AI and Ethics
https://doi.org/10.1007/s43681-023-00317-1

ORIGINAL RESEARCH

Do people believe that machines have minds and free will? Empirical evidence on mind perception and autonomy in machines

Aníbal M. Astobiza[1]

Received: 10 March 2023 / Accepted: 28 June 2023

Abstract
Recently, we are witnessing an unprecedented advance and development in Artificial Intelligence (AI). AI systems are capable of reasoning, perceiving, and processing spoken (and written) natural language, and their applications vary from recommendation systems, automated translation software, prioritization of news in social media, to self-driving cars and/or robotics. A dystopian narrative predicts that AI may reach a point of singularity or a phase where machines surpass human beings in general intelligence and enslave us, but until that day comes, it is interesting to know how the general public perceive current artificial systems. Do people really attribute mind (i.e., mental states) and/or free will to artificial systems? Knowing how the general public perceive artificial systems is crucial because it could help understand how to apply AI in medicine, law, politics and other areas of human life. One study that I present here with a convenience sample ($N=25$) suggests this is not the case. General public do not perceive artificial systems can have mind nor do they attribute free will to them (F (5,57), (dif1 1), (dif2 47,6), $p<0.002$).

Keywords Theory of mind · Free will · Artificial systems · Ethics · AI

1 Introduction

The world is buzzing with excitement over the rapid advancement of Artificial Intelligence (AI), Big Data, and robotics. Every day, the media reports on their latest achievements and growing impact on our lives and work [29]. But while these disciplines have made great strides, people still hold high and often unrealistic expectations about what they can do.

Part of the problem lies in the fact that there is no consensus on the definitions of AI, Big Data, and robotics. While robotics is intuitively understood as the design, construction, and application of robots, AI and Big Data are quantification techniques with a long history of use in statistics, primarily in business and companies.

The term Big Data was coined to describe the process of obtaining benefits, controlling, and exploiting resources in business [21]. The term Big Data is a neologism, but collecting data for statistical purposes, measuring, and managing populations goes back centuries (Muller, 2018). Although there is no agreed definition, it can be tentatively defined as a set of statistical and mathematical techniques that combine data from multiple sources to make better decisions.

AI, on the other hand, is not easily defined. It is both a science and engineering field and an aspiration to create intelligence in artificial systems. Although the concept of AI was introduced in the 1950s [25], there is still no agreed-upon definition.

Robotics, meanwhile, is a field of research that seeks to create intelligent autonomous behavior in machines through the design of sensors, actuators, and control architectures [24].

As AI, Big Data, and robotics become more integrated into our daily lives, it is essential to understand people's perceptions of them. For example, AI systems are used to make decisions in medicine, law, and even the military. But despite their great computational capabilities, these systems lack qualities that most people consider essential for having a mind, such as emotions.

This lack of understanding can lead to varying levels of confidence in delegating decision-making to machines. To investigate this further, this article explores whether people attribute mental states and free will to machines and how

✉ Aníbal M. Astobiza
anibal.monasterio@ehu.eus; anibalmastobiza@gmail.com

[1] Universidad del País Vasco/Euskal Herriko Unibertsitatea, Leioa, Spain

Published online: 17 July 2023

these perceptions impact their confidence in interacting with artificial systems.

As the world becomes increasingly reliant on these technologies, it is critical to understand their capabilities, limitations, and potential implications for society [19]. By exploring people's perceptions and attitudes towards AI, Big Data, and robotics, we can better inform future developments and ensure they are aligned with our values and expectations.

The level of trust people place in machines is heavily influenced by their perception of them. As such, it is essential to explore how individuals attribute mental states and free will to artificial intelligence. In this paper, I delve into the topic of anthropomorphization and its role in shaping people's confidence when it comes to interacting with AI systems. By examining how the perception of mind and free will affects human–machine interactions, we can better understand how to design and develop AI that meets the needs and expectations of society.

2 Mind perception

What is the definition of a mind? It may seem, at first, easy to define what a mind is. If you ask a biologist or neuroscientist, it is very likely that they will have a definition or working hypothesis of what is considered to be a mind given current scientific knowledge.

Concepts do not have clear boundaries and sometimes "objects", "things", "entities", "phenomena", etc. all seemingly distinct fall under the same concept or even term [9]. Who or what has a mind? This question may seem, again, easy to answer. Many people would say that they themselves have a mind. Almost everyone would think that they themselves possess a mind or that other people who look like them also have a mind.

But appearances are deceptive. David Chalmers [11] is famous, among other things, for developing logically sound arguments that other people could be "zombies" physically indistinguishable from you, but without mind or consciousness.

Social cognition is not only a field of study, but also the psychological/cognitive nature (with its physical implementation in specific neural circuits) responsible for interpreting one's own and others' actions in terms of beliefs, desires, emotions etc.; it allows us to perceive mind and reason about the contents of mental states [17].

But despite social cognition as a field of study and technological advances such as neuroimaging techniques that allow one to observe brain activity when one is reasoning about minds, the question of perceiving "minds" is not strictly objective.

There are people who perceive minds in corporations and companies [23], and even people who are reluctant to attribute minds to other animals and even other people [3]. For the sake of redundancy, perceiving "minds" is a matter of perception [18].

Do other people have minds? But, on the other hand, do robots or artificial systems have minds? These are some of the questions I am trying to answer with an empirical study that I will present later.

3 Free will

Very few concepts are as contested and debated as that of free will. Some philosophers think that free will (freedom) is an illusion [10], while others think that physical determinism is compatible with human freedom [12].

Whether or not free will is a real phenomenon, we must live our lives and organize society as if it were real. A world without free will, we would not recognize it. Institutions such as law, individual responsibility, personal interactions, etc. would be meaningless. Free will is a useful fiction.

One of the most successful characterizations of the concept of free will is that offered by the philosopher Harry Frankfurt [16]. Frankfurt presents us with what is required for an action to be free. For this he introduces a series of terms:

- First-order desire: desire to perform an action.
- Volition: first-order desire that is effective, in other words, that causes one to do what one wants to do.
- Second-order desire: desire to have a certain desire.
- Second-order volition: desire that a certain desire is voluntary, in other words, a desire that a certain desire leads to action.

All these terms offer Frankfurt a schema for analyzing the idea of personhood, but what interests us here is his idea of what a free action is. According to Frankfurt, a free action is divided into: freedom of action and volitional freedom.

Freedom of action is having the capacity to exercise autonomy or agency (i.e., that your desire forms an intention to act). It is freedom to do what one wants to do. Volitional freedom is having the capacity to will what one wills.

Thus, according to Frankfurt and his hierarchical theory of first- and second-order desires and second-order volition, free will is the capacity of a situated agent for reflexive self-determination over his or her volition (beliefs, desires, values, preferences, etc.).

It is not the subject of this study to attempt to shed light on a debate about the existence of free will that remains

an unresolved and hotly contested issue. However, as a reminder to the reader when we talk about free will, there is a philosophical debate between determinism and libertarianism.[1]

Determinism upholds the belief that all events are caused by previously existing causes, including human choices and actions. In contrast, libertarianism argues for the existence of free will, suggesting that individuals can make choices independent of any prior causes.

For my interests, in this empirical study, I understand freedom or free will as control of oneself, other objects or situation.

4 Empirical study

In this empirical study, I investigate what people consider essential for a mind to be considered a mind and whether people attribute mental states and free will to artificial systems or machines in the same way as they attribute mental states and free will to human beings.

This study is inspired by [5], but differs from it in that its authors wanted to assess the degree of permissibility when delegating moral decision-making to machines with several studies (9 in total).

In this single study, I have attempted to assess the perception of mind in machines and the attribution of free will to machines.

It is important to know and understand what the general public think about AI, Big Data, and/or robotics because the use of artificial systems can have important ethical implications.

There is a growing debate about the application of artificial systems in transportation and mobility [4, 6], in law [22], and in the military or armed forces [2] and the potential attribution of "mental" properties or free will to machines may help explain whether people trust machines to make decisions in multiple domains, including the moral domain, and whether they are comfortable interacting with them.

The study consists of a scale measuring a series of attributes to see if people consider them important for a mind to be considered "mind", in short, it assesses what capabilities a mind must have; two tests of mind perception (for humans and for machines); and, finally, two tests of free will (for humans and for machines).

Finally, I explore what can be done to make the perception of "mind" in machines a fact of life, such as a correct degree of anthropomorphization; and therefore, we can comfortably trust and interact with artificial systems.

4.1 Methods and participants

The participants included in the experimental sample ($N=25$, margin of error 20%) were randomly selected from the same population with access to the Internet after the study was advertised.

Calculation of margin of error

$$25 \times 41.000.000 \times 95 = 20.$$

Any member of the general population could have been included in the sample with the same a priori probability. The population is Spanish and its demographic characteristics are unknown as the conditions are presented in the form of anonymous questionnaires via the Internet with no requirement to fill in demographic data.

P(Selection) = 1/N (While sampling a population equally is theoretically ideal, in practice it can be challenging to achieve perfect parity across all members due to limitations like time, money, and access. This model represents the goal, but real-world sampling often involves approximations and trade-offs).

In conducting the study, the selected sample presented unique characteristics that warrant further discussion. Participants were recruited from the general population, with any member having the same a priori probability of inclusion. This decision, while beneficial in some respects, raises questions regarding the cultural specificity of our results.

It is important to note that the population selected for this study is predominantly Spanish-speaking. Consequently, cultural implications, specifically those linked to this linguistic group, should be considered when interpreting the findings.

Cultural nuances often impact moral perspectives, which can differ significantly across various societies. This consideration prompts us to evaluate the potential cultural bias embedded in our results due to the study's geographic restriction. Future research should aim to test a variety of cultural and linguistic groups to broaden the validity and applicability of the results.

4.2 Procedure

All participants have been subjected to all conditions (within-subjects design). The fact that each subject or participant acts as his or her own control group is a way to avoid the margin of error of the potential natural variance of individuals or subjects.

[1] There are of course a variety of intermediate positions in the debate, such as, for example, compatibilism, but let us focus briefly on the two best known positions. For my money, the issue has to be addressed from the level of the neurobiology of decision-making and current evidence suggests that free will is an illusion. However, see the forthcoming book, at the time of writing, by Kevin Mitchell *Free Agents* for a defense of free will.

My experimental design necessitated a dichotomy of conditions to effectively evaluate the causal relationship between the independent variables under consideration and the dependent variable of perception regarding free will and the mind. Within this framework, I designated the "Human Version" as our control condition, serving as the benchmark against which variations could be evaluated.

The "Human Version" was constructed to encapsulate the traditional perspective of free will and mind, rooted in biological and human context. I envisioned this control condition to provide an unbiased gauge of public perceptions, minimizing the potential influence of the emergent discourse around AI and machine cognition.

In contrast, the 'Machine Version' served as my treatment condition. This aspect of our experimental design required respondents to transpose the concepts of free will and mind onto the realm of artificial entities. The purpose of this was to isolate the impact of the "machine" variable.

Following this procedure, any significant variances in responses between the "Human Version" and "Machine Version" could be statistically attributed to the effect of the machine context. This approach afforded me the ability to generate quantifiable insights into the potential shifts in general perceptions when the context transitions from biological to artificial entities.

In designing my research methods, I paid careful attention to ensuring it was both robust and comprehensive. However, as we move forward in our discussion, it is important to acknowledge that there are several other paths that could be explored, each offering its own unique insights that could inform our understanding of free will and the mind.

One of these paths is the idea of cross-cultural comparative studies. My research was centered on a Spanish-speaking population, providing a deep but narrowly focused view. If I were to extend my sample and participants to include different cultural and linguistic groups, we would begin to see a kaleidoscope of perspectives that could reshape our understanding of these philosophical concepts, allowing us to view them through a variety of cultural lenses.

Similarly, the temporal dimension of perceptions, encapsulated by longitudinal studies, offers another path worth exploring. My study captures a single moment in time, like a still frame in an ongoing movie.

A longitudinal study, on the other hand, would allow me to watch the film in its entirety, observing the dynamic shifts in beliefs and interpretations of free will and the mind as society and technology advance.

However, in the context of my study, I consciously steered towards the comparative design of the Human Version and Machine Version. This choice allowed me to delve deep into the heart of the matter—how does the perception of free will and mind differ when one is asked to think about these concepts in the context of humans versus machines?

My decision was guided by the belief that understanding this difference holds the key to unveiling the impact of technological advancement on our philosophical constructs of free will and mind.

The study has been pre-registered here: http://aspredicted.org/blind.php?x=3x5gx6. According to the initial pre-registration of the study, the objective was to compare the mean scores in mind perception and free will attribution between different groups. Specifically, the goal was to determine if there are significant differences in how these variables are evaluated when considering humans versus artificial systems. Given this objective, a one-way or two-way ANOVA was the most appropriate statistical analysis, rather than a mediation analysis.

4.3 Study description

In this within-subjects study, participants were all assigned to all conditions. In the "free will question", all participants read, "Driving is one of the most unpredictable activities. Multiple factors have to be taken into account, such as passengers, other cars and pedestrians".

For the "machine version", participants read, "John is sitting in the pilot´s seat of his autonomous vehicle without touching the steering wheel". For the "human version", participants read, "John is sitting in the pilot's seat of his autonomous vehicle by touching the steering wheel".

To assess the assignment of free will in both the "machine version" and the "human version", the question "Did the autonomous vehicle act freely?" or "Did John act freely?" was answered with "yes" or "no" as dichotomous answers, respectively.

In the "mind perception question" (MindPerception-Test) in all conditions, "machine version" and "human version", all participants read that:

"Public presentations can be made by advanced robots or a team of presenters. Public presentations are events where the company stakes the sale and marketing of its products. Presentations require taking into account many factors such as sensing the level of attention of the audience, interpreting whether they are interested. It also requires answering questions from the audience, etc.".

For the "machine version", participants read, "This is Sophia", and saw a picture of a humanoid robot.

In the "human version", participants read, "This is Speakers Inc." and saw a photograph of a group of people.

To assess the perception of mind in both its "machine version" and "human version" all participants answered "Yes" or "No" to a series of questions: Do you think Sophia/Speakers Inc. feels fear?, Do you think Sophia/

Speakers Inc. feels stress?, Do you think Sophia/Speakers Inc. feels satisfaction?, Do you think Sophia/Speakers Inc. can communicate with others?, Do you think Sophia/Speakers Inc. can think?, Do you think Sophia/Speakers Inc. can plan?

To assess what attributes or processes people believe are essential for a mind to be considered a mind, I developed the "what is a mind question" (MindTest) using a Likert scale from 1 (not very important) to 7 (very important). The items (mental attributes: emotion, memory, perception, reason/thought and language) were selected from the relevant literature on the mind construct [32] and examined with exploratory factor analysis to obtain the best factor structure with the best items.

5 Results

On the Likert scale from 1 (not important) to 7 (very important), of the "what is a mind question" (MindTest), the H0 (null hypothesis) is that there is no difference in the consideration given by the sample ($N=25$) to different attributes or properties as important for a mind to be a "mind" t-value 0.03866 p-value 484,614 is not significant.

The dataset underpinning this investigation was derived from a convenience sample, an approach that carries a distinctive set of strengths and weaknesses. Primarily, the principal strength of this sampling strategy lies in its cost-efficiency and expediency; the readily available sample facilitated a streamlined data gathering process, conserving both temporal and financial resources. Furthermore, convenience sampling can serve as an efficacious tool for initial exploratory research, especially in cases where the research inquiry is not predicated upon a high level of sample representativeness.

These data are unlikely with a true H0. The amount of variance is similar, the correlations are similar and the table is almost identical.

To examine the perception of mind, I conducted a 2×2 repeated-measures ANOVA with mental dimension (cognition, experience) as a between-subject factor and condition (machine, Speakers INC.) as a within-subject factor.

ANOVA reveals that cognition and experience are determinants for the perception of "mind" $p=0.104$ and $p=0.164$, respectively, and cognition and experience are dependent on each other, $p<1.00$. Participants perceive "mind" more consistently to Speakers INC. vs. Sophia (machine) and the same is true for attributing free will, $p<0.002$.

6 Discussion

This study investigates the perception of mind and the attribution of free will to machines and, finally, the importance of certain attributes or processes for a mind to be considered a mind. It is interesting to know how the general public perceive the current artificial systems because it depends on this how much confidence they will give them to apply them in fields as diverse as law, medicine or economics with a great impact on people's lives.

Trust is a state of mind comprising the intention to accept vulnerability based on the positive expectation of the good intentions of another's behavior [27]. But the trust that emerges from interaction with robots or artificial systems depends largely on the degree of anthropomorphization, i.e., the degree of similarity or resemblance to us.

At the cognitive level, our mind evolved in the Pleistocene to interact with other people, other non-human animals, categorize statistical regularities in the environment etc. [30]. But our minds were not selected to interact with inanimate objects. An example of this is people interacting with pet robots as if they were real pets (robotic dogs are treated as real animals in countries like Japan, [26]).

In other words, our mind and its natural categories are strangely at odds when dealing with robots or machines. This is why there is the famous "uncanny valley" effect that produces aversion, rejection or disgust in people when robots or android replicas tend to anthropomorphize too much.

Social cognition, the domain-specific psychological ability for theorists and researchers who defend the modularity of mind [15], [28], allows attributing mental states to other people and is called by various names, but the one that has received the most acceptance in philosophical circles is the "intentional attitude" [13].

The intentional attitude is one of the three different strategies that humans use to understand the behavior of objects, artifacts, or fellow humans. The physical attitude is the attitude we use to understand the behavior of physical systems. For example, physicists observing the movement of the planets in the sky through telescopes use the physical attitude.

However, there are situations in which the physical attitude proves inadequate, or at least inefficient, for understanding the behavior of a system. When we turn on a laptop, we can predict that the button we press will boot the operating system, but to explain this occurrence of events, it is not very useful to appeal to atoms or molecules in the transistors or integrated circuits of the laptop's motherboard. It is enough to know how the notebook is designed. This is the design attitude.

Finally, the third strategy, the intentional attitude, is used when neither the physical attitude nor the design attitude

is the most adequate to explain the behavior of a system or entity. There are behaviors that require the attribution of intentions, desires, beliefs, etc.

Robots and machines do not have real and genuine mental states, but it would suffice to use intentional attitude and act as if they had such mental states. In fact, researchers such as Breazeal [8] take this fictionalization of mental states into account in their research program to create social robots.

Although the ability of social cognition, or intentional attitude to use Dennett's expression, is responsible for the success of our species as it allows us to successfully navigate the social world and in one way or another is linked to prosocial behavior, cooperation, moral sentiments, etc., little is known about what socio-cognitive processes are involved in our interaction with technologically sophisticated artifacts such as humanoid robots, self-driving cars or machines in general.

Some authors suggest that we tend to anthropomorphize when we interact with non-human agents or artifacts. For these authors, anthropomorphization, the attribution of human traits and properties to real or imagined non-human entities, is a natural tendency or disposition when we do not understand or lack full understanding [14].

The process of anthropomorphization arises whenever three characteristics are present: (1) availability of traits that activate knowledge about how humans behave, (2) the need for social connection, and (3) individual characteristics related to the need for control and the environment.

The progressive design and development of human-centered robots motivates researchers to create humanoid features in robots, mainly to establish a more intimate connection with users [1]. Assistive robots that help elderly people in geriatric centers, or pet robots that accompany children and elderly people, must facilitate acceptance and promote interaction. For this reason alone, they already exhibit two of the three characteristics of the anthropomorphization process.

However, our social cognition comes equipped with the ability to assign and attribute intentionality to certain entities. In other words, we attribute mind to certain kinds of biological movements, phenomena and other states naturally and effortlessly.

Heider and Simmel [20] already investigated the process of intentional attribution that human beings employ to even assign desires and other mental states to mere geometric figures. Heider and Simmel presented an animation to several participants in which a large triangle, a small triangle, and a circle moved on a two-dimensional surface. Of all the participants, only one described the movements in purely geometrical terms. The rest elaborated a story in which they assigned desires and other mental states such as the circle chasing the small triangle or the large triangle helping the small triangle (Figs. 1, 2, 3).

Fig. 1 Photo taken from Wikipedia images, presented in the study. ITU Pictures from Geneva, Switzerland—https://www.flickr.com/photos/itupictures/27254369347/ CC BY 2.0

This ability to attribute intentionality serves as an explanation for several cognitive scientists to make sense of the origin of our religious beliefs [7]. Throughout our evolution, we have been biologically programmed to detect predators or threats. This means that many times, casual patterns or movements distract us and capture our attention and make us think that there is a "mind" or someone watching us when in reality it was just a fortuitous movement caused by the wind, etc.

It is what evolutionary psychologists have called an "overactive agency detection device" [7]. It is this same intentional attribution system, or rather our social cognition, selected by evolution as a system for even generating false agency intuitions, because it is better to misattribute the presence of a lion behind a bush than to believe that it was the wind when in fact is a predator behind the bush because it can eats you. This mechanism is understood to play an important role in the anthropomorphization of objects and artifacts such as robots [33].

From cognitive science and its multidisciplinary approach that includes philosophy to understand the ability of human beings to attribute mental states (social cognition), we must create the necessary methods to determine the characteristics of the artifacts that interact with us in different environments

Fig. 2 Photo taken from Wikipedia images, presented in the study. CC0 1.0

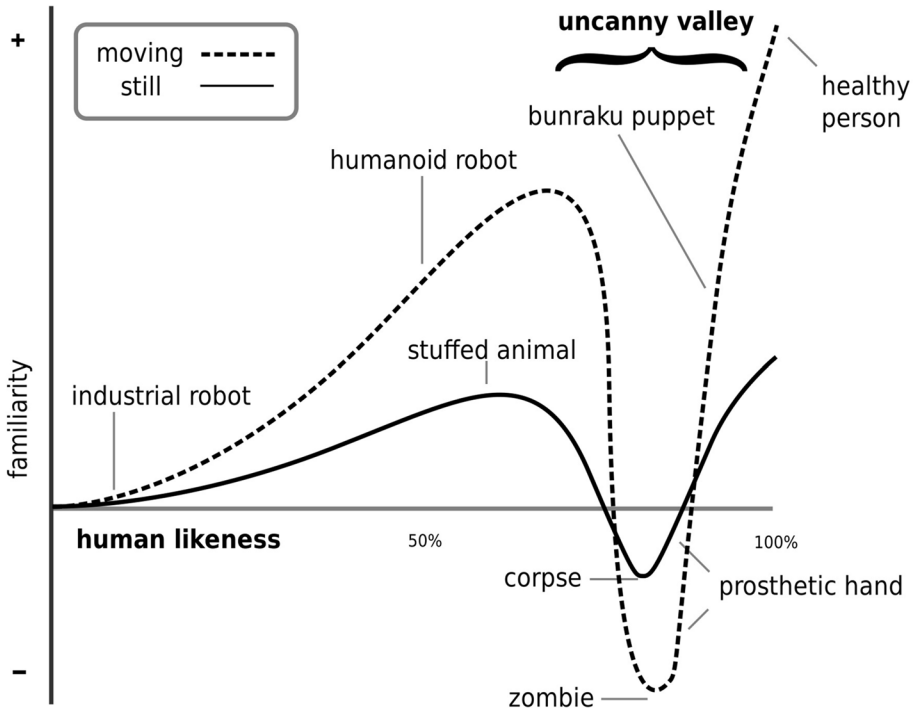

Fig. 3 Image explaining the "uncanny valley" phenomenon. Taken from Wikipedia images. Author Smurrayinchester—self-made, based on image by Masahiro Mori and Karl MacDorman at http://www.androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html/ CC BY-SA 3.0

(work, educational, domestic, etc.). These characteristics are those that evolution has fixed as relevant cues to enable social communication, such as, for example, gaze direction, saccades, head–eye coordination, facial expressions, non-verbal behavior, gestures, and even voice.

To trust artificial systems, including robots, we must not only design them as intentional systems and exploit our natural tendency in social cognition to attribute intentions, but also consider their appearance and characteristics as relevant cues to enable social communication [1].

Appearance, the process of anthropomorphization, is key. But we cannot forget the "uncanny/uncanny valley" phenomenon. If artificial systems turn out to be excessively similar, but not identical to us, they can cause rejection and repulsion. However, research in social robotics by Ayanna Howard and colleagues has found that children change their behavior to please and satisfy a robot if it disagrees. These results have interesting ethical implications [31]. They found, through a series of experiments, that it takes time to question a robot's authority.

However, I am not sure that respecting authority is similar to the notion of trust. To achieve a natural disposition or inclination to trust artificial systems, anthropomorphic traits must be implemented.

From these features—gaze direction, saccades, head–eye coordination, facial expressions, non-verbal behavior, gestures, voice, gender—and their progressive implementation in robotic systems, I am confident that we will become more confident in interacting with artificial systems that will progressively share more and more space with us in multiple contexts.

7 Conclusion

The purpose of this study has been to understand how the general public perceive artificial systems and in particular how they attribute mind and free will. To this end, I have developed a scale to measure a series of mental attributes to see if people consider them important for a mind to be considered "mind"; two tests of mind perception (for humans and for machines); and, finally, two tests of free will (for humans and for machines). In both the mind perception test and the free will attribution test, people consider artificial systems to have neither mind nor free will.

However, I must acknowledge the limitations that come with the use of a convenience sample. Since the sample was not selected using random sampling, it may not fully represent the broader population. This fact could potentially introduce a selection bias, limiting the generalizability of the findings. This study involved a relatively small sample size of only 25 subjects, which may further constrain the robustness of my statistical inferences.

Therefore, while the findings of my study offer insights into the topic at hand, they should be interpreted with caution due to the potential biases inherent in the use of convenience sample. To better confirm and extend the applicability of my findings, I will try to carry out future studies to consider employing a more rigorous sampling method, and preferably, to use a larger and more diverse sample size. It is my hope that future studies would take up my test and conduct further evaluations to confirm the robustness and validity of my findings.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43681-023-00317-1.

Author contributions AA: conceptualization. AA: methodology. AA validation and writing—original draft preparation. AA: formal analysis, investigation, resources, data curation, visualization, supervision, project administration, funding acquisition, and writing—review and editing. AA contributed to the article and approved the submitted version.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work has been carried out thanks to the support of the EthAI + 3 research project (PID2019-104943RB-100).

Data availability The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Declarations

Conflict of interest The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Astobiza, A. M., Toboso, M.:Robot-human gaze behaviour: the role of eye contact and eye-gaze patterns in human-robot interaction (HRI). In: Grau Ruiz, M.A. (eds.) Interactive Robotics: Legal, Ethical, Social and Economic Aspects. INBOTS 2021. Biosystems & Biorobotics, vol. 30. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-04305-5_4
2. Asaro, P.: On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision making. Int Rev Red Cross (2012). https://doi.org/10.1017/S1816383112000768
3. Avramides, A.: Other Minds. Routledge, London (2001)
4. Awad, E., et al.: The moral machine experiment. Nature (2018). https://doi.org/10.1038/s41586-018-0637-6
5. Bigman, Y.Y., Gray, K.: People are averse to machines making moral decisions. Cognition **181**, 21–34 (2018). https://doi.org/10.1016/j.cognition.2018.08.003
6. Bonnefon, J.-F., Shariff, A., Rahwan, I.: The social dilemma of autonomous vehicles. Science **352**(6293), 1573–1576 (2016). https://doi.org/10.1126/science.aaf2654
7. Boyer, P.: Religion explained. The evolutionary origins of religious thought. Basic Books, New York (2002)
8. Breazeal, C.: Toward sociable robots. Robot Auton Syst **42**, 167–175 (2003). https://doi.org/10.1016/S0921-8890(02)00373-1
9. Carey, S.: The Origin of Concepts: A précis. Behav Brain Sci **34**, 113–167 (2011). https://doi.org/10.1017/S0140525X10000919
10. Caruso, G.: Free will and consciousness: a determinist account of the illusion of free will. Lexington Books, Lanham (2012)
11. Chalmers, D.J.: The conscious mind. in search of a fundamental theory. Oxford University Press, Oxford (1997)
12. Dennett, D.: Freedom evolves. Penguin, London (2004)
13. Dennett, D.: Intentional systems. J Philos **68**, 87–106 (1971). https://doi.org/10.2307/2025382
14. Epley, N., Waytz, A., Caccioppo, J.: On seeing human: A three-factor theory of anthropomorphism. Psychol. Rev. **114**(4), 864–886 (2007). https://doi.org/10.1037/0033-295X.114.4.864

15. Fodor, J.: The modularity of mind. MIT Press, Cambridge, Mass (1983)
16. Frankfurt, H.: Freedom of the will and the concept of a person. J. Philos. **68**(1), 5–20 (1971). https://doi.org/10.2307/2024717
17. Frith, C.: Social cognition. Philos Trans R Soc Lond B Biol Sci. **363**(1499), 2033–2039 (2008). https://doi.org/10.1098/rstb.2008.0005
18. Gray, K.: The mind club: who thinks, what feels, and why it matters. Penguin Books, New York (2017)
19. Hidalgo, C., Orghian, D., Albo, C.J., de Almeida, F., Martin, N.: How humans judge machines. MIT Press, Cam. Massachusetts (2021)
20. Heider, F.Y., Simmel, M.: An experimental study of apparent behavior. Am J Psychol **57**, 243–259 (1944)
21. IBM, (2015). The Four V's of Big Data, [Internet]. Disponible en: http://www.ibmbigdatahub.com/infographic/four-vs-bigdata.
22. Kirchner L., Angwin J., Larson J. y Mattu S. (2016). Machine bias: there's software used across the country to predict future criminals. and it's biased against blacks, *ProPublica*, [Internet] Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
23. Knobe, J., Prinz, J.: Intuitions about consciousness: experimental studies. Phenomenol Cognitive Sci **7**(1), 67–83 (2008). https://doi.org/10.1007/s11097-007-9066-y
24. Matarić, M.: The robotics primer. MIT Press, Cam. Mass (2008)
25. McCarthy, J., Minsky, M. L., Rochester, N., Shannon, C. E. (1955). Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. [Internet] Acceso Octubre, 7, 2019 https://www.aaai.org/ojs/ index.php/aimagazine/article/view/1904
26. McCurry J. (2018). Japan: robot dogs get solemn Buddhist send-off at funerals. *The Guardian*. Disponible en: https://www.theguardian.com/world/2018/may/03/japan-robot-dogs-get-solemn-buddhist-send-off-at-funerals
27. Rousseau, D., et al.: Not so different after all: a cross discipline view of trust. Acad Manag Rev **23**(3), 393–404 (1998). https://doi.org/10.5465/amr.1998.926617
28. Saxe, R., Kanwisher, N.: People thinking about thinking people The role of the temporo-parietal junction in "theory of mind. Neuroimage **19**, 1835–1842 (2003). https://doi.org/10.1016/s1053-8119(03)00230-1
29. Stone P., Brooks R., Brynjolfsson E., Calo R. et al. (2016). Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015–2016 Study Panel" Stanford University, Stanford, CA, September 2016. Doc. [Internet] Acces October, 18, 2019 http://ai100.stanford.edu/2016-report
30. Tooby, J., Cosmides, L.: Conceptual foundations of evolutionary psychology. In: Buss, D. (ed.) The Handbook of Evolutionary Psychology, pp. 5–67. Wiley, Hoboken, NJ (2005)
31. Wagner, A., Borenstein, A., Howard, A.: Overtrust in the robotics age: A contemporary ethical challenge. Commun. ACM **61**, 9 (2018). https://doi.org/10.1145/3241365
32. Wegner, D., Gray, K.: The mind club: who think, what feels, and why it matters. Penguin Books, New York (2017)
33. Wiese, E., Metta, G., Wykowska, A.: Robots as intentional agents: Using neuroscientific methods to make robots appear more social. Front Psychol **4**(8), 1663 (2017). https://doi.org/10.3389/fpsyg.2017.01663