ARTIFICIAL INTELLIGENCE AND MENTAL HEALTH

Guido Saraceni[1]

Abstract

Nowadays, artificial intelligence represents the quintessential 'urgent and necessary' theme. To a great extent, this is due to the fact that the advent of new and powerful computer tools – I am thinking particularly of machine learning GAN networks – has unveiled the naivety of many ancient philosophical prejudices, prompting scholars to re-engage in discussions concerning the potential, characteristics, and associated risks of artificial intelligence. The debate has thus become highly engaging, replete with provocations and interdisciplinary exchanges. This article aims to contribute to this context by approaching the issue from a somewhat original perspective: exploring the intricate relationship between artificial intelligence and mental health, considered as a fundamental human right. To this end, after some brief theoretical and historical premises, the text examines the epidemiological impact of AI. Subsequently, it lists its potential in the realm of care and prevention, concluding with reflections on the legal issues and challenges posed by its usage.

Keywords

Artificial Intelligence, Mental Health, Machine Learning, Legal Informatics, Philosophy of Technology.

Summary

0. Introduction. - 1. An intricate theoretical knot. - 1.1. A brief historical excursion: from the chess player to ChatGPT. - 2. The internet addiction disorder. - 2.1. The attentive resources. - 3. AI in the service of mental health. - 4. Conclusion: challenges and issues for the law.

---

[1] Associate Professor of Philosophy of Law and Legal Informatics, Department of Law, University of Teramo – Italy.

0. INTRODUCTION

In recent years, artificial intelligence has become the quintessential 'urgent and necessary' theme, an issue that one truly cannot avoid discussing. Scholars from various disciplines – such as philosophers, sociologists, jurists, psychologists, and computer scientists – have engaged with it, each with different perspectives, objectives, styles, and objectively diverse outcomes. The bibliography o has rapidly and exponentially expanded, becoming truly extensive. Yet, it is a considerably ancient topic whose origins, as we will have the opportunity to consider, date back at least to XVIII century.

The reasons underlying this renewed and increasingly widespread interest are as diverse as the epistemological approaches from which the topic has been pondered. To a small extent, the current, more intense and participatory debate can be attributed to the fact that scientific research, like any other human activity, succumbs to the allure of trends and eras – perhaps initiated by an enlightening essay, a fortunate monograph, or a simple motto that succinctly captures and describes the Zeitgeist perfectly, emphasizing an aspect of fundamental and unexplored importance[2].

However, in this specific case, the viral scope of the topic is coupled with the significant intellectual provocation posed by the latest progress achieved by computer systems. As often happens, technology has broadly preceded philosophical speculation, demanding clear answers from it. Or, perhaps even more fittingly, demanding new doubts.

In the pages that follow, I will attempt to delve into the subject of artificial intelligence in relation to the issue of mental health, considering it as a human right. After establishing certain essential definitional premises, I will critique the etiological impact of artificial intelligence. Subsequently, I will examine its value in terms of health generation, prevention, and treatment,

---

[2] In recent years, it has occurred, for example, with the term "glocalization"; Zygmunt Bauman's "liquidity" or Marc Augé's "non-places".

finally focusing on the challenges that the field of law faces today.

1. AN INTRICATE THEORETICAL KNOT

The discussion that has developed in recent years regarding artificial intelligence is highly engaged and equally intriguing; however, it also appears to be structurally and inevitably 'flawed' due to an inescapable and intricate theoretical knot: the majority of scholars seem to begin from the assumption that the definition of intelligence is obvious.

Consequently, they often take the liberty to discuss various hypotheses concerning artificial intelligence without demonstrating a proper awareness of the immense complexity of the psychological construct named intelligence. Unfortunately (or fortunately), there is no singular, clear, and objective notion of intelligence. Philosophical, pedagogical, and psychological research offers infinite ideas and suggestions. This definitional challenge cannot help but reverberate onto the concept of artificial intelligence – a concept which, moreover, harbors multiple equally problematic definitions[3].

This situation, already quite complex in itself, worsens when we consider

---

[3] Artificial Intelligence can be divided into several categories, including: 1. Narrow AI or Weak AI: Refers to systems capable of performing specific and limited tasks, but lacking general understanding or awareness. Examples include chatbots, recommendation algorithms, and voice recognition software; 2. Strong AI: Refers to systems capable of performing intellectual tasks similar to human abilities, understanding, learning, and solving a wide range of problems. This level of AI is still under development and in-depth research; 3. General AI: Represents a level of AI that can perform any human intellectual activity without limitations. Often referred to as "superior artificial intelligence", it is a highly ambitious goal in this field of research. Applications of AI are widespread across various sectors, including medicine, industrial automation, data analysis, autonomous driving, voice and facial recognition, recommendation systems, games, and many more. AI has demonstrated its ability to tackle complex tasks and improve efficiency in many areas, but it also presents ethical and practical challenges that require attention and regulation.

the concept of mental health. The concept of mental illness is inherently multidimensional, intricate, and mutable. Undoubtedly, a mental distress can induce suffering, compelling the person experiencing it to seek help. However, it can also be perfectly egosyntonic, meaning it aligns completely with the beliefs and lifestyle of the afflicted individual, who, in this case, may not experience any distress and may not feel the need to seek specialist care. It can impede an individual's social functioning, deteriorating interpersonal relationships, academic or work performance. Alternatively, it can remain submerged and inscrutable, without interfering in any way with the social life of the person experiencing it. It might distance the individual from reality, hindering clear thinking, decision-making or rational argumentation of one's own ideas and opinions. Conversely, it might not significantly impact one's grasp of reality, leaving logical capabilities unaffected.

The only certain aspect concerning the characteristics of mental illness seems to be the social stigma that invariably afflicts those with a mental disorder, subjecting them to discrimination, ridicule, and fear from the community. This isolation, often facilitated by societal attitudes, frequently contributes to worsening their condition.

Thus, within the realm of artificial intelligence research lies the inextricable tangle of problems associated with the philosophical, psychological, and pedagogical definition of intelligence, defining mental health is no less intricate and problematic.

The scope and aspirations of this essay do not allow me to delve into both of these themes, dedicating to them the space and depth of analysis that they unquestionably deserve. I must therefore defer the reader to the specific literature on these matters[4]. I also wish to forewarn that in the following pages, I will assume a broad definition of artificial intelligence as a valid one – an area of computer science focused on creating systems and

---

[4] SADIN 2019; BODEN 2021; FLORIDI, DURANTE 2022; KISSINGER, HUTTENLOCHER, SCHMIDT 2023.

technologies capable of performing activities that typically require the typical support or input of human intelligence. In essence, a field of research aimed at developing software or machines capable of imitating and/or enhancing certain cognitive abilities of humans, such as reasoning, learning, adaptation, and problem-solving.

1.1. A BRIEF HISTORICAL EXCURSION: FROM THE CHESS PLAYER TO CHATGPT

Since its inception, automation has been deeply intertwined with the contemplation of the human mind. Indeed, the creation and spread of the first automatons in the 18th century sparked debates in the philosophy of mind, stimulating intense discussions. Enchanting crowds of spectators, those early robots inevitably provided fodder for scholarly discourse. Granted, the functions performed by the first 'thinking machines' were elementary and equally limited – at times, as in the case of the famous Turk that played chess, they relied on mere circus tricks, crafted to exploit public credulity.

However, the fact that a humanoid object could autonomously execute an action already posed the fundamental philosophical problem of the similarity and difference between the mechanisms governing its movement and the human mind.

This theme became even more urgent and necessary when, in the mid-19th century, cybernetics was born – a new science dedicated to the study of control, communication, and regulation in complex systems. In other words, a new science dedicated to how information is processed, communicated, and stored by biological, social, and mechanical systems. In its own way, cybernetics also brought the human mind into focus, amplifying the play of mirrored reflections between human and artificial intelligence that had fascinated scholars in the era of the early automatons.

As for the realm of computer science, Alan Turing, one of the most renowned and rightfully celebrated scholars of the 20th century, delved into

artificial intelligence in a 1950 essay, still considered today as an epochal and indispensable reference point for the field[5]. In that article, Turing devised his famous test, partially refuted by the equally renowned thought experiment of the Chinese Room proposed by John Searle[6]. The Turing Test and the discussions it spawned eloquently demonstrate how computer research and philosophical speculation have framed the problem for about eighty years, laying the foundational groundwork for today's debates from the very beginnings of the earliest mainframes.

Yet, after this initial surge, the discussion on artificial intelligence lay dormant – almost 'karst-like' – until the advent of generative artificial intelligence, especially the GANs systems of machine learning[7], awakened and stirred consciousness, shaking many old and deeply entrenched certainties. In particular, the new computer systems seem to have checkmated the ancient (pre)judgments on purely functionalistic arguments – that's to say, on what machines 'still cannot do'. From here, the debate regained both breath and fire.

Starting from this concise yet necessary historical reconstruction, we can knowingly turn our attention to the present, and more specifically, to the etiopathogenic implications of AI in the era of big data, neural networks, and the Web 3.0.

---

[5] TURING 1950.

[6] SEARLE 1990.

[7] Generative artificial intelligence is a field of artificial intelligence that focuses on creating models, algorithms, or systems capable of producing 'original' content, such as images, texts, or music. These systems are designed to learn from the information present in training data and then generate new content similar to that created by humans. The family of models called GANs (Generative Adversarial Networks) involves placing two neural networks in a collaborative competitive relationship: a 'generator' network creates data, such as images, and another 'discriminator' network tries to distinguish between outputs generated by the first network and those generated by humans. During training, the networks mutually improve, with the goal of creating increasingly compelling and realistic content. Generative artificial intelligence has applications in various sectors, such as visual art, text writing, music creation, product design, and much more.

2. THE INTERNET ADDICTION DISORDER

In 1994, British psychologist Mark Griffiths published an essay on the Psychology Forum dedicated to technological addictions, those dependencies "which involve human machine interaction". In that essay, the author documented a varied phenomenology of pathological behaviors, paying particular attention to gambling addiction (fruit-machine addiction), telephone sex addiction and internet addiction.

Regarding the latter, Griffiths believed that the diagnostic criteria essentially consisted of six points: 1) Salience: the situation where a certain activity becomes the most important in an individual's life, dominating their thoughts, feelings, and behavior; 2) Euphoria: the subjective experience of the dependent individual, described in terms of excitement or relaxation; 3) Tolerance, defined as the need the individual feels to progressively increase their involvement to achieve the desired effect; 4) Withdrawal symptoms: the so-called unpleasant feeling states or physical effects due to abstinence; 5) Conflicts, interpersonal or intrapsychic disagreements that arise due to the dependency; 6) Relapse, the propensity to resume the activity after a period of abstinence, or the inability to recover from the addiction[8].

Just a few months after the publication of this initial essay, American psychiatrist Ivan Goldberg coined the term Internet Addiction Disorder, shaping its diagnostic criteria based on the criteria listed in the DSM-IV for pathological gambling.

About a year later, Kimberly Young, a psychologist and professor at the University of Pittsburgh, founded COLA, a paid online center specialized in treating IAD. According to Young, internet addiction could be diagnosed if five out of the following eight conditions were met for a period of no less than six months: 1) Preoccupation with the internet: the individual

---

[8] CIVITA 2014, 42.

constantly thinks about what they have done or will do online; 2) Increased need to spend time online; 3) Repeated attempts to reduce internet use; 4) Withdrawal when internet use is reduced; 5) Time management problems: the individual tends to stay online for longer periods than initially planned; 6) Relational problems, such as work-related issues, academic performance, or interpersonal problems due to internet use; 7) Lying about the time spent online (the individual hides the amount of time spent online from family or significant others); 8) Mood changes, due to the fact that the patient connects to escape from daily problems, or to alleviate depression, anxiety, guilt, or helplessness[9].

After about twenty years from the publication of these early studies and following extensive research, we can define Internet Addiction Disorder as the distress that afflicts those who spend most of their time and energy online, thereby causing a "significant impairment in major life domains, such as personal, academic, occupational, family, and social"[10] aspects.

As we have considered, the scientific community has been addressing this issue since the mid-1990s. However, IAD was officially recognized only in 2013 when it was included in Section III of DSM-5, alongside conditions requiring further research, such as Internet Gaming Disorder.

It is not yet clear whether there is a single internet addiction or if it is more accurate to speak of multiple addictions, focused at times on gaming, money-seeking, sex, online trading, or social networks. What is certain is that an increasing number of users are beginning to experience the detrimental effects of behavioral addiction, leading to various issues related to physical health, interpersonal relationships, academics, work, or family, which typically accompany it.

It is not entirely implausible to hypothesize that this increasingly concerning spread of IAD is, to a considerable extent, due to the rather malicious use of artificial intelligence by major tech companies and platforms. It's no

---

[9] YOUNG 1996; CIVITA cit., 43; TONIONI 2011, 75.
[10] PIACENTINO, BRUGNOLI 2005, 297.

secret that these companies are harnessing AI's capabilities for self-learning and adaptation to identify users' tastes, preferences, and behaviors. In doing so, they reasonably hope to enhance user engagement, increasing the average time each user spends on the platform.

The situation becomes even more alarming when linked to the phenomenon of big data – the growing volume of data users unwittingly feed into the network, including their health information, mood, and opinions. On one hand, this information is explicitly shared by social media users, whether through explicit content, likes, comments, or shares of content generated by others; on the other hand, a myriad of other equally important insights can be derived from 'silent' yet highly significant behaviors – consider, for example, each user's connection time, duration, and visited sites.

All this data enables the training of artificial intelligence in managing digital platforms, ensuring that each user's experience is as personalized as possible, tailored to their personality traits, mood, or likely life developments.

Digital platforms can thus exploit big data and artificial intelligence to keep users chained to the network, whether in terms of gaming, trading, betting, or mere participation in social networks, thereby facilitating the propagation of true Internet Addiction Disorder. This marks the first and most evident etiological connection between Web 3.0 and mental health. However, it's not the only one.

2.1. THE ATTENTIVE RESOURCES

When we use the term attention, we are normally referring to the human ability to isolate specific stimuli – sensory, cognitive, or mental in nature – separating them from the countless inputs generated by the environment, our own consciousness, or the body. This is done in order to focus our thoughts on these stimuli. It is a psychological resource of great importance, without which nearly any intellectual activity would be almost impossible – such as making decisions, reflecting on the past, or solving problems.

However, like any other human characteristic, quality or resource, attention

also has its limits. As students well know, attention requires a significant amount of mental energy. Its activation and maintenance depend on various psychological and environmental variables – it's certainly easier to stay focused in a relatively quiet environment rather than in the midst of a construction site; when one has managed to get proper rest rather than after a sleepless night. Above all, attention tends to wane over time. Following this inevitable depletion, the mind feels the need to rest, detaches from the chosen object, and begins to wander in search of other content – less demanding or perhaps more urgent.

However, it would be incorrect to assume that there is only one way to understand attention. Psychology teaches us that there are at least two types: the first, called focused attention, is the ability to direct the entire mind to a single object, perfectly isolating it from the context in order to analyze and understand every aspect of it; the second, known as shared attention, is the ability to consider multiple stimuli simultaneously, keeping various needs and goals in mind. Both modes of attention are necessary and valuable. Shared attention is particularly useful for fulfilling the practical tasks of daily life, which are full of constant surprises, deadlines, and demands that inevitably overlap, forcing us to create schedules and establish hierarchies among the urgencies. Focused attention, also known as hunter's attention, is essential for intellectual life – studying, writing, and reasoning.

Maintaining focused attention isn't easy when the context in which we're immersed constantly stimulates our consciousness through endless inputs – in this case, it might be more accurate to say: 'endless notifications', demanding consideration and, above all, requiring real-time responses.

Besides being numerous, the cognitive stimuli provided by computers and smartphones are diverse in terms of sender, form, and content. This means that, in order to consider each of them, we must be able to quickly change our interpretive and expressive code – thereby expending additional important psychological resources.

The (compulsive) use of smartphones thus affects our attentive resources, leading to mental fatigue and reducing our ability to make sensible decisions. It's no coincidence that younger generations have much poorer focused

attention compared to previous ones. However, they have gained more shared attention. Shared attention, also known as multitasking, brings us closer to computers and their artificial intelligence, but for us, it remains, physiologically, shallow.

3. AI IN THE SERVICE OF MENTAL HEALTH

As we have already discussed, artificial intelligence has the potential to seriously jeopardize, if not outright damage, the mental health of internet users by maliciously exploiting the assets of big data available to digital platforms. It can create multiple and personalized 'psycho-traps', such as sending users specific notifications and messages designed to influence their behavior and thoughts. This weakens their reasoning and particularly their decision-making capacity due to the depletion of attentive resources resulting from a bombardment of tailored stimuli – based on the individual's personality traits and lifestyle.

However, artificial intelligence can also provide significant assistance in the realm of mental health, seen from a broad perspective of 'salutogenesis'. Coined by sociologist Aaron Antonovsky in the 1970s, this term refers to a holistic approach to health, focused on promoting and maintaining a healthy life rather than solely treating illness. Salutogenesis thus investigates what contributes to a better quality of life, rather than exclusively analyzing etiological and pathogenic factors.

The World Health Organization (WHO) seems to adopt this concept of health when stating that "health is a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity". Moreover, Antonovsky's great merit was considering the relationship between health and illness as a continuum rather than a binary alternative – an approach congruent with the increasingly widespread dimensional conception of mental distress.

From this specific perspective, it seems evident that artificial intelligence can significantly improve the health of users of digital platforms. For example,

it can use their available data to disseminate prevention and treatment messages among particular categories considered at risk for specific mental disorders based on factors like age, socioeconomic status, or profession. Think of burnout, which often affects people in helping professions; anorexia, which predominantly affects young females; or internet gaming disorder, more common among adolescent males. Furthermore, artificial intelligence can use the data to deliver targeted messages to specific users when they might need information about the risks associated with lack of sleep, poor nutrition, or the dangerous habit of using the phone while driving.

In essence, if the concept of health is considered broadly to include prevention activities, it's evident how much good we can achieve from artificial intelligence in terms of simple dissemination and information.

In a more specific sense, artificial intelligence can be used as a tool for techno-regulation and self-control. In the first case, it can autonomously intervene promptly, for instance, to block access to a gambling site or obscure pornographic content if the user's behavior shows signs of emerging addiction. In the second case, it could respond to a user's request to cut off data traffic after a certain time threshold or prevent the user from exceeding a specific spending limit on gambling websites.

Lastly, there is a myriad of chatbots dedicated to mental health. These chatbots are designed to offer listening and guidance to those in need of emotional support, mental health information, or advice on how to deal with certain existential problems. They can be accessed through online platforms, mobile apps, or websites. Here are some examples:

1. Woebot: Woebot is a chatbot designed to provide emotional support through guided conversations. It employs therapeutic principles based on cognitive-behavioral therapy to help users manage stress, anxiety, and other issues.

2. Wysa: Wysa is a mental health chatbot that uses artificial intelligence to offer support for anxiety, depression, insomnia, and other emotional problems. It also provides self-help exercises and well-being management techniques.

3. Tess: Tess is a mental health chatbot created for U.S. military personnel and veterans, but it is accessible to all. It offers emotional support and tools to cope with stress, depression, and anxiety.

4. Youper: Youper is a chatbot that combines cognitive-behavioral therapy principles and artificial intelligence to help users monitor their emotional state, identify thought patterns, and develop coping strategies.

5. Replika: Replika is an artificial intelligence chatbot that can be used for light conversations as well as discussing personal and emotional issues. It can provide support and companionship.

However, it's important to note that counseling chatbots are not substitutes for professional counseling from qualified therapists or psychologists. They can be useful for short-term support and self-awareness, but for serious problems, consulting a mental health professional is advisable.

4. CONCLUSIONS: CHALLENGES AND ISSUES FOR THE LAW

So far, we have examined the positive and negative consequences that artificial intelligence can bring – and is already bringing – to the mental health of users of digital platforms. In the following, we will reflect on the role that, within this context, falls to legal experts. We will consider the challenges that await legislators, who already appear well aware and determined to regulate the use of artificial intelligence systems[11].

The first point to consider in this regard concerns the classification of mental health as a human right. As we have previously discussed, the concept of psychological distress is quite problematic. However, there is no

---

[11] Following the solicitation by the Commission itself, in 2021, the European Union announced, in June 2023, the forthcoming promulgation of the "first regulation in the world of artificial intelligence" (https://www.europarl.europa.eu/news/it/headlines/society/20230601STO93804/norm ativa-sull-ia-la-prima-regolamentazione-sull-intelligenza-artificiale).

doubt that mental health is a fundamental human right. This should not be overly surprising, considering that the same issue pertains, *mutatis mutandis*, to many other crucial human rights, such as the right to health, human dignity, or the much-debated metaphysical chasm that opens beneath them whenever legal experts employ the term freedom.

The notion that mental health is a human right is clearly deduced from the combined provisions of Article 25 of the Universal Declaration of Human Rights[12], Article 11 of the European Social Charter[13], Article 12 of the International Covenant on Economic, Social, and Cultural Rights adopted by the United Nations General Assembly on December 16, 1966[14], and Article 10 of the Additional Protocol to the American Convention on Human Rights in the Area of Economic, Social, and Cultural Rights signed

---

[12] "1) Every individual has the right to a standard of living adequate for the health and well-being of themselves and their family, including food, clothing, housing, medical care, and necessary social services; and they have the right to security in the event of unemployment, illness, disability, widowhood, old age, or other circumstances beyond their control that result in loss of livelihood; 2) Motherhood and childhood are entitled to special care and assistance. All children, whether born in or out of wedlock, shall enjoy the same social protection".

[13] "To ensure the effective exercise of the right to health protection, the Parties commit to adopting, both directly and in cooperation with public and private organizations, appropriate measures aimed in particular at: 1. Minimizing the causes of impaired health as much as possible; 2. Providing clinics and educational services related to the improvement of health and the development of a sense of individual responsibility in health matters; 3. Preventing, to the extent possible, epidemic, endemic, and other diseases, as well as injuries".

[14] "1. The States parties to the present Covenant recognize the right of every individual to enjoy the highest attainable standard of physical and mental health. 2. Measures which the States parties to the present Covenant undertake to ensure the full realization of this right shall include those necessary for: a) the reduction of stillbirths and infant mortality and for the healthy development of the child; b) the improvement of all aspects of environmental and industrial hygiene; c) the prevention, treatment, and control of epidemic, endemic, occupational, and other diseases; d) the creation of conditions which would assure to all medical service and medical attention in the event of sickness".

in 1988 in San Salvador during the eighteenth regular session of the OAS General Assembly, entering into force on November 16, 1999[15].

For this reason, it is the duty of legislators to protect, hinder, and contain activities, consumer goods, and digital services endowed with a scientifically proven etiological – if not epidemiological – significance.

Regarding the ethical value, even before the legal aspect, of artificial intelligence, we observe that an ancient and widespread philosophical current argues that technology has no intrinsic value – that every technology must be considered, by its very nature, ethically neutral, devoid of a clear axiological connotation. In other words, all tools are, by their nature, devoid of responsibility, attitude, and character; any judgment of their goodness depends on how humans decide to use them. This thesis, although appearing straightforward and sensible, risks confusing the discussion on freedom (and the resulting responsibility) with the different discourse on essence or nature.

While it is true that the tools we use are not free and hence not responsible, it is equally true that they possess a structure, an architecture, an objective form that objectively defines their meaning. What I mean is that the way they are constructed allows us to phenomenologically analyze their essence and thus to reason about what they are – regardless of how humans may choose to use them.

Legal knowledge is well aware of the distinction between the discourse on

---

[15]"1. Everyone has the right to health, understood as the right to enjoy the highest level of physical, mental, and social well-being. 2. In order to ensure the exercise of this right, the States Parties agree to recognize health as a public good and, in particular, to adopt the following measures to guarantee this right: a) primary medical services, meaning medical care available to all individuals and families at the community level; b) expansion of medical services for all individuals under the jurisdiction of the State; c) universal vaccination against major infectious diseases; d) prevention and treatment of endemic and occupational diseases, among others; e) education of the population on the prevention and treatment of health issues; and f) responding to the health needs of higher-risk groups and those made more vulnerable by poverty".

freedom and that on nature. Accordingly, it recognizes that objects and tools also have their own essence, a specific nature. For this reason, the Italian Penal Code considers certain objects equivalent to "real weapons"[16], while others are treated, by the Italian law, as "improper weapons". The former are qualified as objects intended to cause harm, while the latter can be maliciously diverted from their normal use to inflict physical harm on another person. Another example concerns child pornography materials, the production, possession, sale, exchange, and offering of which are prohibited as well as controlled substances.

These regulations do not imply any judgment about the ethical value of an object, substance, or tool – it is not the photograph or video itself that is morally or legally responsible. Instead, they sanction the objective structure of a mean as inherently illicit – regardless of the concrete use humans intend for it.

If we agree to engage on this ground, that is, in the phenomenological realm of essence, we must acknowledge that artificial intelligence, in its essence, is not a virus, drug, or weapon. Rather, it seems to me to be an eminently cultural tool – the highest level achieved by our computational competence, an achievement of which we should be proud, rather than an enemy to fear and combat.

However, even artificial intelligence, like all tools, is susceptible to being misused and diverted for political or profit-driven purposes that are not structurally inherent to it[17]. It becomes clear, then, what the greatest challenge for the field of law in the years to come will be. It will be up to legal experts to build a context, a framework, capable of allowing artificial intelligence to remain true to its cultural vocation.

In particular, national laws and international agreements will need to address the thorniest issues related to the legal liability of choices that artificial

---

[16] See art. 585 Italian Penal Code.
[17] MAYER-SCHONBERGER, CUKIER 2013; CHRISTL, SPIEKERMANN 2016; CLEGG 2017; AMATO MANGIAMELI, CAMPAGNOLI 2020.

intelligence systems can autonomously make. They must clarify intricate matters concerning the protection of users' privacy, the economic exploitation of data collected by digital platforms[18], and responsibility for their unlawful processing.

These legal instruments should prevent automatic chats, especially those dedicated to mental health, from being perceived by users as substitutes for the work of a specialist. This is not so much to safeguard jobs, but rather to protect users from the illusion that they can bypass the human relationship with their therapist – a relationship that inherently holds a significant healing value.

In essence, it will be up to legal experts to establish a set of rules that can restrict the exploitation of artificial intelligence with manipulative and/or harmful outcomes for users' mental health, while also facilitating its increasing use for their benefit and protection.

---

[18] ACQUISTI, TAYLOR, WAGMAN 2016.

REFERENCES

ACQUISTI, Alessandro, TAYLOR, Curtis, WAGMAN, Liad, *The Economics of Privacy*, 442-492, in *Journal of Economic Literature*, 54, 2/2016.

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

BODEN, Margaret A., *L'intelligenza artificiale*, trad. it., Bologna, 2021.

CHRISTL, Wolfie, SPIEKERMANN, Sarah, *Networks of Control. A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy*, Wien, 2016.

CIVITA, Anna, *Un malessere sociale: la dipendenza da internet*, Milan, 2014.

CLEGG, Brian, *Big Data. How The Information Revolution Is Transforming Our Lives*, London, 2017.

FLORIDI, Luciano, DURANTE, Massimo, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Milan, 2022.

KISSINGER, Henry, HUTTENLOCHER, Daniel, SCHMIDT, Eric, *L'era dell'intelligenza artificiale. Il futuro dell'identità umana*, trad. it, Milan, 2023.

MAYER-SCHÖNBERGER, Viktor, CUKIER, Kenneth, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, trad. it., Milan, 2013.

PIACENTINO, Daria, BRUGNOLI, Roberto, *Le dipendenze comportamentali: introduzione*, in CARETTI, Vincenzo, LA BARBERA, Daniele, *Le dipendenze patologiche. Clinica e psicopatologia*, Milan, 2005.

RALPH, Brandon C.W., THOMSON, David R., CHEYE, James A., Smilek D., *Media multitasking and failures of attention in everyday life*, 661-669, in *Psychological Research*, 78, 5/2014.

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

SEARLE, John R., *Is the Brain's Mind a Computer Program?*, 26-31, in *Scientific America*, 262, 1/1990.

TONIONI, Federico, *Quando internet diventa una droga. Ciò che i genitori devono sapere*, Turin, 2011.

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

TURING, Alan M., *I calcolatori digitali possono pensare?*, 21-26, in *Sistemi intelligenti*, 10, 1/1998.

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

YOUNG, Kimberly S., *Internet Addiction: The Emergence of a New Clinical Disorder*, 237-244, in *CyberPsychology & Behavior*, 1, 3/1996.