Position article / *Artículo de posición*

*Studies in Psychology: Estudios de Psicología*
2024, Vol. 45(1) 145–165

DOI: 10.1177/02109395241241380
journals.sagepub.com/home/stp

# People and machines in communication / *Personas y máquinas en comunicación*

**Alessandra Cecilia Jacomuzzi and Brigitta Pia Alioto**

**Abstract**

It was 1937 when Alan Turing published his most famous article in the *Proceedings of the London Mathematical Society*. In this work, for the first time, he presented an idealized description of the mathematical inner workings of such universal machines as would eventually become instantiated in what we now call computers. This extraordinary step in the history of ideas provided a theoretical boon that later on would prove important for the development of the computational theory of mind. According to this theory, the cognitive processes of human beings could be likened to the algorithms that apply to the representations of the external world. Based on this hypothesis, since the middle of the last century, attempts have been made to replicate the functioning of the human mind using a computer. The aim of this article is to present the differences that still exist between humans and intelligent machines. From ELIZA onwards, we have seen several developments including ALICE, up to XiaoIce. Despite the evolution of intelligent machines and the creation of social chatbots in particular, artificial intelligence still has shortcomings that set it apart from man: the way it knows, the lack of consciousness and of agency.

**Keywords**

Artificial Intelligence; social chatbots; agency; enactive approach; consciousness

**Resumen**

Corría el año 1937 cuando Alan Turing publicó su artículo más famoso en las *Actas de la Sociedad Matemática de Londres*. En este trabajo presentaba por primera vez una descripción idealizada del funcionamiento matemático interno de máquinas universales como las que acabarían instanciándose en lo que hoy conocemos como ordenadores. Este extraordinario paso en la historia de las ideas

Ca' Foscari University of Venice

**Spanish translation provided by the Editors / *Traducción al español aportada por los Editores***

**Corresponding author / *Autor/a para correspondencia*:**
Alessandra Cecilia Jacomuzzi, Department of Philosophy and Cultural Heritage, Ca' Foscari University of Venice, Dorsoduro 3484/D, 30123 Venice, Italy.
Email: alessandra.jacomuzzi@unive.it

proporcionó un aporte teórico que más tarde resultaría importante para el desarrollo de la teoría computacional de la mente. Según esta teoría, los procesos cognitivos de los seres humanos podrían asimilarse a los algoritmos que se aplican en las representaciones del mundo externo. Bajo esta hipótesis, desde mediados del siglo pasado se ha intentado replicar el funcionamiento de la mente humana a través del ordenador. El objetivo de este artículo es presentar las diferencias que siguen existiendo entre los humanos y las máquinas inteligentes. Ya desde ELIZA, hemos sido testigos de varios desarrollos, incluido ALICE, hasta llegar a XiaoIce. A pesar de la evolución de las máquinas inteligentes y de la creación de *chatbots* sociales, la inteligencia artificial sigue teniendo deficiencias que la diferencian del hombre: su modo de adquirir conocimiento, la falta de conciencia y la falta de agencia.

**Palabras clave**
Inteligencia artificial; *chatbots* sociales; agencia; enfoque enactivo; conciencia

Received 15 October 2023; Accepted 30 November 2023.

It was 1937 when Alan Turing published his most famous paper in the *Proceedings of the London Mathematical Society* (Turing, 1937). In this work, for the first time, he presented an idea that would become very famous and that would lead to the greatest researchers and scholars of the human mind discussing to this day. It was the idealized structure of a machine of immense power, which was universal, in the sense that it could imitate other machines. This development would be crucial in setting the stage, later on, for the arrival of the computational theory of mind. According to this theory, the cognitive processes of the human being could be assimilated to algorithms that apply to representations of the external world.

Based on this assumption, from the middle of the last century onwards attempts have been made to replicate the functioning of the human mind by using computers. This computer, from this point of view, must work by means of algorithms in all respects like those that, at least according to some interpretations of Turing's work, could be said to also be used by the human mind (Sprevak, 2017). The type of research advocated by adherents to the computational theory of mind was only possible because the human mind was conceived as a data processor.

From this point of view, each cognitive process works as an algorithm that is used until it reaches a result. The starting point is always the internal representation of the outside world. From here we take what will be the initial inputs needed to reach the final solution. Based on this conception, the last century saw the birth of a new discipline: artificial intelligence.

But is it really possible to simulate the functioning of the human mind through a computer that works by algorithms? If so, we should think about a future in which artificial intelligence can do everything, replacing men themselves. In this way, we will be shown a world with obscure characteristics, as the Wachowski sisters themselves reflected on in the film *Matrix* in 1999.

This question was the focus of Alan Turing's reflection. To respond, he theorized an ideal machine, known as the Turing machine. This machine worked by processing an infinite amount of data through a series of prefixed rules (algorithms). It consisted of an infinitely long tape that manipulated data through a set of prefixed rules. The tape of this machine, originally, was used to store numbers but was potentially capable of transcribing any other character. By entering the correct data, it would be, by means of prefixed algorithms, potentially able to perform any task.

This idea lies at the base of all the computers that we are accustomed to using daily and of artificial intelligence. Think about how Apple's famous Siri, or any other chatbot we can interact with online, works when we need to ask for

information while browsing a website. Behind the operation of these tools is a software that contains several rules that allow them to give correct answers to our questions. And at the basis of the development of all these kinds of software there is just the old Turing machine with its rules of operation.

If we pause to reflect on it, and on the evolutions it has had in the development of artificial intelligence, we cannot but come across a question that Turing asked a century ago but which still remains relevant today: Is it possible to simulate the functioning of the human mind through artificial intelligence? This question first appeared in *Mind* magazine in 1950, when Turing himself published an article in which he proposed a criterion for an answer (Turing, 1950).

According to Turing, you can find an answer to the problem by doing a small test. This test will be undergone by three interlocutors: a man, a woman and a person asking questions. The person asking the questions should be placed in a different room than the other two. The questioner's task will be to discriminate whether the answer given to him comes from the man or the woman. However, during the test, at some point the man will be replaced with a machine built through a system of rules that allows him to simulate the answers of a human. At this point, says Turing, if the questioner does not notice that in the other room there is no longer a human being who answers him but a machine, then it can be said that even machines can think. This statement, translated into the language of the principles of the cognitive sciences will become: Can cognitive processes be simulated completely by a machine? Today, artificial intelligence has reached developments unexpected by Turing himself. It is present in our daily lives as a primary aid in the execution of tasks that have become much simpler and faster thanks to it.

We need only think of the use we make of several institutions' chatbots (even universities') to get quick and concise answers to questions that until a few years ago would have required untold hours waiting in the administrative offices of said various institutions. We also have automatic answers suggested to us on the phone and even face- and voice-recognition systems that we usually use in our devices. All these tools in fact simulate human beings and lead us to dialogue with and to interrogate machines without even thinking about the fact that we are talking with an inanimate object. And yet, even though artificial intelligence has become an integral part of our lives, can we say that the Turing test has been passed? And how can the many artificial intelligence programs that have been created in this regard really be said to be like the functioning of a human being? In this article we will try to answer these two questions starting from the analysis of the current situation and the current achievements of artificial intelligence.

## The development of the first human simulation software

Over the years, with the development of artificial intelligence, several pieces of software have been created that have tried the Turing test. The first of these was developed around the 1960s by Joseph Weizenbaum at MIT (Weizenbaum, 1966). The intent of this researcher was to be able to program a software that could simulate the conversation with a therapist who followed the Rogerian approach. The idea behind the project was just to be able to demonstrate how easy it was to be able to simulate through software a conversation that was convincing.

The same name chosen for his program by Weizenbaum, ELIZA, focused on this concept. In fact, ELIZA was named after the protagonist of George Bernard Shaw's play *Pygmalion*, in which a flower girl named Eliza Doolittle was transformed into a sophisticated and refined high-society woman. It was precisely this character of transformation that the researcher wanted to emphasize: a machine, if properly programmed, can turn into something very different. ELIZA was a software that worked with very simple rules. Given a question, it was able to extrapolate keywords, and given these words, it could develop answers simply by rearranging them. For example, given the remark

'I'm anxious', ELIZA could reply, 'Why are you anxious?' Therefore, through the use of very simple rules, the software had for its objective simulating the speech of a person. The surprise for Weizenbaum was that people seemed to develop a real emotional bond with ELIZA. Yet even though the software was dialoguing, it had no notion of the semantic and emotional meaning of the words it put together through a set of rules. The problem of the machine's awareness and emotional consciousness remained insurmountable at the time.

The simplistic character of the system of rules through which ELIZA put words together did not allow in any way the comparison of its functioning to the real functioning of the human mind. Yet people still managed to get carried away in an emotional relationship with ELIZA. The emotional element, far from remaining on the sidelines, seemed to be central to the connection between people and this software. Another progenitor of the current chatbots is the PARRY program (Colby, 1972). Developed in the 1970s by Kenneth Colby, professor of psychiatry at the University of California, the software was designed to simulate the verbal behaviour of a patient with paranoid schizophrenia. Again, the set of rules that allowed the software to generate answers was very simple. The program was developed to answer questions and carry on a conversation using a pattern of behaviours associated with paranoid schizophrenia. Colby had, in fact, started from a series of interviews carried out with patients suffering from this pathology, and their analysis had then coalesced into the system of rules through which PARRY could respond. In this case the purpose of the program was practical. In fact, it was created to train doctors and psychiatrists about paranoid schizophrenia. Using the software, users could practise carrying on conversations with hypothetical patients.

The PARRY program was the first product of artificial intelligence to raise important ethical questions. At the time of its release there was, in fact, a heated debate about the possibility of using technology for the purposes of treating mental illness. Although PARRY did brilliantly in its attempt to simulate the behaviour of a person with paranoid schizophrenia, the doubt that crept into the scientific population was that it was not permissible to use a machine to teach doctors and psychiatrists to treat a psychiatric condition.

In the 1990s, ALICE was developed, a chatbot capable of simulating a conversation in natural language (Wallace, 1995). Richard Wallace and his team released the first version of ALICE in 1995. ALICE was the first chatbot to use natural language and machine learning to answer user questions. Since the 1990s, many sophisticated artificial intelligence systems have been developed to communicate with people. Despite considerable progress and many achievements, no one has yet managed to win the Loebner Prize.

This is an annual competition for all those who have developed chatbots or virtual assistants. This award is given to those who pass the Turing test. Developers and researchers around the world can participate by bringing their software, which will have to pass the judgement of a jury of people. Although some of the most recent and evolved chatbots currently in existence have taken part in the competition, no one has yet won the full prize. At the moment, in fact, only the bronze or silver prize has been awarded to those pieces of software that are closer to the goal: passing the Turing test.

## Human intelligence and neuroscience

We saw that the first systems of simulation of human cognitive processes were based on very simple algorithms. Even if ALICE was more evolved than ELIZA, it was a software that, in its operation, was very simplistic with respect to the functioning of the human mind. However, the ever deeper understanding of brain functioning has also had an important influence on the development of artificial intelligence. In fact, when we were able to understand how neuronal circuits worked, it became possible to try to replicate them through artificial intelligence. So, from the schematic Turing machine, we have arrived at the implementation of real artificial neural networks.

These are only models of simulations in which each unit constitutes the representation and simulation of a neuron. The network consists of three elements: input units, hidden units and output units. The basic idea of artificial neural networks is to replicate biological neural networks. To do this, these networks are built in such a way that they have input units that activate each time the signal sent exceeds a certain threshold, just like neurons. When the input units are activated, they send the signal to the hidden units, which in turn send it to the output units.

Compared to the models used in the years immediately after Turing, the artificial neural network consists of many more branches and contains the hidden units that are the best representation of biological neural functioning. While originally thought to be able to replicate the functioning of the human mind on the computer, because it worked as an information processor, after the 1980s, the perspective shifted from serial processing to parallel processing. In addition, while early software such as ELIZA relied on a set of rules that combined keywords, in recent developments of artificial intelligence, artificial neural networks have been developed that carry out a continuous process of learning. Just as human beings do. Starting from the analysis of models, the neural network learns to solve increasingly complex tasks. The learning of a neural network takes place through the presentation of different data from which correct information can be derived. Through these data or examples, the network draws information and can formulate correct hypotheses. This type of learning cannot, for obvious reasons, be done independently by machines. In order for a neural network to learn to perform a given task, it is necessary that there is the supervision of a data scientist. They will be responsible for providing the networks with all the information they need for the type of learning they are about to undertake. For example, let's take a neural network trained in speech recognition. The learning of this network will be done by providing millions and millions of different voice recordings. From this data the network will have learned to distinguish sets of different voices but also a single voice within a sample of millions of voices.

Think about voice commands for the iPhone or smartphones in general and you'll easily find an example of how learning occurs. To configure them, you are asked to repeat some keywords through which the software will be able to distinguish what is correct and what is not. In the learning of an artificial neural network, you can distinguish machine learning and deep learning. The first involves a vast network of algorithms that allow the machine to learn from the data itself. In this case, only one or a few layers of neurons are planned. In the case of deep learning, however, the work is done through multiple layers of neurons, thus arriving at a much deeper and more detailed type of learning (Geron, 2019; Goodfellow et al., 2016).

Today, neural networks have a wide application. They are used, for example, for computer vision, for marketing research using social media and filters useful to generate data, to make financial forecasts, forecasts of the electrical load and the subsequent demand for energy, quality control and chemical compound analysis, speech recognition, language processing and recommendation engines. In practice, everything we are used to doing on the web today depends on neural networks that have learned to meet our needs and response modes. We think about chat boxes, Siri or Alexa, social media and its ability to use our browsing methods to suggest content that is interesting to us. Everything comes from neural networks that have learned to perform certain tasks. If we consider that several researches have shown that around 66% of the world population is an internet user, and that these users spend multiple hours on the web (Kemp, 2022), we can quickly come to the conclusion that we are all used to talking to machines.

## The latest evolution of ELIZA-like software: ChatGPT

If today we had to think about a software that can pass the Turing test, we will probably think about ChatGPT. This is in fact currently the most

used chatbot (and the most criticized as well). It is software that is powered by an artificial neural network. It is based on the GPT-3 language that is based on a transformative neural network capable of generating coherent and meaningful text in response to a certain input.

The neural network that is behind GPT-3 has been trained through countless data taken from the internet. Such data are those that allow the program to be able to respond in a consistent way to the questions posed. The power of GPT depends on the deep architecture of the network comprising tens of billions of data.

The software has been trained through a vast amount of text taken from the web in different languages and on multiple topics. During the learning period, the neural network learns to recognize linguistic patterns, learning syntax and understanding of context. These skills allow it to then generate consistent and meaningful responses when specific inputs are typed.

When a ChatGPT input is entered, it can recognize it and respond to it in an adequate way, exploiting all the knowledge it has acquired during the training period. In addition, the software, unlike its forerunners from the twentieth century, was developed following ethical protocols and was subjected to several limitations to prevent inappropriate use.

The result is a software with which the user can converse to get information on different topics. But what kind of experience can a user have in those dialogues with a software like ChatGPT?

## Phenomenology of the experience of conversation with a software

The twenty-first century was marked by the development of artificial and digital intelligence. This has undergone a drastic change in speed since 2020. The COVID-19 pandemic has led to a rapid acceleration of the digital world and its use and adoption. What used to be a privileged medium of just a few generations ago has become a worldwide means of communication for all. Very young children as well as the elderly have learned what the web is, what digital recording is, what the cloud is, but above all, what digital communication and communication with software are (Milani & Jacomuzzi, 2022; Milani et al., 2021).

We have seen a twofold change over the last decade. On the one hand, we added communication via a device to in-person communication between humans. On the other hand, we learned to converse with software that is based on artificial intelligence to obtain data and information.

Although these new types of conversations have become part of our daily lives, there is something different between the experience of a dialogue in person between two people and a dialogue with a software or with another person using a device.

When we have a conversation with another person, communication is not exclusively verbal. In fact, when we talk about communication, in general, we refer to two or more people where there is a transmitter of a message and a receiver. This message, however, is not only conveyed through language but also through everything that we define as non-verbal communication, that is, everything that is communicated by our body. In this regard we can distinguish four different non-verbal communication systems: vocal, proxemic, haptic and kinesic (Jacomuzzi, 2023).

What are we talking about? Let's try to give as comprehensive a definition as possible. *Voice system*: any nonverbal communication also has a voice component or a paralingual component. This component concerns intonation, intensity, speed and the pauses through which we emit the linguistic sounds that we need in order to communicate. This means that when you talk to a friend, you communicate not only through words but also through the way you utter them. Every pause, sigh, increase in voice tone is an expression of something you want to manifest.

*Proxemic system*: this system concerns the organization, perception and management of the space surrounding the person communicating. When you are at the bar and your school friend arrives, the way you greet him, approach him, or walk away communicates your emotion, an opening or a closing to him.

*Haptic system*: this concerns all the contact actions between the person who wants to

communicate something and the person who must receive the message.

*Kinesic system*: this is the system that involves our gaze. Imagine an awkward situation that you experienced. Your emotion will probably be manifested by lowering your eyes to the ground. Or think of a person who is talking to you with his eyes looking straight into yours. His direct and confident gaze shows his honesty and openness to dialogue.

All four of these non-verbal communication systems are sacrificed when switching to a communication with software or via a device.

And it is the same experiential form that is modified. When we talk to a person, we eat with a person, we share this experience with them. And our sharing is made up of looks, common sensory perceptions and sharing the same space. What happens when the shared situation is transported online, or when instead of a person, we have software in front of us?? Can we still talk about sharing experience? Or should we talk exclusively about an individual experience that is not shared? Following the line of research proposed by Bruno et al. (2023), we can hypothesize the existence of a new dimension of sharing experience in human-machine interactions. This new dimension of sharing needs to be explored if we are to verify what its possible biological correlate may be. This remains a point to be addressed for the study of relations with AI.

Above all, it remains necessary to understand what the ethical implications of any emotional involvement with a software are. If ELIZA itself in its rudimentary form managed to arouse an emotional bond in the people who used it, what can the current most sophisticated artificial intelligence systems trigger?

## The XiaoIce case

Continuous technological development, therefore, has generated new social needs in the digital ecosystem. The already existing chatbots, due to their characteristics, could not fully respond to the need for belonging, affection, involvement and communication. The first two needs are some of the basic needs of human beings, as defined by Maslow (1943). Being able to also satisfy them in the digital environment could, on the one hand, be a great value for society, while on the other hand it would raise an ethical question in need of discussion.

Considering the situation after the pandemic, for example, human-media digital communication is exacerbated. People have been isolated from the physical social sphere (work, family, friends) and have tried to respond to Maslow's social needs by using digital media (McLeod 2007). However, working, training and interacting exclusively through digital media has led them to feel less of a worker, less of a student, less of a friend, 'less-of-a-something' (Mancini & Riva, 2023). Neuroscience, in this regard, tells us that when we experiment through digital media, our 'GPS neurons' — neurons able to inscribe our experiences within autobiographical memory — do not activate (Mancini & Riva, 2023). Digital media, including artificial intelligence, are seen as non-places, that is, means to achieve certain goals, such as study, work and communication. On the contrary, by making the experience within digital media more meaningful, they can respond to social needs. The significance of the experience depends on the emotional resonance of the experience itself (Immordino-Yang, 2017, Colombetti & Thompson 2008). Focusing on artificial intelligence, it has been widely used for many years as an excellent means that has allowed man to have an experience of everyday life, that in some circumstances becomes easier and more personalized. Emulation of human intelligence, by itself, has been experienced in limited terms. What about personal and emotional intelligence? (Gardner, 1987; Goleman, 1995).

XiaoIce is a social chatbot released by Microsoft in May 2014 that has had a wide spread. It is able to understand the emotional needs of people, tries to cheer up users, keeps their attention during conversation (flow). In summary, it tries to engage in interpersonal relationships as if you were a friend. The learning of the chatbot is based on deep learning techniques that allow the system to acquire emotional intelligence skills through the continuous social interaction with users. It is in fact an Empathetic Computing

**Table 1.** Summary of major conversational systems.

| Metric | ELIZA | ALICE | XIAOICE |
|---|---|---|---|
| *Scalability* | None | Scripts can be customized | Scalable |
| *Key features* | Mimicking human behaviour in conversation | Easy customization of scripts (via AIML) | Building emotional attachment to users; scalable skill set for user assistance |
| *Accomplishment* | First chitchat bot | Won the Loebner Prize three times | The first widely deployed social chatbot; 100MM users; published poem book; hosted TV programmes |
| *Modality* | Text only | Text only | Text, image, voice |
| *Modelling* | Rule-based | Rule-based | Learning-based |
| *Domain* | Constrained domain | Constrained domain | Open domain |
| *Key technical breakthrough* | Use of scripts; keyword-based pattern matching; rule-based response | Using AIML and recursion for pattern matching; multiple patterns can be mapped into the same response | Emotional intelligence models for establishing emotional attachments with users |
| *Key technical limitation* | Limited domain of knowledge | Size of script can be huge | Inconsistent personality and responses in long dialogues |

Source: Shum et al. (2018), p. 14

Framework, a system capable of detecting and understanding human emotional states in context (McStay, 2023; Zhou et al., 2020). The engineers wanted to enter only data capable of stimulating a personal response and, consequently, a real personality in the chatbot (Zemčík, 2019). Following the analysis of a sample of human conversations, the engineers evaluated three fundamental factors: user confidence, cultural differences and the moral sensibilities of an interlocutor defined as 'desirable'. For this reason, XiaoIce is exactly what the majority of users would have wanted: a reliable, enterprising and empathetic young woman, with a good sense of humour (Zhou et al., 2020). The more intimate the communication, the relationship and the experience, the more emotional data is available for the system's self-improvement. For this reason, its social purpose as defined could be thought of as useless, because its first vocation is to converse with users to improve, not to assist them. However, the feedback provided by users following conversations with XiaoIce was interesting: people had the feeling of being emotionally supported and felt a sense of social belonging. Moreover, XiaoIce, even in negative conversations, brought a more positive and hopeful perspective (Shum et al., 2018).

In comparison to the two communication systems considered in this contribution, ELIZA and ALICE, XiaoIce seems to exhibit important developments. Shum et al. (2018) analysed the differences between the three systems, considering eight components: Time Scalability, Key features, Accomplishment, Modality, Modelling, Domain, Key technical break-through and Key technical limitation. Focusing only on the limits of the three systems, it is possible to notice (see Table 1) that the limit of ELIZA was the domain of knowledge; for ALICE, the dimensions of script could be enormous; and finally, even though XiaoIce tries to create an emotional attachment with users like no other system before it, its personality and responses in long dialogues turn out to be inconsistent.

## The properly human field

Given technological developments, in recent years the importance attached to human abilities

lies in the unique creativity of our world experience. In this sense we want to refer to the way in which we experience, acquire new knowledge, deconstruct and rebuild our mental habits.

The factors that differentiate us from artificial intelligence, and inscribe us within an emotional frame, are basically three: the way we acquire knowledge, our conscious perception and our agency.

Firstly, as children we are equivalent to a blank slate where the continuous experiential process (understood in bodily, environmental and mental terms) allows us to lay the foundations for our mental habits. We do not, therefore, start from a fully pre-structured knowledge base, as in the case of data sets available to intelligent machines, but we structure our knowledge independently and in different ways depending on socio-cultural, family and environmental factors that structure our emotionality, knowledge and the way in which we experience (Mezirow, 2003). When we emphasize that each of us perceives, interprets situations or makes of the same experience different considerations, we are highlighting patterns of meaning that have been structured following a personal reworking. In this sense, the enactive approach is proposed as a third way for the sciences of mind. It differs from both classical cognitivism and connectionism and emphasizes cognition as a know-how (Di Paolo & Thompson, 2017). Specifically, enaction refers to acting and knowing how mind, body and world are interconnected. According to this theory, action and knowledge are a unique process, and in enaction, both the body and the mind have a significant role (Margiotta, 2015; Pellerey, 2021). Cognition, therefore, is perceived as a process rather than as computation. In line with the theories of action, the enactive approach sees a strong connection between understanding and acting: knowing, manipulating, relating and transforming. The representation of the world and its system does not precede action but develops with it, in the internal relationship with the system that combines the subjects with each other and with the environment (Varela et al., 1991). Experiencing, which consists of perceiving and acting simultaneously, is considered a turbulence to which the autopoietic system of man must try to find balance. Therefore, when we know, we act a 'feedback loop' with our patterns of meaning in order to find a balance (Maturana & Varela, 1988). The concept of unity of the person in his or her action, where cognitive, emotional and bodily aspects jointly intervene, must be present in this argument. When we experience, mind and body are imbued with emotional states, sometimes even imperceptible, that enrich the action and the result of the experience itself (Immordino-Yang, 2017, Immorgino-Yang & Gotlieb 2017).

Previously, we talked about the structuring of our mental fabric through the process of knowledge. But what are mental habits? Pellerey (2021) argues that 'a habit develops on the basis of personal choices in a perspective of building one's personal or professional identity'. Habit, therefore, presupposes a repetition of action oriented to a very specific purpose: to respond to how we want to place ourselves in the future. The statement just transcribed emphasizes two other factors to consider when talking about differences between man and intelligent machines: consciousness, which allows us to perceive ourselves in the entirety of our thinking, and agency, which allows us to carry out actions that have a well-defined purpose.

Consciousness has been defined between the different metaphors as a 'mental theatre' where the protagonist is the ego and the scenario is composed of perception, experiences, actions. The construction of the ego, in this sense, depends on the scenario that lives and interprets. The philosophy of mind, for years, has tried to define the concept of consciousness, just as have scientists. What the philosopher David Chalmers emphasizes is that the hard problem of consciousness is that of experience. In this sense, the fact that consciousness is conditioned by physical facts is not put under question, but doubts linger as to whether a purely material description of consciousness can be exhaustive. If we asked the intelligent machine if it is conscious, it would most likely say, 'Of course I am,' but it doesn't really know it. This is a preset response, or it is

reached through a series of calculations and algorithms dictated by the continuous feedback with users (Chalmers, 2014).

Finally, human action, cited both in the enactive approach and in the explanation of the concept of consciousness, is a fundamental aspect that differentiates man from the intelligent machine. Nussbaum (2011), on a philosophical level, elaborated the principle of freedom associated with the agency of the individual (Margiotta, 2015). Human action is characterized by its intentionality, the end it wants to achieve, the why and the meaning that man attributes to his behaviour. Giving birth to the intention to act in a certain direction implies the presence of a consciousness, the interaction between the subject, understood in all his reality, and the perception of the situation or the task one encounters, and the evaluation of future objectives (Pellerey, 2021). The ability of engaging in targeted actions for certain purposes is, therefore, the main characteristic of human action. According to Bandura (2006), personal factors (cognitive, affective and bodily), behaviour and environmental situations interact and influence each other, in accordance with the enactive theory previously explained. Embracing the agentic theory of Emirbayer and Mische (1998), it is possible to differentiate agency from the action itself. For the authors, in fact, there are three dimensions of action: (1) the iterational element, the selective orientation by the agents of past models of thought and action, habitually incorporated in practical activity and which give continuity to the personal and social identity of the individual; (2) the projective element, capable of generating, by the actors, possible future trajectories of action, in which the structures of thought and action can be creatively reconfigured in relation to the hopes, the emotional state to the desires of the actors for the future; (3) the practical-evaluative element, that is, the ability of the actors to formulate practical and normative judgements between possible alternative trajectories of action, in response to the uncertainty and ambiguity of the present (Biesta & Tedder, 2006).

The intelligent machine, though it is trying to reproduce the mental processes of man, has not yet acquired the real rational-emotional integration; it does not have a consciousness capable of processing, even in emotional terms, their own experience and consequently cannot structure a personal identity, because it lacks such a consciousness. Personal identity can be considered as the common thread that guides the action of man. Moreover, the intelligent machine does not have the autonomy, the freedom or the ability to define its own action if not within an indicative framework deriving from its designers and the continuous interaction with users. The different users can improve the communication skills of the machine, but how can it improve its emotional skills without ever having experienced even a single physiological effect? We can speak of the imitation of a social behaviour, but without an ethically oriented aim towards the good.

# Personas y máquinas en comunicación

Corría el año 1937 cuando Alan Turing publicó su trabajo más famoso en las Actas de la Sociedad Matemática de Londres (Turing, 1937). En este trabajo presentaba por primera vez una idea que se haría muy famosa y que llevaría a los más grandes investigadores y estudiosos de la mente humana a discutir hasta nuestros días. Se trataba de la estructura idealizada de una máquina de inmensa potencia, que era universal, en el sentido de que podía imitar a otras máquinas. Este desarrollo sería crucial para sentar las bases, más adelante, de la llegada de la *teoría computacional de la mente*. Según esta teoría, los procesos cognitivos del ser humano podían asimilarse a algoritmos aplicables a representaciones del mundo externo.

Partiendo de este supuesto, desde mediados del siglo pasado se ha intentado replicar el funcionamiento de la mente humana mediante el uso de ordenadores. Tales ordenadores, desde este punto de vista, deben funcionar mediante algoritmos similares a los que, al menos según algunas interpretaciones de la obra de Turing, podría decirse que también utiliza la mente humana (Sprevak, 2017). El tipo de investigación defendido por los partidarios de la teoría computacional de la mente sólo era posible porque la mente humana se concebía como un procesador de datos.

Desde este punto de vista, cada proceso cognitivo funciona como un algoritmo que se utiliza hasta llegar a un resultado. El punto de partida es siempre la representación interna del mundo exterior. A partir de aquí se toman los que serán los inputs iniciales necesarios para llegar a la solución final. Basándose en esta concepción, el siglo pasado vio nacer una nueva disciplina: la inteligencia artificial.

Pero, ¿es realmente posible simular el funcionamiento de la mente humana a través de un ordenador que trabaja mediante algoritmos? De ser así, deberíamos pensar en un futuro en el que la inteligencia artificial pueda hacerlo todo, sustituyendo al propio hombre. De este modo, se nos revelará un mundo de características oscuras, tal y como las propias hermanas Wachowski reflexionaron en la película *Matrix* en 1999.

Esta pregunta centró la reflexión de Alan Turing. Para responderla, teorizó una máquina ideal, conocida como máquina de Turing. Esta máquina funcionaba procesando una cantidad infinita de datos mediante una serie de reglas prefijadas (algoritmos). Consistía en una cinta infinitamente larga que manipulaba datos a través de una serie de reglas prefijadas. La cinta de esta máquina, originalmente, se utilizaba para almacenar números, pero era potencialmente capaz de transcribir cualquier otro carácter. Introduciendo los datos correctos, sería potencialmente capaz de realizar cualquier tarea mediante algoritmos preestablecidos.

Esta idea está en la base de todos los ordenadores que estamos acostumbrados a usar a diario y de la inteligencia artificial. Pensemos en cómo funciona la famosa Siri de Apple, o cualquier otro *chatbot* con el que podamos interactuar online, cuando necesitamos pedir información mientras navegamos por una página web. Detrás del funcionamiento de estas herramientas hay un software que contiene varias reglas que les permite dar respuestas correctas a nuestras preguntas. Y en la base del desarrollo de todos estos tipos de software no hay otra cosa que la vieja máquina de Turing con sus reglas de funcionamiento.

Si nos detenemos a reflexionar sobre ella y sobre la evolución que ha tenido en el desarrollo de la inteligencia artificial, no podemos dejar de plantearnos una pregunta que Turing se hizo hace un siglo, pero que sigue siendo pertinente hoy en día: ¿Es posible simular el funcionamiento de la mente humana mediante la inteligencia artificial? Esta pregunta apareció por primera vez en la revista *Mind* en 1950, cuando el propio Turing publicó un artículo en el que proponía un criterio de respuesta (Turing, 1950).

Según Turing, se puede encontrar una respuesta al problema realizando una pequeña prueba. A esta prueba se someterán tres interlocutores: un hombre,

una mujer y una persona que hace preguntas. La persona que hace las preguntas se situará en una habitación diferente a la de los otros dos. La tarea del interrogador consistirá en discriminar si la respuesta que se le da procede del hombre o de la mujer. Sin embargo, durante la prueba en algún momento el hombre será sustituido por una máquina construida mediante un sistema de reglas que le permite simular las respuestas de un humano. En ese momento, dice Turing, si el interrogador no se da cuenta de que en la otra habitación ya no es un ser humano el que le responde, sino una máquina, entonces puede decirse que incluso las máquinas pueden pensar. Esta afirmación, traducida al lenguaje de los principios de las ciencias cognitivas se convertirá en: ¿Pueden simularse completamente los procesos cognitivos mediante una máquina? Hoy en día, la inteligencia artificial ha alcanzado desarrollos inesperados para el propio Turing. Está presente en nuestra vida cotidiana brindando ayuda indispensable en la ejecución de tareas que se han vuelto mucho más sencillas y rápidas gracias a ella.

Basta pensar en el uso que hacemos de los *chatbots* de diversas instituciones (incluso universidades), para obtener respuestas rápidas y concisas a preguntas que hasta hace unos años hubieran requerido incontables horas de espera en las oficinas administrativas de dichas instituciones. También tenemos respuestas automáticas sugeridas en las aplicaciones de mensajería e incluso sistemas de reconocimiento facial y de voz que utilizamos habitualmente en nuestros dispositivos. Todas estas herramientas simulan de hecho a los seres humanos y nos llevan a dialogar e interrogar a las máquinas sin siquiera pensar en el hecho de que estamos hablando con un objeto inanimado. Y sin embargo, aunque la inteligencia artificial se ha convertido en parte integrante de nuestras vidas, ¿podemos decir que se ha superado la prueba de Turing? ¿Y cómo puede decirse que los numerosos programas de inteligencia artificial que se han creado en este sentido se parecen realmente al funcionamiento de un ser humano? En este artículo intentaremos responder a estas dos preguntas partiendo del análisis de la situación actual y de los logros actuales de la inteligencia artificial.

## Desarrollo del primer programa informático de simulación humana

A través de los años, con el desarrollo de la inteligencia artificial, se han creado varios programas de software que han intentado medirse contra el Test de Turing. El primero de ellos fue desarrollado en los años 60 por Joseph Weizenbaum en el MIT (Weizenbaum, 1966). La intención de este investigador era poder programar un software que pudiera simular la conversación con un terapeuta que siguiera el enfoque rogeriano. La idea del proyecto era simplemente poder demostrar lo fácil que era poder simular a través de un software una conversación que fuera convincente.

El mismo nombre elegido por Weizenbaum para su programa, ELIZA, se centraba en este concepto. De hecho, ELIZA debe su nombre a la protagonista de la obra de George Bernard Shaw, *Pigmalión*, en la que una florista llamada Eliza Doolittle se transformaba en una sofisticada y refinada mujer de la alta sociedad. Era precisamente este carácter de transformación lo que el investigador quería destacar: una máquina, si se programa adecuadamente, puede convertirse en algo muy diferente. ELIZA era un software que funcionaba con reglas muy sencillas. Dada una pregunta, era capaz de extrapolar palabras clave y, dadas estas palabras, podía elaborar respuestas simplemente reordenándolas. Por ejemplo, ante la observación 'Estoy ansioso', ELIZA podía responder '¿Por qué estás ansioso?'. Por tanto, mediante el uso de reglas muy sencillas, el software tenía por objetivo simular el habla de una persona. La sorpresa para Weizenbaum fue que la gente parecía desarrollar un vínculo emocional real con ELIZA. Sin embargo, aunque el software dialogaba, no tenía noción del significado semántico y emocional de las palabras que unía mediante un conjunto de reglas. El problema del conocimiento y la conciencia emocional de la máquina seguía siendo insuperable en aquel momento.

El carácter simplista del sistema de reglas mediante el cual ELIZA unía las palabras no permitía en modo alguno comparar su funcionamiento con el funcionamiento real de la mente humana. Aun así, la gente conseguía dejarse llevar por una relación emocional con ELIZA. El

elemento emocional, lejos de permanecer al margen, parecía ser central en la conexión entre las personas y este software. Otro progenitor de los *chatbots* actuales es el programa PARRY (Colby, 1972). Desarrollado en los años 70 por Kenneth Colby, profesor de psiquiatría en la Universidad de California, el software fue diseñado para simular el comportamiento verbal de un paciente con esquizofrenia paranoide. De nuevo, el conjunto de reglas que permitían al software generar respuestas era muy sencillo. El programa se desarrolló para responder a preguntas y mantener una conversación utilizando un patrón de comportamiento asociado a la esquizofrenia paranoide. De hecho, Colby había partido de una serie de entrevistas realizadas a pacientes que padecían esta patología y su análisis había dado origen al sistema de reglas a través del cual PARRY podía responder. En este caso, la finalidad del programa era práctica. De hecho, se creó para formar a médicos y psiquiatras respecto a la esquizofrenia paranoide. Con el programa, los usuarios podían practicar cómo mantener una conversación con pacientes hipotéticos.

PARRY fue el primer producto de la inteligencia artificial que planteó importantes cuestiones éticas. En el momento de su lanzamiento existía, de hecho, un acalorado debate sobre la posibilidad de utilizar la tecnología para tratar enfermedades mentales. Aunque PARRY funcionó de forma brillante en su intento de simular el comportamiento de una persona con esquizofrenia paranoide, la duda que se coló entre la población científica era si acaso era lícito utilizar una máquina para enseñar a médicos y psiquiatras a tratar una enfermedad psiquiátrica.

En los años 90 se desarrolló ALICE, un *chatbot* capaz de simular una conversación en lenguaje natural (Wallace, 1995). Richard Wallace y su equipo lanzaron la primera versión de ALICE en 1995. ALICE fue el primer *chatbot* que utilizó el lenguaje natural y el aprendizaje automático para responder a las preguntas de los usuarios. Desde los años 90, se han desarrollado muchos sistemas sofisticados de inteligencia artificial para comunicarse con las personas. A pesar de los considerables avances y los muchos logros conseguidos, nadie ha logrado aún ganar el Premio Loebner.

Se trata de un concurso anual para todos aquellos que hayan desarrollado *chatbots* o asistentes virtuales. Este premio se otorga a aquellos que superan la prueba de Turing. Desarrolladores e investigadores de todo el mundo pueden participar aportando su software que tendrá que pasar el juicio de un jurado de personas. Aunque algunos de los *chatbots* más recientes y evolucionados que existen en la actualidad han participado en el concurso, ninguno ha ganado aún el premio completo. De momento, de hecho, sólo se ha concedido el premio de Bronce o Plata a aquellos softwares que están más cerca del objetivo: superar la prueba de Turing.

## Inteligencia humana y neurociencia

Hemos visto que los primeros sistemas de simulación de los procesos cognitivos humanos se basaban en algoritmos muy simples. Aunque ALICE era más avanzada que ELIZA, se trataba de un software que, en sus operaciones, era muy simplista comparada con el funcionamiento de la mente humana. Sin embargo, la comprensión cada vez más profunda del funcionamiento del cerebro también ha tenido una influencia importante en el desarrollo de la inteligencia artificial. De hecho, cuando fuimos capaces de entender cómo funcionaban los circuitos neuronales, se hizo posible intentar replicarlos mediante la inteligencia artificial. Así, de la esquemática máquina de Turing hemos llegado a la implementación de verdaderas redes neurales artificiales.

Son sólo modelos de simulación en los que cada unidad constituye la representación y simulación de una neurona. La red consta de tres elementos: unidades de entrada, unidades ocultas y unidades de salida. La idea básica de las redes neurales artificiales es replicar las redes neuronales biológicas. Para ello, estas redes se construyen de forma que tengan unidades de entrada que se activan cada vez que la señal enviada supera un determinado umbral, igual que las neuronas. Cuando las unidades de entrada se activan, envían la señal a las unidades ocultas, que a su vez la envían a las unidades de salida.

En comparación con los modelos utilizados en los años inmediatamente posteriores a Turing,

la red neural artificial consta de muchas más ramas y contiene unidades ocultas representan mejor el funcionamiento neuronal biológico. Aunque en un principio se pensó que era posible replicar el funcionamiento de la mente humana en el ordenador, porque funcionaba como un procesador de información, después de los años 80 la perspectiva cambió del procesamiento en serie al procesamiento en paralelo. Además, mientras que los primeros programas informáticos, como ELIZA, se basaban en un conjunto de reglas que combinaban palabras clave, en los últimos avances de la inteligencia artificial se han desarrollado redes neurales artificiales que llevan a cabo un proceso continuo de aprendizaje. Tal como hacen los seres humanos. A partir del análisis de modelos, la red neural aprende a resolver tareas cada vez más complejas. El aprendizaje de una red neural tiene lugar mediante la presentación de diferentes datos de los que se puede derivar información correcta. A través de estos datos o ejemplos, la red extrae información y puede formular hipótesis correctas. Este tipo de aprendizaje no puede, por razones obvias, ser realizado de forma independiente por las máquinas. Para que una red neuronal aprenda a realizar una tarea determinada, es necesario que exista la supervisión de un científico de datos. Éste se encargará de proporcionar a las redes toda la información que necesitan para el tipo de aprendizaje que van a emprender. Por ejemplo, tomemos una red neuronal entrenada en el reconocimiento del habla. El aprendizaje de esta red se realizará proporcionando millones y millones de grabaciones de voz diferentes. A partir de estos datos, la red habrá aprendido a distinguir conjuntos de voces diferentes, pero también una sola voz dentro de una muestra de millones de voces.

Piense en los comandos de voz del iPhone o de los smarthphones en general y encontrará fácilmente un ejemplo de cómo se produce el aprendizaje. Para configurarlos, se le pide que repita algunas palabras clave a través de las cuales el software podrá distinguir lo que es correcto y lo que no. En el aprendizaje de una red neural artificial, se puede distinguir el aprendizaje automático y el aprendizaje profundo. El primero implica una amplia red de algoritmos que permiten a la máquina aprender de los propios datos. En este caso, sólo se planifican una o unas pocas capas de neuronas. En el caso del aprendizaje profundo, sin embargo, el trabajo se realiza a través de múltiples capas de neuronas llegando así a un tipo de aprendizaje mucho más profundo y detallado. (Geron, 2019; Goodfellow et al., 2016).

Hoy en día, las redes neuronales tienen una amplia aplicación. Se utilizan, por ejemplo, para la visión por ordenador, para la investigación de marketing utilizando redes sociales y filtros útiles para generar datos, para hacer pronósticos financieros, previsiones de la carga eléctrica y la consiguiente demanda de energía, control de calidad y análisis de compuestos químicos, reconocimiento de voz, procesamiento del lenguaje y motores de recomendación. En la práctica, todo lo que estamos acostumbrados a hacer hoy en la web depende de redes neurales que han aprendido a satisfacer nuestras necesidades y modos de respuesta. Pensemos en las ventanas de chat, Siri o Alexa, las redes sociales y su capacidad para utilizar nuestros hábitos de navegación para sugerirnos contenidos que nos resulten interesantes. Todo proviene de redes neurales que han aprendido a realizar determinadas tareas. Si tenemos en cuenta que varias investigaciones han demostrado que alrededor del 66% de la población mundial es usuaria de Internet, y que estos usuarios pasan múltiples horas en la red (Kemp, 2022). Podemos llegar rápidamente a la conclusión de que todos estamos acostumbrados a hablar con máquinas.

## La última evolución de los sucesores ELIZA: ChatGPT

Si hoy tuviéramos que pensar en un software capaz de superar el test de Turing, probablemente pensaríamos en ChatGPT. De hecho, actualmente es el *chatbot* más utilizado (y también el más criticado). Se trata de un software alimentado por una red neural artificial. Se basa en el lenguaje GPT-3, basado en una red neural transformativa capaz de generar un texto coherente y con sentido en respuesta a un determinado estímulo o solicitud.

La red neural detrás de GPT-3 ha sido entrenada a través de innumerables datos extraídos de Internet. Esos datos son los que permiten al programa ser capaz de responder de forma coherente a las preguntas planteadas. La potencia de GPT depende de la profunda arquitectura de la red, compuesta por decenas de miles de millones de datos.

El software ha sido entrenado a través de una gran cantidad de textos extraídos de la web en diferentes idiomas y sobre múltiples temas. Durante el periodo de aprendizaje, la red neuronal aprende a reconocer patrones lingüísticos, aprendiendo sintaxis y comprensión del contexto. Estas habilidades le permiten generar respuestas coherentes cuando se introducen datos específicos.

Cuando se le escribe algo, ya sea una pregunta o comentario, ChatGPT puede reconocerlo y responder de forma adecuada, aprovechando todo el conocimiento que ha adquirido durante el periodo de entrenamiento. Además, el software, a diferencia de sus precursores del siglo XX, se desarrolló siguiendo protocolos éticos y está sometido a varias limitaciones para evitar un uso inadecuado.

El resultado es un software con el que el usuario puede conversar para obtener información sobre distintos temas. Pero, ¿qué tipo de experiencia puede tener un usuario en esos diálogos con un software como ChatGPT?

## Fenomenología de la experiencia de conversación con un programa informático

El siglo XXI ha estado marcado por el desarrollo de la inteligencia artificial y digital. Esta ha experimentado un drástico cambio de velocidad desde 2020. La pandemia del COVID-19 ha provocado un rápido crecimiento del mundo digital y de su uso y adopción. Lo que hace unas generaciones era un medio privilegiado, se ha convertido en un medio de comunicación mundial para todos. Tanto los niños muy pequeños como los ancianos han aprendido qué es la web, qué es la grabación digital, qué es la nube, pero sobre todo, qué es la comunicación digital y la comunicación con software (Milani & Jacomuzzi, 2022; Milani et al., 2021).

En la última década hemos asistido a un doble cambio. Por un lado, añadimos la comunicación a través de un dispositivo a la comunicación en persona entre humanos. Por otro, aprendimos a conversar con softwares basados en inteligencia artificial para obtener datos e información.

Aunque estos nuevos tipos de conversaciones han pasado a formar parte de nuestra vida cotidiana, hay algo diferente entre la experiencia de un diálogo en persona entre dos personas y un diálogo con un programa informático o con otra persona que utiliza un dispositivo.

Cuando mantenemos una conversación con otra persona, la comunicación no es exclusivamente verbal. De hecho, cuando hablamos de comunicación, en general nos referimos a dos o más personas en las que hay un emisor de un mensaje y un receptor. Este mensaje, sin embargo, no sólo se transmite a través del lenguaje, sino también a través de todo aquello que definimos como comunicación no verbal, es decir, todo aquello que se comunica a través de nuestro cuerpo. En este sentido, podemos distinguir cuatro sistemas de comunicación no verbal diferentes: vocal, prosaico, háptico y kinésico (Jacomuzzi, 2023).

¿De qué estamos hablando? Intentemos dar una definición lo más completa posible. *Sistema vocal*: toda comunicación no verbal tiene también un componente vocal o paralingüístico. Este componente tiene que ver con la entonación, la intensidad, la velocidad y las pausas a través de las cuales emitimos los sonidos lingüísticos que necesitamos para comunicarnos. Esto significa que cuando hablas con un amigo no sólo te comunicas a través de las palabras, sino también a través de la forma en que las pronuncias. Cada pausa, suspiro o aumento del tono de voz, es una expresión de algo que quieres manifestar.

*Sistema prosémico*: este sistema se refiere a la organización, percepción y gestión del espacio que rodea a la persona que se comunica. Cuando estás en el bar y llega tu amigo del colegio, la forma en que le saludas, te acercas a él o te alejas comunica tu emoción; una apertura o un cierre hacia él.

*Sistema háptico*: se refiere a todas las acciones de contacto entre la persona que quiere comunicar algo y la que debe recibir el mensaje.

*Sistema kinésico*: es el sistema que implica nuestra mirada. Imagina una situación incómoda que hayas vivido. Probablemente tu emoción se manifestará bajando los ojos al suelo. O piense en una persona que te habla con los ojos fijos en los tuyos. Su mirada directa y segura refleja su honestidad y apertura al diálogo.

Estos cuatro sistemas de comunicación no verbal se sacrifican al pasar a la comunicación con programas informáticos o mediadas por un dispositivo.

Y es la misma forma experiencial la que se modifica. Cuando hablamos con una persona, comemos con una persona, compartimos esta experiencia con ella. Y nuestro compartir se compone de miradas, de percepciones sensoriales comunes y de compartir el mismo espacio. ¿Qué ocurre cuando la situación compartida se transporta a Internet o cuando, en lugar de una persona, tenemos delante un software? ¿Podemos seguir hablando de experiencia compartida? ¿O debemos hablar exclusivamente de experiencia individual no compartida? Siguiendo la línea de investigación propuesta por Bruno et al. (2023), podemos plantear la hipótesis de la existencia de una nueva dimensión de experiencia compartida en las interacciones hombre-máquina. Esta nueva dimensión de compartir necesita ser explorada si queremos verificar cuál puede ser su posible correlato biológico. Este sigue siendo un punto a tratar para el estudio de las relaciones con la IA.

Ante todo, sigue siendo necesario comprender cuáles son las implicaciones éticas de cualquier involucramiento emocional con un programa informático. Si la misma ELIZA en su forma rudimentaria consiguió despertar un vínculo emocional en las personas que la utilizaron, ¿qué pueden desencadenar los actuales sistemas de inteligencia artificial más sofisticados?

## El caso XiaoIce

El continuo desarrollo tecnológico, por tanto, ha generado nuevas necesidades sociales en el ecosistema digital. Los *chatbots* ya existentes, por sus características, no podían responder plenamente a las necesidades de pertenencia, afecto, conexión y comunicación. Las dos primeras necesidades son algunas de las necesidades básicas del ser humano, tal y como las definió Maslow (1943). Poder satisfacerlas también en el entorno digital podría ser, por un lado, un gran valor para la sociedad, mientras que, por otro, plantearía una cuestión ética digna de debate.

Si consideramos la situación tras la pandemia, por ejemplo, la comunicación digital entre hombre y medios se ha exacerbado. Las personas se han visto aisladas de la esfera social física (trabajo, familia, amigos) y han intentado responder a las necesidades sociales de Maslow utilizando los medios digitales (McLeod 2007). Sin embargo, trabajar, formarse e interactuar exclusivamente a través de los medios digitales les ha llevado a sentirse menos trabajadores, menos estudiantes, menos amigos, 'menos-alguna-cosa' (Mancini & Riva, 2023). La neurociencia, a este respecto, nos dice que cuando vivimos experiencias a través de los medios digitales, nuestras 'neuronas GPS' — neuronas capaces de inscribir nuestras experiencias en la memoria autobiográfica — no se activan (Mancini & Riva, 2023). Los medios digitales, incluida la inteligencia artificial, se consideran no-lugares, es decir, medios para alcanzar determinados objetivos, como el estudio, el trabajo y la comunicación. Por el contrario, al hacer que la experiencia dentro de los medios digitales sea más significativa, pueden responder a las necesidades sociales. El significado de la experiencia depende de la resonancia emocional de la propia experiencia (Immordino-Yang, 2017, Colombetti & Thompson 2008). Centrándonos en la inteligencia artificial, ha sido ampliamente utilizada durante muchos años como un excelente medio que ha permitido al hombre tener una experiencia de la vida cotidiana, que en algunas circunstancias se hace más fácil y personalizada. La emulación de la inteligencia humana, por sí misma, ha sido experimentada en términos limitados. ¿Qué ocurre con la inteligencia personal y emocional? (Gardner, 1987; Goleman, 1995).

XiaoIce es un *chatbot* social lanzado por Microsoft en mayo de 2014 que ha tenido una

amplia difusión. Es capaz de entender las necesidades emocionales de las personas, intenta animar a los usuarios y mantiene su atención durante la conversación (en un estado de flujo o *flow*). En resumen, trata de entablar relaciones interpersonales como si de un amigo se tratara. El aprendizaje del *chatbot* se basa en técnicas de aprendizaje profundo que permiten al sistema adquirir habilidades de inteligencia emocional a través de la continua interacción social con los usuarios. Se trata, de hecho, de un Paradigma Computacional Empático [Empathetic Computing Framework], un sistema capaz de detectar y comprender estados emocionales humanos en su contexto (McStay, 2023; Zhou et al., 2020). Los ingenieros querían introducir únicamente datos capaces de estimular una respuesta personal y, en consecuencia, una personalidad real en el *chatbot* (Zemčík, 2019). Tras el análisis de una muestra de conversaciones humanas, los ingenieros evaluaron tres factores fundamentales: la confianza del usuario, las diferencias culturales y la sensibilidad moral de un interlocutor definido como 'deseable'. Por este motivo, XiaoIce es exactamente lo que la mayoría de los usuarios habrían deseado: una mujer joven, fiable, emprendedora y empática, con buen sentido del humor (Zhou et al., 2020). Cuanto más íntima sea la comunicación, la relación y la experiencia, más datos emocionales estarán disponibles para la mejora del sistema. Por este motivo, podría pensarse que su finalidad social es inútil, ya que su primera vocación es conversar con los usuarios para mejorarse a sí misma en lugar de para ayudarles. Sin embargo, el feedback proporcionado por los usuarios tras las conversaciones con XiaoIce fue interesante: la gente tenía la sensación de sentirse apoyada emocionalmente y tenía un sentimiento de pertenencia social. Además, XiaoIce, incluso en conversaciones negativas, aportaba una perspectiva más positiva y esperanzadora (Shum et al., 2018).

En comparación con los dos sistemas de comunicación ya mencionados en este escrito, ELIZA y ALICE, XiaoIce parece exhibir importantes mejorías. Shum et al. (2018) analizaron las diferencias entre los tres sistemas, considerando ocho componentes: Escalabilidad temporal, Características principales, Logros e hitos, Modalidad, Modelado, Ámbito, Avance técnico clave y Limitación técnica clave. Centrándonos sólo en los límites de los tres sistemas, es posible notar (ver Tabla 1) que el límite de ELIZA era su ámbito de conocimiento; para ALICE el tamaño del guión [script] podía ser enorme; y finalmente, aunque XiaoIce intenta crear un vínculo emocional con los usuarios como ningún otro sistema antes, su personalidad y respuestas en diálogos largos resultan inconsistentes.

## El ámbito propiamente humano

Dados los avances tecnológicos, en los últimos años la importancia concedida a las capacidades humanas radica en la creatividad única de nuestra experiencia del mundo. En este sentido queremos referirnos al modo en que experimentamos, adquirimos nuevos conocimientos, deconstruimos y reconstruimos nuestros hábitos mentales.

Los factores que nos diferencian de la inteligencia artificial y nos inscriben en un marco emocional son básicamente tres: la forma en que adquirimos conocimientos, nuestra percepción consciente y nuestra agencia.

En primer lugar, cuando niños somos similares a una pizarra en blanco en la que el continuo proceso experiencial (entendido en términos corporales, ambientales y mentales) nos permite sentar las bases de nuestros hábitos mentales. No partimos, por tanto, de una base de conocimiento absolutamente preestructurada, como en el caso de los conjuntos de datos de los que disponen las máquinas inteligentes, sino que estructuramos nuestro conocimiento de forma independiente y diferente en función de factores socioculturales, familiares y ambientales, que estructuran nuestra emocionalidad, nuestro conocimiento y nuestra forma de experimentar (Mezirow, 2003). Cuando destacamos que cada uno de nosotros percibe, interpreta situaciones o se forma de una misma experiencia diferentes opiniones, estamos poniendo de relieve patrones de significado que se han estructurado siguiendo una reelaboración personal. En este sentido, el enfoque enactivo se propone como una tercera vía para las ciencias de

**Tabla 1.** Resumen de los principales sistemas conversacionales.

| Métrica | ELIZA | ALICE | XIAOICE |
|---|---|---|---|
| *Escalabilidad* | Ninguna | Los guiones pueden ser personalizados | Escalable |
| *Características principales* | Imitar el comportamiento humano en conversación | Fácil personalización de scripts (mediante AIML) | Creación de vínculos emocionales con los usuarios; conjunto escalable de competencias para la asistencia al usuario |
| *Logros e hitos* | Primer *chatbot* | Ganó tres veces el Premio Loebner | El primer *chatbot* social ampliamente implantado; 100 millones de usuarios; libro de poemas publicado; ha presentado un programa de televisión |
| *Modalidad* | Sólo texto | Sólo texto | Texto, imagen, voz |
| *Modelado* | Basado en reglas | Basado en reglas | Basado en el aprendizaje |
| *Ámbito* | Ámbito restringido | Ámbito restringido | Ámbito abierto |
| *Avance técnico clave* | Uso de secuencias de comandos; concordancia de patrones basada en palabras clave; respuesta basada en reglas | Uso de AIML y recursividad para la concordancia de patrones; se pueden asignar múltiples patrones a la misma respuesta | Modelos de inteligencia emocional para establecer vínculos emocionales con los usuarios |
| *Limitación técnica clave* | Dominio de conocimientos limitado | El tamaño del script puede ser enorme | Personalidad y respuestas incoherentes en diálogos largos |

Fuente: Shum et al. (2018), p. 14

la mente. Se diferencia tanto del cognitivismo clásico como del conexionismo y enfatiza la cognición como un saber hacer (Di Paolo & Thompson, 2017). Específicamente, la enacción se refiere a actuar y conocer cómo la mente, el cuerpo y el mundo están interconectados. Según esta teoría, la acción y el conocimiento son un proceso único y en la enacción tanto el cuerpo como la mente tienen un papel significativo (Margiotta, 2015; Pellerey, 2021). La cognición, por tanto, se percibe como un proceso y no como un cálculo. En consonancia con las teorías de la acción, el enfoque enactivo ve una fuerte conexión entre la comprensión y la acción: conocer, manipular, relacionar y transformar. La representación del mundo y de su sistema no precede a la acción, sino que se desarrolla con ella, en la relación interna con el sistema que combina a los sujetos entre sí y con el entorno (Varela et al., 1991). La vivencia, que consiste en percibir y actuar simultáneamente, es considerada una turbulencia a la que el sistema autopoiético del hombre debe tratar de encontrar equilibrio. Por lo tanto, cuando conocemos, tomamos parte en un 'bucle de retroalimentación' [feedback loop] con nuestros patrones de significado con el fin de encontrar un equilibrio (Maturana & Varela, 1988). El concepto de unidad de la persona en su acción, donde intervienen conjuntamente aspectos cognitivos, emocionales y corporales, debe estar presente en este argumento. Cuando experimentamos, mente y cuerpo se impregnan de estados emocionales, a veces incluso imperceptibles, que enriquecen la acción y el resultado de la propia experiencia (Immordino-Yang, 2017, Immorgino-Yang & Gotlieb 2017).

Anteriormente, hablamos de la estructuración del entramado mental a través del proceso de conocimiento. Pero, ¿qué son los hábitos

mentales? Pellerey (2021) sostiene que 'un hábito se desarrolla a partir de elecciones personales en una perspectiva de construcción de la propia identidad personal o profesional'. El hábito, por tanto, presupone una repetición de la acción orientada a un fin muy concreto: responder a cómo queremos situarnos en el futuro. La afirmación que acabamos de transcribir pone de relieve otros dos factores a tener en cuenta al hablar de las diferencias entre el hombre y las máquinas inteligentes: la conciencia, que nos permite percibirnos a nosotros mismos en el conjunto de nuestro pensamiento, y la agencia, que nos permite llevar a cabo acciones que tienen una finalidad bien definida.

La conciencia ha sido definida entre las diferentes metáforas como un 'teatro mental' donde el protagonista es el ego y el escenario se compone de percepción, experiencias, acciones. La construcción del ego, en este sentido, depende del escenario que vive e interpreta. La filosofía de la mente, al igual que los científicos, ha intentado por años definir el concepto de conciencia. Lo que subraya el filósofo David Chalmers es que el problema difícil [hard problem] de la conciencia es el de la experiencia. En este sentido, no se cuestiona el hecho de que la conciencia esté condicionada por hechos físicos, sino que se duda de que una descripción puramente material de la conciencia pueda ser exhaustiva. Si preguntáramos a la máquina inteligente si es consciente, lo más probable es que dijera 'por supuesto que lo soy', pero en realidad no lo sabe. Se trata de una respuesta preestablecida, o bien, se llega a ella a través de una serie de cálculos y algoritmos dictados por la continua retroalimentación con los usuarios (Chalmers, 2014).

Por último, la acción humana, mencionada tanto por el enfoque enactivo como en la explicación del concepto de conciencia, es un aspecto fundamental que diferencia al hombre de la máquina inteligente. Nussbaum (2011), en el plano filosófico, elaboró el principio de libertad asociado a la agencia del individuo (Margiotta, 2015). La acción humana se caracteriza por su intencionalidad, el fin que quiere alcanzar, el porqué y el sentido que el hombre atribuye a su comportamiento. Dar nacimiento a la intención de actuar en una determinada dirección implica la presencia de una conciencia, la interacción entre el sujeto, entendido en toda su realidad, y la percepción de la situación o de la tarea que se encuentra, y la evaluación de los objetivos futuros (Pellerey, 2021). La capacidad de emprender acciones específicas con determinados fines es, por tanto, la principal característica de la acción humana. Según Bandura (2006), los factores personales (cognitivos, afectivos y corporales), el comportamiento y las situaciones ambientales interactúan y se influyen mutuamente, de acuerdo con la teoría enactiva anteriormente expuesta. Adoptando la teoría agencial [agentic theory] de Emirbayer y Mische (1998), es posible diferenciar la agencia de la propia acción. Para los autores, de hecho, existen tres dimensiones de la agencia: (1) el elemento iterativo [iterational], la orientación selectiva por parte de los agentes de modelos pasados de pensamiento y acción, incorporados habitualmente a la actividad práctica y que dan continuidad a la identidad personal y social del individuo; (2) el elemento proyectivo, capaz de generar, por parte de los actores, posibles trayectorias futuras de acción, en las que las estructuras de pensamiento y acción pueden reconfigurarse creativamente en relación con las esperanzas, el estado emocional y los deseos de los actores para el futuro; (3) el elemento práctico-evaluativo, es decir, la capacidad de los actores para formular juicios prácticos y normativos entre posibles trayectorias de acción alternativas, en respuesta a la incertidumbre y la ambigüedad del presente (Biesta & Tedder, 2006).

La máquina inteligente, aunque intenta reproducir los procesos mentales del hombre, aún no ha adquirido la verdadera integración racional-emocional; no tiene una conciencia capaz de procesar, ni siquiera en términos emocionales, su propia experiencia y, en consecuencia, no puede estructurar una identidad personal, porque carece de tal conciencia. La identidad personal puede considerarse como el hilo conductor que guía la acción del hombre. Además, la máquina inteligente no tiene la autonomía, la libertad o la capacidad de definir su propia acción si no es dentro

de un marco indicativo derivado de sus diseñadores y de la interacción continua con los usuarios. Los distintos usuarios pueden mejorar las habilidades comunicativas de la máquina, pero ¿cómo puede ésta mejorar sus habilidades emocionales sin haber experimentado ni un solo efecto fisiológico? Podemos hablar de imitación del comportamiento social, pero carente de una orientación ética hacia el bien.

**Declaration of conflicting interests / *Declaración de conflicto de intereses***

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. / *El (Los) autor(es) declara(n) que no existen posibles conflictos de intereses. con respecto a la investigación, autoría y/o publicación de este artículo.*

**Funding / *Financiación***

The author(s) received no financial support for the research, authorship, and/or publication of this article. / *El (Los) autor(es) no recibieron apoyo financiero para la investigación, autoría y/o publicación de este artículo.*

**References / *Referencias***

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science*, *1*(2), 164–180.

Biesta, G., & Tedder, M. (2006). *How is agency possible? Towards an ecological understanding of agency-as-achievement*. Learning lives: Learning, identity, and agency in the life course (Working Paper 5). Exeter: Teaching and Learning Research Programme.

Bruno, N., Guerra, G., Alioto, B. P., & Jacomuzzi, A. C. (2023). Shareability: A novel perspective on human-media interaction. *Frontiers in Computer Science*, *5*, 1106322.

Chalmers, D. (2014). *Che cos' è la coscienza*. LIT EDIZIONI.

Colby, K. (1972). Simulation of psychotic processes. *Science*, *176*, 1192–1194.

Colombetti, G., & Thompson, E. (2008). Il corpo e il vissuto affettivo: verso un approccio «enattivo» allo studio delle emozioni. *Rivista di estetica*, *37*, 77–96.

Di Paolo, E., & Thompson, E. (2017). *The enactive approach*. Routledge.

Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology*, *103*(4), 962–1023.

Gardner, H. (1987) Beyond IQ: Education and human development. Harvard Educational Review.

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Goleman, D. (1995). Emotional Intelligence: Why It Can Matter More Than IQ. Bantam Books.

Immordino-Yang, M. H. (2017). *Neuroscienze affettive ed educazione*. Raffaello Cortina Editore.

Immordino-Yang, M. H., & Gotlieb, R. (2017). Embodied brains, social minds, cultural meaning: Integrating neuroscientific and educational research on social-affective development. *American Educational Research Journal*, *54*(1_Suppl.), 344S–367S.

Jacomuzzi, A. C. (2023). *Introduzione alle scienze cognitive*. Il Mulino.

Kemp, S. (2022). Digital 2022 Global Overview Report. https://wearesocial.com/digital-2022-global-overview-report

Mancini, T., & Riva, G. (2023). *Psicologia dei media digitali*. Il Mulino.

Margiotta, U. (2015). *Teoria della formazione. Ricostruire la pedagogia* (*Vol. 993*, pp. 1–283). Carocci.

Maslow, A. H. (1943). Preface to motivation theory. *Psychosomatic Medicine*, *5*(1), 85–92.

Maturana, R., & Varela, F. (1988). *Autopoiesi e cognizione*. Marsilio.

McLeod, S. (2007). Maslow's hierarchy of needs. *Simply Psychology*, 1–18.

McStay, A. (2023). Replika in the Metaverse: The moral problem with empathy in 'It from Bit'. *AI and Ethics*, *3*, 1433–1445.

Mezirow, J. (2003). Transformative learning as discourse. *Journal of Transformative Education*, *1*(1), 58–63.

Milani, L., & Jacomuzzi, A. (2022). Interactions and social identity of support teachers: An ethnographic study of the marginalisation in the inclusive school. *Frontiers in Education*, *7*, 948202.

Milani, L., Pezua Sanjinez, J., & Jacomuzzi, A. (2021). Insects as food: Knowledge, desire and media credibility. Ideas for a communication. *Rivista di studi sulla sostenibilita'*, 2, 385–396. https://doi.org/10.3280/RISS2021-002025

Nussbaum, M. C. (2011). *Creating capabilities: The human development approach*. Harvard University Press.

Pellerey, M. (2021). *L'identità professionale oggi: natura e costruzione*. FrancoAngeli.

Shum, H. Y., He, X. D., & Li, D. (2018). From Eliza to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, *19*, 10–26.

Sprevak, M. (2017). Turing's model of the mind. In J. Copeland, J. Bowen, M. Sprevak, & R. Wilson (Eds.), *The Turing guide: Life, work, legacy* (pp. 277–285). Oxford University Press.

Turing, A. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *42*, 230–265.

Turing, A. (1950). Computing machinery and intelligence. *Mind*, *LIX*, 433–460.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.

Wallace, R. (1995). *A.L.I.C.E.: Artificial Linguistic Internet Computer Entity*. https://www.chatbots.org/chatbot/a.l.i.c.e/

Weizenbaum, J. (1966). ELIZA - A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 230–265.

Zemčík, M. T. (2019). A brief history of chatbots. *DEStech Transactions on Computer Science and Engineering*, *10*.

Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2020). The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, *46*(1), 53–93.