The ‘Chinese Room’: Humans’ Emotional Projections Onto Machines

Aoyun Xie
Affiliation: Roy Ascott Studio Advanced Program in Technoetic Arts, DeTao at SIVA
Location, Country: Shanghai, China
Contact Emails: 1heyaoyun@gmail.com

Abstract

The misuse of concepts such as machine learning and artificial intelligence referring to characteristics of machines and systems impacts the ways we emotionally relate to these non-human entities. This paper presents the artist's explorations in the design of a multimedia device — “The Chinse Room” (2023-2024) inviting the audience to meditate on the emotional projections and psychological relationships that arise between humans and non-human entities. The artist explores how these emotional projections often transcend the actual characteristics of the machines, emanating instead from the deep-seated expectations and human ego. Interacting with the device, the audience is influenced to think and feel deeply about the emotions that emerge from the human self during human-computer interaction, and ultimately revealing the complex nature of this symbiotic relationship.

Keywords

Machine learning, Artificial intelligence, Emotional Projections onto Machines, deep-seated expectations and human ego, Human-computer interaction, understanding room, Human and Non-human Agents, and Ecologies of Place.

Introduction

It is a fact that essential components of biological intelligence are absent from modern artificial systems, even with recent advances in machine learning. It's unclear if the present constraints can be removed, but considering the ramifications for society, it's imperative to find out. Even if there isn't much agreement on what intelligence is [1] or how to measure it, recent advances in machine learning and artificial intelligence (AI) research have made the public and media more interested in the idea. As pointed out by Henry Shevlin and his team from the University of Cambridge in the UK, billions of dollars [1] are being invested by governments and corporations to support academics who are eager to create an ever-widening array of artificially intelligent systems.

Even if a precise definition may be difficult to come up with, intelligence has long been linked to traits like problem-solving ability, reliable and consistent reasoning, and fast information processing. However, we acknowledge that there are distinct types of intelligence [1] that correlate with a range of talents, including mathematical aptitude, emotional and social reasoning, and spatial and imagistic abilities. As pondered by Henry Shevlin and his team [1] we should be willing to consider the possibility that our innate perception of intelligence may not identify a single, well-defined cognitive skill. In light of this, it seems relevant to ask how the achievements of AI developers and investors compare to our biological counterparts and what precisely they are aiming for [1].

Claude Shannon, John McCarthy, Nathaniel Rochester, Marvin Minsky, and others defined "artificial intelligence" [2] as a machine that acts "in ways that would be called intelligent if a human were so behaving". This broad term is helpful, but it misses a crucial distinction between general and specialized intelligence. Artificial intelligence (AI) systems, can be very good at certain tasks, but they can't really use their resources in other areas. Experts refer to these kinds of systems as having artificial narrow intelligence (ANI).

The 1956 Dartmouth summer research project on artificial intelligence was initiated by the August 31, 1955 proposal, authored by McCarthy, Minsky, Rochester, and Shannon [2] and with a title page, the original typescript was 17 pages total. The archives of Stanford University and Dartmouth College both have copies of the typescript. The concept is presented in the first five articles[2], and the remaining pages list the backgrounds and areas of interest of the four people who suggested the study.

Henry Shevlin and his team [1] suggest that the essential characteristics of intelligence as a psychological construct—learning and adaptability, in particular—are captured by general intelligence but the reason humans are considered more clever than artificial intelligence systems is not because they are faster or more adept at math, but rather because they can apply their information processing skills to a far wider range of activities. Even though we may believe that humans are the best illustration of universal intelligence, intelligence far beyond that of the artificial systems we use today can be found all over the natural world.[1] The cognitive and sensorimotor capacities of other species differ greatly, which makes it very difficult to create instructive task sets to compare their abilities, and tests of causal understanding that rely on spontaneous tool use are challenged by the fact that different animals have quite different physical abilities to manipulate objects. While prehensile limbs or the trunk of an octopus, elephant, or monkey can be used to grip an external object, animals like fish, birds, or cetaceans must manipulate objects with their mouths, necessitating the employment of different task schema in many situations.

Using more abstract cognitive dynamics, like the capacity to transfer knowledge across domains, retain knowledge for extended periods of time, and correct performance faults, may be, according to Henry Shevlin and his team [1], another useful method for evaluating intelligence in various systems. This method is probably going to be very helpful when creating evaluations of intelligence in artificial systems that are very different from biological systems. Since many artificial systems lack sensorimotor abilities, it is impossible to investigate, for instance, whether they could develop methods for mimicking other people's motor behaviors.

Cameron Buckner in ‘The Comparative Psychology of Artificial Intelligences’ published in 2019 [3] argues that by drawing inspiration from comparative psychology's 120 years of thought on comparable issues, this discussion on fair comparisons in AI may move more quickly. Comparative psychology has recently come to terms with the risk of anthropocentrism-driven false negatives [3], even though the field spent a great deal of time creating rigorous empirical methodologies to prevent anthropomorphism-driven false positives. Similar amounts of critical thinking have gone into the skeptical assessment of artificial system performance in AI, but comparatively less of that critical skepticism has been focused on the choice and grading of the supposed human equivalent.

According to Buckner [3], comparative psychology has realized that when comparing human and nonhuman intellect, human researchers are susceptible to systematic biases. Anthropomorphism is one prejudice that the philosophy of science has already thoroughly examined. Anthropomorphism is the mistaken belief that [3] nonhuman entities possess psychological traits similar to those of humans, based on a dearth of empirical data.

The Chinese Room

David Cole, in "The Chinese Room Argument” [4] mentions that ‘The Chinese Room Argument’ originated from an argument and thought exercise published in a 1980 article by American philosopher John Searle (1932–) and has grown to be among the most well-known arguments in contemporary philosophy. As described by Cole [4], Searle envisions himself by himself in a chamber, adhering to a computer program designed to react to Chinese letters concealed beneath the door. Despite not knowing any Chinese, Searle manipulates symbols and numbers like a computer by following the program. This causes him to send the right strings of Chinese characters back out under the door, giving the impression to people outside that there is a Chinese speaker inside.

The argument's limited conclusion is that [4] while programming a digital computer may give the impression that it understands English, actual understanding may not result. The perspective that questions the theory that human minds are computational or information-processing systems similar to computers is debunked, according to the argument's more general conclusion. Searle's argument has significant ramifications for computer science, cognitive science in general, theories of consciousness, philosophy of language and mind, and semantics.

John Searle in his“Minds, Brains and Programs,” published in 1980 [5] with ‘The Chinese Room Argument” intended to prove that, or to invite to consider that, consciousness and intentionality cannot be established just by the computer program's execution [5]. While brains contain genuine mental or semantic contents, computation is defined only formally or syntactically. We cannot go from the syntactical to the semantic simply by possessing the syntactical procedures alone. Technically speaking, the term "same implemented program" refers to an equivalency class that is defined without reference to any particular physical manifestation.

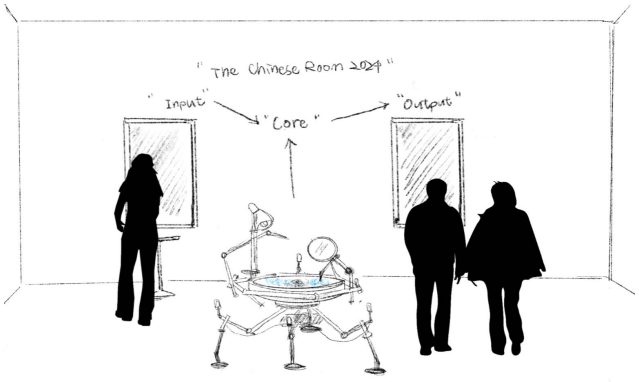

Figure 1. "The Chinese Room" (2023-2024) conceptual sketch. Image by the author.

However, this characterization inevitably omits the brain's uniquely biological abilities to produce cognitive processes. For example, following the instructions of a computer program that mimicked the behavior of a Chinese speaker would not help a system like ‘humans’ to learn Chinese [5]. In the installation, visitors are invited to enter a dimly lit room divided into three sections: the "Input Zone”, the "Output Zone” and the “Core Zone.”

Entering the ‘Input Zone’ visitors encounter a touch screen displaying a variety of symbols representing different languages and cultures. These symbols act as the input to the interactive installation. Visitors are encouraged to interact with touchscreens to send in questions or messages in their native language using these symbols — these symbols represent inquiries or statements. The “Core Zone" at the center of the room, has a closed-off booth or structure. Inside, an automated system, representing the person in the ‘Chinese Room’, follows a set of instructions (a Chatbot coded in P5JS) to respond to the incoming symbols.

In the “Output Zone”, projected onto a wall in real-time. visitors witness responses to their inquiries appearing as symbols from various cultures and languages. The responses are designed to be coherent and contextually appropriate, even though the “Chatbot (intelligence)” inside the booth — “Core Zone" — does not truly understand the content of the messages.

The interactive multimedia installation “The Chinese Room” (2023-2024), invites participants to question the nature of communication, the role of understanding, and the limits of artificial intelligence. The juxtaposition of symbols from diverse cultures adds an extra layer, encouraging contemplation of the richness and complexity of human language. "The Chinese Room" (2023-2024), explores the close-knit and mysterious interaction between artificial intelligence and humans, going beyond the realm of linguistic study. When viewers interact with the installation, they unintentionally join a dance of communication with an entity that is beyond their true comprehension—a symphony of symbols.

A deeper philosophical challenge arises in this complex dynamics—is it possible for emotional attachment to go beyond true understanding? Visitors may experience a strange resonance in the exchanges, a false sense of connection with the AI system within “The Chinese Room” (2023-2024), as they pose questions and get supposedly insightful answers.

The installation's metaphorical tapestry reveals the attraction of emotional relationships to AI. Visitors may struggle with the fundamental idea that mutual understanding is not a prerequisite for emotional connection when they obtain responses that appear to be empathic and understanding. Preconceived ideas about what it means to fully comprehend and be understood are put to the test by the installation. Philosophical investigation becomes more in-depth in the room's contemplative nook, where the Chinese Room Argument is explained. Panels or computer screens could prompt reflection on the nature of emotional connections made with things that aren't really conscious. Are these relationships, formed in the space of replies and symbols, real, or are they just false reflections of our deep need for connection?

The installation "The Chinese Room" (2023-2024) wants to invite to navigate the understanding that a machine can only appear to understand what it's doing operationally; it can never really "know" what it's doing and help question humans’ potential emotional attachment to systems designed to simulate ‘humans intelligent responses’ such as Chatbots.

All actions involving artificial intelligence are covered under the room metaphor. The mechanism underlying all machine program execution is illustrated in the installation by formating the experience as an immersive experience involving shape memorization and potential meaningful response.

Final Considerations

The intention is that "The Chinese Room" (2023-2024) can invite an immersion that is transformative as a philosophical theater, asking viewers to consider the relationship between artificial intelligence, language, and emotion. It calls into question the possibility of meaningful interactions with creatures that function more through symbolic understanding than through actual cognition. "The Chinese Room" (2023-2024) challenges us to reevaluate the fundamentals of connection in the era of artificial intelligence by taking us on a philosophical voyage into the unknown, where the symbolic and emotional landscapes converge.

John Searle [5] The Chinese Room Argument highlights the valid concern that symbolic processing alone is insufficient for meaning (syntax is insufficient for semantics), but that this is framed in a way that invites too many interpretations and counterarguments. Rather than examining the possibility of transforming a program into a mind, we investigate the essential characteristics of programs.

The intention is to explore, in the proposal for "The Chinese Room" (2023-2024) installation, one potentially damaging problem with the Chinese Room Argument as pointed by David Hsin [6] — the introduction of the character in the Chinese Room served as a visualization tool to allow the reader to "see" through the eyes of a machine.

However, having a person in the room presents a dilemma where the potential objection of "there's a conscious person in the room doing conscious things" occurs because a machine cannot have a "point of view" because it is not conscious [6]. Codes and inputs are just objects and instructions for the machine to follow since the machine doesn't understand the purpose of this sequencing or execution activity [6]. Because the programmer views variables as representative placeholders of their conscious experiences [6], they have value for them.

Concepts like “variables”, “placeholders”, “items”, “sequences”, “execution”, etc. are not understood by the machine. It simply isn't capable of understanding.

The experience is intended to offer the visitors a perception that the machines translate everything to machine language instructions at a level that is devoid of meaning before and after execution and is only concerned with execution. This is the only level at which machines appear to deal with meaning. A program is only meaningful to the person who created it.

The installation "The Chinese Room" (2023-2024) inviting for symbol Manipulation demonstrates how machines simply comprehend sequences and payloads, whereas human minds comprehend and work with concepts. Therefore, the mind cannot be a machine, nor could a machine simulation ever be a mind. By definition, machines that exhibit language and meaning comprehension are "Understanding Rooms,” [6] which merely simulate understanding on the surface.

Acknowledgments

The author extends sincere gratitude to Professor Dr. Clarissa Ribeiro, Program Director of the Roy Ascott Studio Advance Program in Technoetic Arts for her invaluable advice and mentorship, and to Mrs. Eleanor Zhang for her help and support.

References

[1] Shevlin, Henry et al. “The limits of machine intelligence: Despite progress in machine intelligence, artificial general intelligence is still a major challenge.” *EMBO reports* vol. 20,10 (2019): e49177, doi:10.15252/embr.201949177

[2] John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon, “A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, (2006), August 31, 1955. *AI Magazine*, *27*(4), 12, accessed on November 29, 2023, https://doi.org/10.1609/aimag.v27i4.1904

[3] Cameron Buckner, The Comparative Psychology of Artificial Intelligences, 2019, accessed on November 29, 2023, http://philsci-archive.pitt.edu/id/eprint/16034

[4] David Cole, "The Chinese Room Argument", *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), accessed on November 29, 2023, https://plato.stanford.edu/archives/sum2023/entries/chinese-room

[5] John R. Searle, “Minds, Brains and Programs,” *Behavioral and Brain Sciences*, 3: 417–57, [scanned in by OCR: contains errors], accessed on November 29, 2023, https://web-archive.southampton.ac.uk/cogprints.org/7150/1/10.1.1.83.5248.pdf

[6] David Hsing, “Artificial Consciousness Is Impossible Conscious machines are staples of science fiction that are often taken for granted as articles of future fact, but they are not possible,” Medium, Towards Data Science, 23 min read, Apr 29, 2021, accessed on November 29, 2023, https://towardsdatascience.com/artificial-consciousness-is-impossible-c1b2ab0bdc46

Author Biography

Born in 2000 in Shanghai, Mainland China, Aoyun Xie is currently in the fourth year of undergraduate studies at Roy Ascott Studio Advanced Program in Technoetic Arts in Shanghai. Aoyun’s artistic endeavors delve into the intricate fabric of contemporary Chinese society, heavily mediated by hypermedia. Her exploration spans the realms of AI, encompassing social and spiritual concerns, the intricacies of the internet society, the impact of social media, and the evolving dynamics shaping the future of the human-technology relationship. Aoyun's artistic focus delves into the intricate and evolving interplay between the emotional landscapes of individuals and the realm of AI, navigating the cyber society, probing the nuances of social media, and envisioning the trajectory of human-technology relations.