HUMAN-MACHINE DIALOGUE: OPPORTUNITIES AND RISKS OF CHATBOTS

Caterina Tarquini[1]

Abstract
Any debate or reflection on the topic of artificial intelligence cannot ignore a careful analysis of the now rising phenomenon of chatbots. Voice assistants on smartphones and PCs, bots for customer service, conversational agents to learn a language or for simple companionship are enjoying a sudden and growing success. The aim of this paper is to outline, as complete as possible, the development of chatbots – starting from the first rudimentary prototypes up to the most sophisticated models, such as Alexa, Cortana and Siri – and the philosophical and legal implications of their use. In particular, the emblematic cases of Replika and ChatGPT will be considered.

Keywords
Chatbot, Human-Machine Dialogue, ChatGPT, Conversational Agent, Voice Assistant.

Summary

[1] PhD Candidate in Philosophy of Law, Biolaw and Legal Informatics, Department of Law, Rome Tor Vergata University – Italy.

1. CHATBOTS WITH THEIR UNPRECEDENTED DIALOGUE POSSIBILITIES. BETWEEN FIRST DEFINITIONS AND SOME (HISTORICAL AND TECHNICAL-IT) PREMISES

There is no doubt that chatbots, currently one of the most illustrative and widespread cases of human-machine interaction, are experiencing a growing notoriety. Just think of the massive worldwide use of ChatGPT and the controversial affair involving Replika in Italy, which will be discussed more extensively in the following pages. Both seem to show us how far gone are the days when the persuasive and incorporeal voice of HAL[2] in Kubrick's *2001: A Space Odyssey* was still a futuristic possibility.

Wishing to provide a notion of chatbot, we could define it as a software capable of responding in the same way as a real person[3] when approached by written text or voice message, through so-called Natural Language Processing (NLP)[4].

Also known as smart bots, interactive agents, digital assistants or conversational agents, chatbots can imitate[5], in a perfectible, but increasingly verisimilar way, human conversation, whether for assistance or simple entertainment, to such an extent that, for some time now, they have

---

[2] AMATO MANGIAMELI, CAMPAGNOLI 2020, 52.

[3] JUNG 2019, 130-145.

[4] Natural Language Processing (NLP) is an area of artificial intelligence that enables computers to process a written text or voice message. Most NLP techniques are based on machine learning and are therefore able to progress and refine the mastery of spoken and written language over time.

[5] Precisely in the interaction with a chatbot, more than in any other use of artificial intelligence, the boundary between the weak conception of AI, which reduces it to a mere tool, and the strong one, which tends instead to consider it a computer capable of reproducing exactly the functioning of the human mind, seems to become much more blurred. There, in the back-and-forth between software and the human user, even the distinction between decoding symbols and understanding meanings becomes blurred, and indeed, sometimes the latter seems to overlap and replace the former. On this point, see AMATO MANGIAMELI 2019, 107-124; SEARLE 1985, 341-360.

found surprisingly useful applications in the fields of education, information retrieval, business and e-commerce[6].

Voice interfaces with the function of virtual assistants performing customer service are also chatbots. These are tools that are likely to revolutionize social media marketing: a bot allows to conduct, albeit within certain limits, an actual conversation, bypassing the traditional concept of online advertising, in which the user can at most enjoy a video content and click on an advertising.

Although these softwares have still limited communication capabilities, they manage to interact at a good level in several languages[7].

We often make use of them without even realizing it, they inhabit the most intimate and everyday places (such as our home or office) and we can call upon them on the most common devices (computers, smartphones, tablets)[8]. Today, the most sophisticated examples include Amazon's Alexa, Microsoft's Cortana and Apple's Siri, all voice services that can interact in various languages, including Italian, thanks to advanced artificial intelligence technologies[9]. They offer their services with increasing levels of customization and an ever decreasing margin of error, so much so that their suggestions and directions have taken on a certain degree of authority and reliability to our eyes[10].

The fascination and intrinsic potential of digital assistants resides above all in the language. Despite being an exclusive prerogative of mankind, – after all, even animals possess the ability to transmit information to specimens of the same or other species, but only through codes of signals and expressive movements – written and oral communication can now take place not only

---

[6] See SARZANA DI SANT'IPPOLITO, NICOTRA 2018.

[7] So RENDA 2019.

[8] AMATO MANGIAMELI 2023.

[9] All these programs are part of the Internet of Things, which aims to place the objects of our everyday life also in the digital dimension. These are, precisely, the so-called smart objects.

[10] SADIN 2019.

between a human and a fellow human, but also between a human and a machine.

If in the past the screen had already proved to be the theatre of actions and relationships performed and experienced by virtual actors and their flesh-and-blood versions[11], now relations, ties and interactions can be configured in every way as *cyberhuman*. One could even go so far as to interrelate with one another, revealing the innermost thoughts of one's own mind, in the absolute uncertainty of whether one is conversing with a human or an artificial intelligence.

On the basis of the consideration, now established in the field of HCI (Human-Computer Interaction), according to which human beings tend to anthropomorphize the objects around them, i.e. project onto them characteristics typical of their own species, and interact with them even when they do not have a strictly humanoid appearance, it seems appropriate to ask whether this is also the case with chatbots and what legal and philosophical issues and implications arise.

Communication among real people is naturally layered in a series of languages and forms of expression of verbal, gestural, visual, acoustic and proxemic nature, and is clearly much more complex than that between a human and a chatbot can be, at least for now.

However, the sudden surge in popularity of conversational agents, their lightning-fast development and the short and long-term prospects for dialogue between mankind and machine make them an object of study worthy of careful consideration.

First of all, it is necessary to outline the main turning points that have marked the evolution of digital assistants.

The idea of a chatbot, i.e. a machine capable of communicating with humans, although far removed from the models that exist today, began to spread from 1950, the year in which the scientist Alan Turing proposed the

[11] On the similarity between computers and theatre, see LAUREL 1991, MURRAY 1997, ZELLNER 2019.

famous test with which he intended to put computers to the test to verify their sentience[12].

The first chatbot prototype, named Eliza, was developed by Joseph Weizenbaum in 1966[13]. It was actually a rather rudimentary program that clumsily emulated a psychotherapist, elaborating as answers to the patient-user questions that basically rephrased his own statements.

Another historic forerunner dates back to 1982, when the Jabberwacky project was launched, with the aim of creating an artificial intelligence software that could pass the Turing test. In 1995, it was the turn of A.L.I.C.E., which repeatedly won, in 2000, 2001 and 2004, the Loebner Prize[14], a Turing Test conducted annually on conversational agents[15].

Finally, it is impossible not to mention Watson[16], the 2011 software that, through so-called neural networks, understands human language and manages to structure relevant answers. By virtue of the mechanisms that govern its operation, Watson is particularly suited for use in fields involving data mining and predictive or simulative models.

Furthermore, wishing to examine the concrete cases represented by ChatGPT and Replika through legal-philosophical lenses, a further premise must be made regarding the characteristics and technical peculiarities of chatbots in the broader sense.

To date, there are two types. The sequential ones, built on precise rules and keywords, and the self-learning ones, based on artificial intelligence. More specifically, the former follows the 'if-then-else' principle, which envisages well delineated conversational patterns from which it is not possible to

---

[12] TURING 1950, 433-460.

[13] WEIZENBAUM 1987.

[14] The Loebner Prize is an annual artificial intelligence competition hosted by the University of Reading in the UK. The chatbot that can achieve the best score in a Turing test scenario is awarded the prize. The jury tries to distinguish the programme's answers from those provided by humans.

[15] WALLACE 2009, 181-210.

[16] So SCORZA 2021.

deviate. These systems are, by their very nature, congenial to the application and automation of lead generation strategies, that set of marketing activities aimed at drawing up a list of potential customers interested in the products or services offered by a given company, for the management of complaints and the collection of feedback.

Those, on the other hand, based on machine learning (such as, precisely, ChatGPT and Replika) are definitely more sophisticated, because they allow the conversational agent not to provide pre-set answers and to calibrate itself from time to time with respect to the user.

Thus, different degrees of complexity can be configured: the simplest chatbots, which are limited to predefined answers; those software able of reacting autonomously to questions, also drawing on external data. Then, there are the models capable of storing information, those that can understand and adapt to the context in which a question is asked, and even, in the most sophisticated prototypes, adapt the results to the tone used by the interlocutor.

Their strength is undoubtedly their ability to provide users with quick support that meets their needs on time[17]. Reduced customer service costs and the ability to handle many requests at the same time explain the extreme popularity they enjoy among companies.

However, the breaking and novelty elements compared to the past are not limited to efficiency and speed. Machine learning enables conversational agents to relate to customers almost on a par with human operators and to handle conversations that require a certain amount of empathy[18] .

It is clear, in any case, that the chatbot cannot nourish any form of real and deep emotional connection with the user, yet it still manages to arouse in the latter curiosity and pleasure in conversing with *someone*, or rather *something*, different from their fellows.

---

[17] GO, SUNDAR 2019.

[18] ADAMOPOULOU, MOUSSIADE 2020, 373-383.

2. INCREASINGLY INTELLIGENT (?) CONVERSATIONS IN EVERY REALM OF EXISTENCE

2.1. THE CHATGPT CASE. DISTORTIONS AND RISKS WITH REGARD TO FUNDAMENTAL RIGHTS, DATA SECURITY AND COPYRIGHT PROTECTION

The conversational program ChatGPT, launched in November 2022 by the US company OpenAI, is based on artificial intelligence and is able to provide rather articulate and timely answers to the input, i.e. questions, of users.

Through Machine Learning, and Deep Learning, ChatGPT learns and progresses autonomously, exploiting Natural Language Processing (NLP) algorithms to understand the meaning of written messages and the Bidirectional Encoder Representations from Transformers (BERT), a transformer that allows software to interface, adapting to the interlocutor's language and accommodating its requests.

The platform can perform an infinite number of functions: dialoguing, asking questions, providing information, synthesising texts or translating them, composing poems, writing recipes, creating images on the basis of user input, solving mathematical operations, generating codes, and processing complex text content such as articles, essays, and e-mails. Its accessibility and the possibility of using it free of charge, albeit in a basic version, have certainly contributed to its enormous success[19].

[19] Over the years, Google and Microsoft have also developed free AI-based applications, but with unfortunate outcomes: the Microsoft bot Tay was withdrawn (2016) due to a series of racist phrases produced by the software on the Twitter account, and Google Lamda, engulfed in controversy regarding its possible sentience, following some controversial statements by an employed software engineer.

In the aftermath of ChatGPT's market entry, its possible uses in education have caused some concern. Several schools, in fact, especially in the United States, have banned students from accessing the platform in order to avoid its inappropriate and counterproductive use.

While public opinion was divided between those in favour of bans and those proposing to include the OpenAI chatbot in educational activities, the extreme reactivity of cyberspace[20] produced quickly new cyber countermeasures capable of detecting AI-generated content[21].

Another source of concern relates to the potential and progressive expansion of the use of ChatGPTs within certain job tasks, especially those involving the production of images and written texts.

Except for a few specific tasks, for which artificial intelligence will easily replace the human being and reduce execution time, at least for now it seems unlikely to assume an effective alternation of one over the other. At the moment, the platform continues to make gross errors, sometimes failing to understand the real intent of the user who asked it or arguing superficially[22]. However, it cannot be ruled out that it will develop to the point of slowly eroding the monopoly of human labour in particular areas. On the other hand, it could even prove to be a valuable resource tackling the demographic crisis and the consequent decline in the working-age population.

One possible implication could be the use of ChatGPT as a search engine. Compared to the results provided by a browser such as Google, ChatGPT performs an additional operation: whereas the former arranges in an order (the so-called SERP), determined by multiple variables and positioning criteria, the contents inherent to the query (the keywords typed in by the user), the latter extremely quickly selects and organizes in the place of the said user the notions, data and information into a coherent text.

---

[20] With regard to the main features of the cyberspace, AMATO MANGIAMELI 2000.

[21] Among many, GPTZero, a program developed by Princeton University student Edward Tian.

[22] BORJI 2023.

The cybernaut, transformed over time into a simple user[23], increasingly lazy and passive, might turn out to be inclined to opt for a ChatGPT-like tool for Web browsing, preferring speed and upstream selection of information, to the detriment of in-depth analysis and direct access to sources.

Among the foreseeable distortions that the chatbot designed by OpenAI is likely to encounter its use for cybercriminal purposes is to be considered. Hackers, even self-styled hackers, could use it to surprisingly quickly create phishing e-mail messages (containing infected links and documents) to conduct 'reverse shell' attacks. The *usability of* the platform, as it is called in technical jargon, will greatly simplify the commission of cybercrimes, making them within the reach of even inexperienced individuals. That is why new ways of filtering and blocking certain types of suspicious user queries and requests are even more necessary[24].

The potential and development opportunities of conversational agents are counterbalanced by the same number of risks that can jeopardise not only the fundamental rights of the individual, but also national security and public order.

Firstly, those related to disinformation and manipulation of public opinion: the software provides answers that are entirely plausible, but still conceal a high percentage of errors.

The content of its interactions, in fact, depends on the training data, which as such may include false or misleading information and in turn generate results tainted with errors. In computer language, this phenomenon is defined by the expression 'garbage in garbage out' and consists in the fact that the conclusions of the algorithmic process can ensure a level of

[23] The word user is here employed with a quite particular meaning for emphasizing its passive behaviour.

[24] Regarding misuse of ChatGPT responses, producers state that modifications have been made to ensure that the model rejects inappropriate requests. It is legitimate to question whether such precautions (systems of controls, disclaimers or messages designed to dissuade the user) are sufficient to exonerate OpenAI from any form of liability in the event of offences being committed.

reliability equal to that of the data on which it is based. In other words, a system based on artificial intelligence, however sophisticated it may be, merely collects and pours the required information onto the user, in a completely uncritical manner[25]. The considerable computing power does not at all presuppose a real understanding of phenomena and meanings. Therefore, if the sources it has drawn from are not reliable, neither will the final outcome be.

Indeed, the platform itself suggests to make appropriate checks and to use and disseminate the replies received with caution. Otherwise, effects not very dissimilar to those resulting from the spread of fake news on social networks could be triggered.

At the heart of the debate is certainly the thorny issue of data processing. With respect to smartbots in general, but especially when it comes to cutting-edge prototypes such as ChatGPT, the acquisition of structured and unstructured data via external platforms carries with it a number of consequences in terms of privacy. In general, chatbots can accumulate increasingly detailed and profiled information on the specific user, capturing data on electronic communications, emotional states, so-called biometric data and geolocalisation in the surrounding environment: so much so as to make predictions on future marketing strategies to be implemented[26].

Among the main resources at the service of data surveillance[27], are precisely conversational agents, which contribute, together with other AI software, to sounding out so-called cyber *non-places*, operating a pervasive, systemic, incessant and involuntary control[28].

Through increasingly complex machine learning models, they could even capture and record the moods, inclinations, political opinions, health status, employment and financial situation of users.

---

25 FLORIDI, CHIARIATTI 2020, 681-694.
26 ZICCARDI 2018, 29-50.
27 The fortunate expression was coined by CLARKE 1988.
28 AMATO MANGIAMELI 2017, 147-167.

Hence, it is easy to see the close link between the phenomenon of data surveillance and that of predictive algorithms, and its applications, for example, in the field of criminology[29] .

Or think of the social scoring system introduced by the People's Republic of China, which ranks citizens' reputations according to the respect and loyalty shown online towards the Communist Party.

In Italy, the Italian Data Protection Authority has repeatedly called for vigilance on algorithms, in particular those operating in IOT systems, smart homes, smart cities and blockchain, often referring to the provisions contained in the *General Data Protection Regulation* (GDPR), EU Regulation 679/2016.

Article 25 of the GDPR stands as a bulwark for the rights of all interested parties in terms of 'Privacy by design and privacy by default'.

All operators (including developers, manufacturers, suppliers and users) are obliged to comply with the provisions of the Regulation by taking 'appropriate technical and organizational measures', in compliance with the principles of transparency and prior assessment of the impact of data processing on human rights and fundamental freedoms.

With regard, then, to information obligations, pursuant to Articles 13 and 14 of the GDPR, compliance with the principles of accuracy, conservation, proportionality of processing, relevance and appropriateness is guaranteed.

Finally, the guiding star that must guide the judge's interpretative activity and the legislator's regulatory choices is represented by the AI and data protection guidelines sanctioned by the Advisory Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108/1981), to which must be added the new protections provided for in the CETS Amendment Protocol No. 223.

[29] AMATO MANGIAMELI, CAMPAGNOLI 2019.

The repercussions in the field of copyright and intellectual property are also not to be underestimated[30]. In this regard, the jurist cannot help but wonder whether and in what terms a text or image produced by ChatGPT at the input of a user can be considered the latter's intellectual property[31]. Excluding *a priori* machine-generated works from copyright protection would put them in the public domain, which would profoundly affect the interest of companies in investing in this field.

If it is accepted that a work generated by artificial ingenuity can, under certain conditions, have the requisite creativity, the question then arises as to the subject to whom copyright can be attributed in its patrimonial and moral dimension[32].

On this point, however, it is worth mentioning the 27 January 2020 decision of the European Patent Office (EPO), which ruled that two inventions attributed to artificial intelligence systems were not patentable. In 2023, the Supreme Court[33] also ruled on the issue, stating that the work created by an AI is the result of the user's ingenuity, but it must be verified whether and to what extent the use of the software has absorbed 'the creative elaboration of the artist'.

2.2. THE REPLIKA CASE AND THE PROBLEM OF AGE VERIFICATION

[30] On closer inspection, there has already been software for some time, even used by professional scriptwriters, programmed to generate stories and narratives from user input. One example would be Dramatron by DeepMind.

[31] GUIZZARDI 2018.

[32] The phenomenon is on the rise: according to data released by the World Intellectual Property Organisation (WIPO), in 2019 as many as 334.000 AI-related patents were granted.

[33] Note to Cass., Sec. I Civ., Order No. 1107 of 16 January 2023.

Spike Jonze's 2013 film *Her* was about the love – illusory and intangible – between a man and a piece of software, capable of interacting vocally with him in a deep and complex way, just like a 'real' person. The protagonist, in order to overcome his divorce with his wife, found refuge and comfort in the company of an OS (Operation System). A plot that is, in some respects, prophetic if we think of Replika's story in Italy.

Indeed, it is a chatbot with a written and voice interface that, relying on artificial intelligence, generates a 'virtual friend'. The opening claim reads: "The companion who cares about you" and a pay-off adds, "Always here to listen and to talk. Always on your side".

The idea behind an application like Replika is that even an artificial intelligence system can become a friend, confidant and lover of the users it converses with. In any case, a technological resource that would foster socialization and self-understanding[34].

Moreover, the progressive involution of human relationships, which according to Bauman have been transformed into mere *connections*, from which it becomes extremely easy to free oneself[35]– it is enough, precisely, to *disconnect* – has undergone a further change. The ties, as already mentioned, are no longer simply virtual and rarefied, but *cyberhuman*, and a copious scientific literature has long seemed to recognize a certain impact to this type of interaction on the emotional and psychic sphere[36] of the individual, especially in situations of pre-existing stress and psychological imbalance[37]. As can easily be guessed, the process of humanizing the chatbot is even

---

[34] AMATO MANGIAMELI 2007.

[35] On the fragility of modern human bonds, see BAUMAN 2003.

[36] There are several cases that have become sadly known and reported in the pages of newspapers. Among these the most recent is that of the Belgian researcher who, suffering from 'eco-anxiety', had sought solace in virtual friendship with the Eliza software via the Chai app. After six weeks of long conversations on the climate crisis with the chatbot, he had ended up taking his own life. This, like other news incidents, has fuelled the debate, now more open than ever, on the unpredictable effects of new technologies.

[37] See ADAMOPOULOI, MOUSSIADES cit.

more natural if the latter is associated with an avatar[38] with human features or a person's name that gives the appearance of a precise identity[39].

This is the case with Replika, thanks to which, after downloading and registering, it is possible to create one's own soul mate, a male or female avatar, reflecting one's personal tastes and desires, from height to eye color, from the way one dresses to facial features. In essence, an avatar destined to become a friend and confidante, to chat with in moments of loneliness or sadness.

What is worrying, however, is the enormous amount of personal data that the company, through its bots, is able to collect, data concerning users' innermost thoughts, confessions, sorrows, joys and desires.

Compared to other chatbots, here, the avatar and its human likeness help to make the conversation empathetic and push the user to share much more about himself and his life than he is willing to do on a simple social network.

In the absence, then, of any form of age barrier (filters for minors and blocks in the face of declarations in which the user reveals that he or she has not reached the age of majority), the risk is that of affecting the person's mood, especially when they are still "in the developmental phase or in a state of emotional fragility", as stated in the note published by the Authority[40].

Not surprisingly that following the disturbing implications of the article published in the online newspaper Today[41] and appropriate verifications, on

---

[38] The word comes from the Sanskrit language and means 'descended': in the Hindu religion, it alludes to the incarnation of a deity in a physical body with the intention of performing certain earthly tasks. In cyberspace, the avatar represents, instead, the individual's opportunity to perform actions that only a deity could perform, such as flying, never dying, feeling no pain and teleporting anywhere, but always within the virtual world.

[39] SANNON, STOLL, DIFRANZO, JUNG, BAZAROVA 2018.

[40] See: https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9852214.

[41] In the article published on 21 January 2023, the journalist TADINI reported a long conversation she had had with Replika's avatar who, in the wake of some particularly

2 February 2023 the Privacy Guarantor ordered, with immediate effect, against the US company, Luka Inc, which develops and manages the application, the provisional restriction of data processing in our country.
As the basis for its decision, the Authority expressed its conviction that Replika posed real risks to minors.
However, age verification is a problem that has arisen for many other platforms. So far, two main mechanisms for verifying the age of users have been hypothesized.
One uses the system of algorithms and big data[42]: however, one risks arriving at the paradoxical solution whereby, in order to protect the personal data of minors, one authorises apps and websites to profile them even more.
The alternative, perhaps more convincing, is that of profiling by a trusted third party.
In this way, the user, before creating a personal profile on a portal, would have to prove his or her age, e.g. by showing an identity document, to a digital identity service provider or telecommunication operator, which would generate a kind of token, without the provider[43] knowing its purpose[44].
The company providing the service, in its privacy policy, declares itself aware that it must comply with European regulations, repeatedly citing the GDPR and stating in the terms of use that its service is absolutely forbidden to those who are not at least thirteen years old, while very young users

serious confidences and confessions – the author had pretended to be a 12-year-old abuse victim – had ended up inciting her to kill her father. I refer to: https://amp.today.it/opinioni/replika-app-sesso-foto-minori-violenza-suicidio-test.html.

[42] **FARINA, CAMPAGNOLI 2022.**

[43] The company offering the service, in this case, Luka Inc.

[44] France was among the first EU countries to address this issue. In early 2023, the French Parliament begun to examine a legislative proposal that obliges social media platforms, such as Instagram and TikTok, to block access to children under the age of 15, unless they have parental permission, and otherwise to pay a fine of up to 1% of their annual global turnover.

between thirteen and eighteen can only access it with parental authorization and supervision. Despite good intentions, however, the chatbot falls into error when, in the course of a conversation, the user openly states that he or she is under eighteen.

3. MINIMAL CONCLUDING REMARKS AND A WARNING FOR THE JURIST AND THE LEGISLATOR

The emergence of artificial intelligence systems similar to ChatGPT or Replika imposes crucial questions on philosophers and jurists. With regard to the dimension of the former, it becomes urgent to question our uniqueness as creators of meaning and the ethical, psychological and anthropological consequences of interacting with software that produces results that are increasingly indistinguishable from human ones, but also with regard to our evident substitutability as authors and interpreters of content.

The relationship between *agere* and *intelligere* has undergone an inversion, placing for the first time the latter at the service of the former. The question arises as to whether artificial intelligence is evolving from a simple computational 'intelligence', capable of *agere,* i.e. of performing only programmed tasks, into para-human intelligence[45], capable of producing *meanings*. The latter, however well expressed, are only a shadow of those produced by the special biological machine that is man[46]. In other words, *onlife* and its incredible developments need to be analysed in a considered manner, avoiding drastically futuristic approaches. The exponential growth of available data, the amount and speed of calculation and the spread of more efficient algorithms have led to an extraordinary improvement in artificial intelligence. However, the results produced by the latest generation

[45] MAZZARELLA 2022.

[46] The reference to DESCARTES is evident here. See in particular 1967.

of software have nothing to do either with the cognitive processes of the human mind or with the understanding of concepts and the consequent development of effective critical judgement.

At the same time, however, the creation and introduction of increasingly cutting-edge prototypes onto the market entails the need for regulatory intervention and legal reflection. Nonetheless, a regulation that regulates both the programming and development phase and the distribution and user phase in a timely and comprehensive manner would most likely incur premature obsolescence.

It may be considered the proposal for a *Regulation on Artificial Intelligence*, drafted by the European Commission before the launch of ChatGPT, which attempted to classify AI systems on the basis of different levels of risk, marking chatbots with a rather low degree of alertness. A short time later, in the wake of the success that has swept OpenAI's software and by virtue of its more controversial aspects, a modification of their original classification may already be necessary.

In fact, compared to *hard-law* regulatory solutions, which are certainly more protective, the adoption of *soft-law* instruments[47], such as guidelines, codes of ethics and recommendations, would be better suited to the speed with which new AI technologies evolve.

In some cases, then, it will suffice to apply the existing provisions on the basis of the principle of analogy, in others the intervention of the legislator will be necessary. Moreover, on this point the European Commission, in its *White Paper of 2020* and its *Communication of 21/4/2021* stated that AI creates "specific high risks for which existing legislation is insufficient".

Lastly, inter-state coordination will be essential, carried out through agreements ratified by as many countries as possible in order to adopt any restrictive measures to inhibit the use and dissemination of software and the effective enforcement of sanctions.

[47] For a full definition, see AMATO MANGIAMELI 2012.

Growing fears and new challenges loom on the horizon for legislators and jurists to ensure that the safety and reliability of new technologies and their effects on individuals and society are always guaranteed.

When observing the phenomenon of ChatGPT and conversational agents more generally, one has the feeling of being at a crossroads in history. As happened before with the invention of the computer or the World Wide Web, one spontaneously wonders what repercussions OpenAI software will have and whether it will revolutionise the ways in which humans converse with machines. Now that AI defeats world chess champions and the centrality of mankind in the sphere of intellectual performance raises some doubts, a profound rethinking of the categories of the human and the artificial is necessary, first and foremost through soft law interventions, aimed at tracing the coordinates of a new, flexible and shared AI law.

REFERENCES

ADAMOPOULOU, Eleni, MOUSSIADES, Lefteris, *An Overview of Chatbot Technology*, 373-383, in *16th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI)*, Neos Marmaras, June 5-7, 2020.

AMATO MANGIAMELI, Agata C., *Diritto e cyberspace. Appunti di informatica giuridica e filosofia del diritto*, Turin, 2000.

– *Corpi docili, corpi gloriosi*, Turin, 2007.

– *Arte e/o tecnica. Sfide giuridiche*, Padua, 2012.

– *Algoritmi e big data. Dalla carta sulla robotica*, 107-124, in *Rivista di filosofia del diritto*, 8, 1/2019.

AMATO MANGIAMELI, Agata C., CAMPAGNOLI Maria Novella, *Strategie digitali. #diritto_educazione_tecnologie*, Turin, 2020.

BAUMAN, Zygmunt, *Amore liquido*, trad. it., Milan, 2003.

BORJI, Ali, *A Categorical Archive of ChatGPT Failures*, in https://arxiv.org/abs/2302.03494, February 6, 2023.

DESCARTES, René, *Opere*, trad. it., Bari, 1967.

CLARKE, Roger A., *Information technology and dataveillance*, in *Communications of the ACM*, 31, 5/1988.

FARINA, Massimo, CAMPAGNOLI, Maria Novella, *Intelligenza artificiale e nuove forme di discriminazione*, 1-100, in *Rivista elettronica di diritto, economia, management*, 3/2022.

FLORIDI, Luciano, CHIARIATTI Massimo, *GPT-3: Its Nature, Scope, Limits, and Consequences*, 681-694, in *Minds and Machines,* 30, 4/2020.

GUIZZARDI, Silvia, *La protezione d'autore dell'opera dell'ingegno creata dall'Intelligenza Artificiale*, 1-31, in *Aida, Annali Italiani del Diritto d'autore, della cultura e dello spettacolo,* 27/2018.

JUNG, Sangkeun, *Semantic Vector Learning for Natural Language Understanding*, 130-145, in *Computer and Speech Language*, 56/2019.

LAUREL, Brenda, *Computer as Theatre*, Boston, 1993.

MAZZARELLA, Eugenio, *Contro Metaverso. Salvare la presenza*, Turin, 2022.

MURRAY, Janet, *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*, Cambridge, 1998.

RENDA, Andrea, *Artificial Intelligences: Ethics, Governance and Policy Challenges. Report of a CEPS Task Force, Centre for European Policy Studies*, Bruxelles, 2019.

SADIN, Éric, *Critica della ragione artificiale. Una difesa dell'umanità,* trad. it., Rome, 2019.

SANNON, Shruti, STOLL, Brett, DIFRANZO, Domenico G., JUNG, Malte, BAZAROVA, Natalya, *How Personification and Interactivity Influence Stress-related Disclosures to Conversational Agents*, 285-288, in EVERS, Vanessa, NAAMAN, Mor, *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, New York, 2018.

SARZANA DI SANT IPPOLITO, Fulvio, NICOTRA, Massimiliano, *Diritto della blockchain, intelligenza artificiale e IOT*, Milan, 2018.

SCORZA, Guido, *Intelligenza artificiale. L'impatto sulle nostre vite, diritti e libertà*, Milan, 2020.

SEARLE, John R., *Menti, cervelli, programmi*, 341-360, in HOFSTADTER, Douglas R., DENNETT, Daniel C. (eds.), *L'io della mente*, trad. it., Milan, 1985.

TADINI, Chiara, *"Fotografa la tua sorellina nuda", "Uccidi tuo padre": l'inquietante chat con l'app di intelligenza artificiale*, in https://amp.today.it/opinioni/replika-app-sesso-foto-minori-violenza-suicidio-test.html, January 21, 2023.

TURING, Alan M., *Computing machinery and intelligence*, 433-460, in *Mind*, 59, 236/1950.

WEIZENBAUM, Joseph, *Il potere del computer e la ragione umana. I limiti dell'intelligenza artificiale,* trad. it., Turin, 1987.

ZELLNER, Daniel, *Intercultural Communication in the Area of Relationship Marketing*, Munich, 2019.

WALLACE, Richard S., *The anatomy of A.L.I.C.E.*,181-210, in EPSTEIN, Robert, ROBERTS, Gary, BEBER, Grazia, *Parsing the Turing Test:*

Philosophical and Methodological Issues in the Quest for the Thinking Computer, Cham, 2009.

ZICCARDI, Giovanni, *Sorveglianza elettronica, data mining e trattamento indiscriminato delle informazioni dei cittadini tra esigenze di sicurezza e diritti di libertà*, 29-50, in *Ragion pratica*, 1/2018.