial emotion recognition using deep learning: review and insights. *Procedia Computer Science*, *175*, 689–694.

Misra, D., & Gaj, K. (2006). Face recognition captchas. In *Advanced int'l conference on telecommunications and int'l conference on internet and web applications and services* (p. 122). IEEE.

Müller, V. C., & Ayesh, A. (2012). Revisiting Turing and his test: Comprehensiveness, qualia, and the real world.

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, *18*(5), 544–551.

Nagarhalli, T. P., Vaze, V., & Rana, N. (2020). A review of current trends in the development of chatbot systems. In *2020 6th international conference on advanced computing and communication systems* (pp. 706–710). IEEE.

Nagels-Coune, L., Riecke, L., Benitez-Andonegui, A., Klinkhammer, S., Goebel, R., De Weerd, P., et al. (2021). See, hear, or feel–to speak: a versatile multiple-choice functional near-infrared spectroscopy-brain-computer interface feasible with visual, auditory, or tactile instructions. *Frontiers in Human Neuroscience*, *15*, Article 784522.

Natale, S. (2021). *Deceitful media: Artificial intelligence and social life after the Turing test*. USA: Oxford University Press.

Neufeld, E., & Finnestad, S. (2020). In defense of the Turing test. *AI & Society*, *35*, 819–827.

Northoff, G. (2003). Qualia and the ventral prefrontal cortical function 'neurophenomenological' hypothesis. *Journal of Consciousness Studies*, *10*(8), 14–48.

Olague, G., Olague, M., Jacobo-Lopez, A. R., & Ibarra-Vazquez, G. (2021). Less is more: pursuing the visual Turing test with the Kuleshov effect. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1553–1561).

OpenAI (2023). OpenAI. URL: https://openai.com.

Oppy, G., & Dowe, D. (2003). The Turing test.

Pan, J., Wang, L., Huang, H., Xiao, J., Wang, F., Liang, Q., et al. (2022). A hybrid brain–computer interface combining P300 potentials and emotion patterns for detecting awareness in patients with disorders of consciousness. *IEEE Transactions on Cognitive and Developmental Systems*, *15*(3), 1386–1395.

Parkhi, O., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In *BMVC 2015-proceedings of the british machine vision conference 2015*. British Machine Vision Association.

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology*, *56*, 89–114.

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews. Neuroscience*, *9*(2), 148–158.
Pessoa, L. (2018). Understanding emotion with brain networks. *Current Opinion in Behavioral Sciences*, *19*, 19–25.
Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, *14*(2), 198–202.
Phelps, E. A. (2006). Emotion and cognition: insights from studies of the human amygdala. *Annual Review of Psychology*, *57*, 27–53.
Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*, *48*(2), 175–187.
Phelps, E. A., & Sharot, T. (2008). How (and why) emotion enhances the subjective sense of recollection. *Current Directions in Psychological Science*, *17*(2), 147–152.
Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research*, *209*, 481–493.
Pizzi, R. (2020). An artificial neural network compares neurophysiological events triggered by mutually associated sensory and cognitive stimuli. *International Journal of Engineering Research and Applications*, *10*(9 (Serie 6.)), 45–49.
Pizzi, R., Cino, G., Gelain, F., Rossetti, D., & Vescovi, A. (2007). Learning in human neural networks on microelectrode arrays. *Biosystems*, *88*(1–2), 1–15.
Pizzi, R., de Curtis, M., & Dickson, C. (2003). Evidence of chaotic attractors in cortical fast OscillationsTested by an artificial neural network. In *Soft computing applications* (pp. 11–22). Springer.
Pizzi, R., Musumeci, M., et al. (2017a). *Artificial neural networks, dynamical systems and self-organization*. CreateSpace Independent Publishing Platform.
Pizzi, R., Musumeci, M., et al. (2017b). Coding mental states from EEG signals and evaluating their integreted information content: a computational intelligence approach. *Internationa Journal of Circuits, System and Signals Processing*, *11*(4464), 464–470.
Pizzi, R. M., Rossetti, D., Cino, G., Marino, D., Vescovi, A. L., & Baer, W. (2009). A cultured human neural network operates a robotic actuator. *Biosystems*, *95*(2), 137–144.
Proudfoot, D. (2020). Rethinking Turing's test and the philosophical implications. *Minds and Machines*, *30*(4), 487–512.
Qin, J., Zong, L., & Liu, F. (2024). Exploring inner speech recognition via cross-perception approach in EEG and fMRI. *Applied Sciences*, *14*(17), 7720.
Ramachandran, V. S., Marcus, Z., & Chunharas, C. (2020). Bouba-Kiki: Cross-domain resonance and the origins of synesthesia, metaphor, and words in the human mind. In *Multisensory perception* (pp. 3–40). Elsevier.
Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*.
Reisenzein, R., Hudlicka, E., Dastani, M., Gratch, J., Hindriks, K., Lorini, E., et al. (2013). Computational modeling of emotion: Toward improving the inter-and intradisciplinary exchange. *IEEE Transactions on Affective Computing*, *4*(3), 246–266.
Rimmele, U., Davachi, L., Petrov, R., Dougal, S., & Phelps, E. A. (2011). Emotion enhances the subjective feeling of remembering, despite lower accuracy for contextual details. *Emotion*, *11*(3), 553.
Ritter, H., & Schulten, K. (1986). On the stationary state of Kohonen's self-organizing sensory mapping. *Biological Cybernetics*, *54*(2), 99–106.
Ritter, H., & Schulten, K. (1988). Convergence properties of Kohonen's topology conserving maps: fluctuations, stability, and dimension selection. *Biological Cybernetics*, *60*(1), 59–71.
Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161.
Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin*, *110*(3), 426.
Salimpoor, V. N., Van Den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., & Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science*, *340*(6129), 216–219.
Sallam, M., Salim, N., Barakat, M., & Al-Tammemi, A. (2023). ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive study highlighting the advantages and limitations. *Narra J*, *3*(1), e103.
Salvagno, M., Taccone, F. S., Gerli, A. G., et al. (2023). Can artificial intelligence help for scientific writing? *Critical Care*, *27*(1), 1–5.
Sarvakar, K., Senkamalavalli, R., Raghavendra, S., Kumar, J. S., Manjunath, R., & Jaiswal, S. (2023). Facial emotion recognition using convolutional neural networks. *Materials Today: Proceedings*, *80*, 3560–3564.
Saxena, A., Khanna, A., & Gupta, D. (2020). Emotion recognition and detection methods: A comprehensive survey. *Journal of Artificial Intelligence and Systems*, *2*(1), 53–79.
Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815–823).
Sharma, N. A., Ali, A. S., & Kabir, M. A. (2024). A review of sentiment analysis: tasks, applications, and deep learning techniques. *International Journal of Data Science and Analytics*, 1–38.
Sharot, T., & Phelps, E. A. (2004). How arousal modulates memory: Disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 294–306.
Shen, S., Liu, F., Wang, H., Wang, Y., & Zhou, A. (2024). Temporal shift module with pretrained representations for speech emotion recognition. *Intelligent Computing*, *3*, 0073.
Shepley, A. J. (2019). Deep learning for face recognition: a critical analysis. ArXiv preprint arXiv:1907.12739.
Singh, Y., & Biswas, A. (2022). Robustness of musical features on deep learning models for music genre classification. *Expert Systems with Applications*, *199*, Article 116879.
Singh, V. P., & Pal, P. (2014). Survey of different types of CAPTCHA. *International Journal of Computer Science and Information Technologies*, *5*(2), 2242–2245.
Stark, L. (2019). Affect and emotion in digitalSTS. *DigitalSTS: A Field Guide for Science & Technology Studies*, 117–135.
Steriade, M., Gloor, P., Llinas, R. R., Da Silva, F. L., & Mesulam, M.-M. (1990). Basic mechanisms of cerebral rhythmic activities. *Electroencephalography and Clinical Neurophysiology*, *76*(6), 481–508.
Sterrett, S. G. (2000). Turing's two tests for intelligence. *Minds and Machines*, *10*(4), 541–559.
Stevens, F. L., Hurley, R. A., & Taber, K. H. (2011). Anterior cingulate cortex: unique role in cognition and emotion. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *23*(2), 121–125.
Strange, B. A., Hurlemann, R., & Dolan, R. J. (2003). An emotion-induced retrograde amnesia in humans is amygdala-and β-adrenergic-dependent. *Proceedings of the National Academy of Sciences*, *100*(23), 13626–13631.
Sukhani, K., Sawant, S., Maniar, S., & Pawar, R. (2021). Automating the bypass of image-based CAPTCHA and assessing security. In *2021 12th international conference on computing communication and networking technologies* (pp. 01–08). IEEE.
Suzuki, Y., & Tanaka, S. C. (2021). Functions of the ventromedial prefrontal cortex in emotion regulation under stress. *Scientific Reports*, *11*(1), 18225.
Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science*, *14*(5), 455–461.
Tallon-Baudry, C., Kreiter, A., & Bertrand, O. (1999). Sustained and transient oscillatory responses in the gamma and beta bands in a visual short-term memory task in humans. *Visual Neuroscience*, *16*(3), 449–459.
Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.-T., et al. (2022). Lamda: Language models for dialog applications. ArXiv preprint arXiv:2201.08239.
Torse, D. A., Khanai, R., Pai, K., & Iyer, S. (2022). Hardware implementation of automated seizure detection system using EEG signals and edge computing. In *2022 6th international conference on trends in electronics and informatics* (pp. 472–477). IEEE.
Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, [ISSN: 0026-4423] *LIX*(236), 433–460.
Turing, A. M. (2009). *Computing machinery and intelligence*. Springer.
Tye, M. (1994). Qualia, content, and the inverted spectrum. *Noûs*, *28*(2), 159–183.
Tye, M. (2017). Philosophical problems of consciousness. *The Blackwell Companion To Consciousness*, 17–31.
Von Ahn, L., Blum, M., Hopper, N. J., & Langford, J. (2003). CAPTCHA: Using hard AI problems for security. In *Advances in cryptology—EUROCRYPT 2003: international conference on the theory and applications of cryptographic techniques, warsaw, Poland, May 4–8, 2003 proceedings 22* (pp. 294–311). Springer.
Wan, L., Liu, N., Huo, H., & Fang, T. (2017). Face recognition with convolutional neural networks and subspace learning. In *2017 2nd international conference on image, vision and computing* (pp. 228–233). IEEE.
Wang, M., & Deng, W. (2021). Deep face recognition: A survey. *Neurocomputing*, *429*, 215–244.
Wang, P., Gao, H., Guo, X., Xiao, C., Qi, F., & Yan, Z. (2023). An experimental investigation of text-based CAPTCHA attacks and their robustness. *ACM Computing Surveys*, *55*(9), 1–38.
Wang, H., Li, B., Wu, S., Shen, S., Liu, F., Ding, S., et al. (2023). Rethinking the learning paradigm for dynamic facial expression recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 17958–17968).
Wang, X., Peng, J., Zhang, S., Chen, B., Wang, Y., & Guo, Y. (2022). A survey of face recognition. ArXiv preprint arXiv:2212.13038.
Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., et al. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*, *83*, 19–52.
Wang, W., Xu, K., Niu, H., & Miao, X. (2020). [Retracted] emotion recognition of students based on facial expressions in online education based on the perspective of computer simulation. *Complexity*, *2020*(1), Article 4065207.
Watanabe, T., Yagishita, S., & Kikyo, H. (2008). Memory of music: roles of right hippocampus and left inferior frontal gyrus. *Neuroimage*, *39*(1), 483–491.
Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.
Wheeler, M. (2020). Deceptive appearances: The Turing test, response-dependence, and intelligence as an emotional concept. *Minds and Machines*, *30*(4), 513–532.
Williams, N. S., King, W., Mackellar, G., Randeniya, R., McCormick, A., & Badcock, N. A. (2023). Cowdsourced EEG Experiments: A proof of concept for remote EEG acquisition using EmotivPRO Builder and EmotivLABS. *Heliyon*, *9*(8).
Williams, N. S., McArthur, G. M., de Wit, B., Ibrahim, G., & Badcock, N. A. (2020). A validation of Emotiv EPOC Flex saline for EEG and ERP research. *PeerJ*, *8*, Article e9713.

Winecoff, A., Clithero, J. A., Carter, R. M., Bergman, S. R., Wang, L., & Huettel, S. A. (2013). Ventromedial prefrontal cortex encodes emotional value. *Journal of Neuroscience*, *33*(27), 11032–11039.

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., et al. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, *10*(5), 1122–1136.

Wu, Y., & Li, J. (2023). Multi-modal emotion identification fusing facial expression and EEG. *Multimedia Tools and Applications*, *82*(7), 10901–10919.

Xiao, J., & Wu, J. (2023). Effectiveness of the neuroimaging techniques in the recognition of psychiatric disorders: A systematic review and meta-analysis of RCTs. *Current Medical Imaging*, *20*(1), E260523217379.

Xolmurotova, S., & Adilova, S. (2023). Shaxs ijtimoiylashuvida ratsional-emotiv psixoterapiyaning samaradorligi. *Interpretation and Researches*, *2*(3).

Xu, Y., Lin, Y.-S., Zhou, X., & Shan, X. (2024). Utilizing emotion recognition technology to enhance user experience in real-time. *Computing and Artificial Intelligence*, *2*(1), 1388.

Xu, X., Liu, L., & Li, B. (2020). A survey of CAPTCHA technologies to distinguish between human and computer. *Neurocomputing*, *408*, 292–307.

Zaman, K., Sah, M., Direkoglu, C., & Unoki, M. (2023). A survey of audio classification using deep learning. *IEEE Access*.

Zhang, W., Deng, Y., Liu, B., Pan, S. J., & Bing, L. (2023). Sentiment analysis in the era of large language models: A reality check. ArXiv preprint arXiv:2305.15005.

Zhang, N., Ebrahimi, M., Li, W., & Chen, H. (2022). Counteracting dark web text-based CAPTCHA with generative adversarial learning for proactive cyber threat intelligence. *ACM Transactions on Management Information Systems (TMIS)*, *13*(2), 1–21.

Zimmerman, D. (2016). Thinking with your hypothalamus: Reflections on a cognitive role for the reactive emotions. In *Free will and reactive attitudes* (pp. 255–272). Routledge.