

## Urszula Jarecka\*

ORCID: 0000-0002-2667-2898

Theory of Culture Unit, Institute of Philosophy and Sociology PAS

## Paweł Fortuna

ORCID: 0000-0002-0633-4453

The Catholic University of Lublin

## SOCIAL MEDIA IN THE FUTURE: UNDER THE SIGN OF UNICORN...

In this essay case studies pointing to problems related to the use of AI in shaping the virtual world are discussed. AI algorithms helps to shape and control the conventional web behaviour and speech of today's media users, mostly teenagers and adults. Considering the development of software, social media may constitute a separate virtual world in the future. AI also shapes the image of this world and human relationships. The essay begins with an analysis of the future of social media against the background of truth; later, case studies show problems caused by AI to media users and the community. The authors attempt to answer questions such as: What kind of attitudes and abilities will be shaped in social media? What network ethics does AI dictate? What kind of attitudes and thinking will be promoted in the social media of the future?

Keywords: emotions, social media, Artificial Intelligence, virtual world, algorithms, network ethics, utopian media

Social media uses Artificial Intelligence (AI) support. Systems based on AI learn the human world and influence its shape, confirming the thesis of Stanisław Lem (1988) that technology is a variable independent of civilization. Given the development of software, social media may come to constitute a separate virtual world in the future. AI also shapes the image of this world and interpersonal relationships in a specific way<sup>1</sup>. As in the mythical

---

\* Theory of Culture Unit, Institute of Philosophy and Sociology PAS, ul. Nowy Świat 72, 00-330 Warsaw, Poland; e-mail: [ujarecka@ifispan.edu.pl](mailto:ujarecka@ifispan.edu.pl) (corresponding author)

<sup>1</sup> ∞Meta company assures that. "However, through self-supervised learning, machines are able to learn about the world just by observing it and then figuring out the structure of images, speech or text. This is a more scalable and efficient approach for machines to tackle new complex tasks, such as understanding text for more spoken languages. Self-supervised learning algorithms for images, speech, text or other modalities function in very different ways, which has limited researchers in applying them more broadly", <https://about.fb.com/news/2022/01/first-self-supervised-algorithm-for-speech-vision-text/> [4.02.2022].

world of unicorns, whose horns held magical properties (as antidotes to poison) AI in social media tracks “poisoned” words and false information. There is no falsehood in the world of unicorns, no aggression nor violence; those who are able to see them have pure hearts. In a reality shaped that way, one can assume there will be space for joy and happiness to be experienced and not only expressed in the form of emojis, photos, and comments. This discrepancy between experience and its expression is reflected in the saying: “We are a sad generation with happy pictures” (Fig. 1).



Figure 1. Meme from the net

In various ways, AI has become a part of our lives for good. However, there is still a need for improvement and rethinking of the role of AI in our future social lives. Young people brought up in times of widespread digitalization, common usage of smartphones, and not fully understood consequences of their constant use, will live in the Internet of Things, in a reality at least partly designed and controlled by AI. What kinds of attitudes and abilities will be shaped by that fact, and which ones will be needed in the future? Conventional web behavior and speech currently shaped by AI algorithms affect today’s media users, who are mostly teenagers and adults.

In this essay, we will consider case studies pointing to problems related to the use of AI in shaping the virtual world. We will start by considering the future of social media and the truth in order to show, through case studies, what problems AI is facing, and which media users and communities it affects. What does the truth of emotions, shaped on the web, look like? Finally, we will consider what network ethics are dictated by AI and what kind of attitudes and thinking will be promoted in the social media of the future.

THE FUTURE AND THE TRUTH: NOT ONLY THE METAVERSE<sup>2</sup>

The agony of portals such as MySpace (and other platforms shaped locally in various countries, in Poland e.g., grono.net, Naszaklasa.pl) was a natural phenomenon with the appearance of Facebook, a website that better senses the demand for varied and globalized contacts on the web<sup>3</sup>. Instagram completed image-related tasks, focusing primarily on the visual aspect of communication. Additional possibilities offered by audio calls, video calls, and text chats on both websites ensured the possibility of multi-level contacts. On the other hand, it seems that social media do not develop along with human social development (see Hjorth and Hinton, 2019, 36–52; Lovink, 2016), but functions instead as an independent technical communication structure to which the unfortunate name has stuck. At times social media is even called “anti-social” (Vaidhyanathan, 2018). One can also wonder if the fabric of reactions and interpersonal relationships shaped by AI has anything to do with the real world<sup>4</sup>.

More and more “fossilized,” a predictable and expanded Facebook, along with a visually overloaded Instagram, belongs to pioneering media, that is, media of the breakthrough. On the other hand, other entities are also developing among social media outlets such as SnapChat and TikTok, assuming the role of a kind of safety valve for the “classic” and more structured Facebook and Instagram<sup>5</sup>. As with other messengers, using TikTok does not encourage community building, although such a goal is written in its mission. It mainly serves as an entertainment wizard<sup>6</sup>; at times, it fulfills this task at a very poor level. Well, to be friendly and to satisfy everyone is a tricky task and one needs need to be truly Machiavellian to do that. To “bring joy” is also a very general assumption, in some circles it is treated as a joke because joy can be understood in various ways. When it is applied in a sarcastic way, it can serve to explain vulgar words or violations of political correctness and the dignity of people belonging to various groups, e.g., women (brutal behavior, mockery of women’s skills, dissemination of stereotypical roles and activities). One can always say, “it’s just a joke”, although in such situations the joke always has a victim.

<sup>2</sup> <https://edition.cnn.com/2021/10/28/opinions/zuckerberg-facebook-meta-rushkoff/index.html>; <https://www.npr.org/2021/10/28/1050280500/what-metaverse-is-and-how-it-will-work?t=1643920950933> [6.02.2022].

<sup>3</sup> On MySpace and media changes (Levinson, 2013, pp. 126–131).

<sup>4</sup> The dichotomy between “social media” and the “real world” in this paper is used technically, as it is rhetorically useful. The meaning of social media for users is unquestionable. However, it’s good to remember that globally as well as locally the percentage of people who do not use the Internet, including social media, is still quite big. On the scale of digital exclusion, for Internet non-users (offline population), social media can’t build their reality, as one can read in these reports: <https://www.capgemini.com/pl-pl/wp-content/uploads/sites/15/2020/05/Infographic-%E2%80%93-Digital-Divide-10.pdf>; <https://www.iriss.org.uk/resources/esss-outlines/digital-inclusion-exclusion-and-participation>; in Poland the problem was noticed during the POLPAN research results: <https://www.computerworld.pl/news/Blisko-1-3-Polakow-nadal-cyfrowo-wykluczona,434287.html>; all sites accessed [4.04.2022].

<sup>5</sup> New generations of users reach for various messaging services, not related to the metaverse. Nota bene, if the development of the media was sinusoidal, children born in the ‘20s of the 21st century would return to ∞Meta.

<sup>6</sup> “TikTok’s mission is to inspire creativity and bring joy. We are building a global community where people can create and share, discover the world around them, and connect with others across the globe. As we grow, we are committed to maintaining a supportive environment for our community. Our Community Guidelines define a set of norms and common code of conduct for TikTok; they provide guidance on what is and is not allowed to make a welcoming space for everyone”, <https://www.tiktok.com/community-guidelines?lang=en> [4.02.2022].

In contemporary social media, focused on advertisers, the user seems to be only a tenant of a place that stores their content. Social media do not shape the community in order to help young people to mature. As Manfred Spitzer reminds us:

“... social behavior [we learn] – in mutual relations with other people. For that to happen, you need to be in the company of other people. This is exactly where digital media hits social competences: they crowd out real virtual contacts” (Spitzer, 2016, p. 315).

Young people do not learn to build communities on social media. For many users, social media is a one-person broadcasting station, with a counter of viewers and likes as a measure of social “success.” Social media has evolved into a system of one-way messaging. It has become a source of isolation or – to a point – even “anti-social”<sup>7</sup>.

The first problem that we want to draw attention to is the immaturity of AI in assessing the statements of network users, especially when it comes to content composed of images and text. The algorithm is not prepared for non-standard reasoning. The algorithm that keeps order and cares about the “truth” often extinguishes individuality and does not allow for “leaning out.” The algorithm cares about the truth<sup>8</sup> in terms of understanding political correctness, which is not a set of good manners, but an ideological version of language, with the assumption that there are minority groups that can be easily discriminated against with the use of language. Political correctness as a set of limitations and censorship prejudices, however, does not result from the actual state of the social order; from the current state of language, instead, it is a demand for change where problems may arise (see Duingnan and Gann, 1995). Political correctness is also a temporary phenomenon: in a few years from now, neutral words may be considered offensive or insensitive and other groups may be considered minorities. One may wonder whether political correctness prevents exclusion or reminds us of that exclusion. The cultural aspirations of AI creators may raise concerns because the world cannot be developed according to a project, even if it is the most sublime one, which should be remembered by Plato’s famous experiment in Syracuse.

## THE ALGORITHM KNOWS BETTER

Algorithms are a part of a “package” of new technologies known as STARA (Brougham and Haar, 2018). We use the term algorithm to denote any automated formula, rule, or set of instructions that are used to process data or perform a specific task (Germann, 2018). The development of these formulas is a key element in the improvement of AI that was intended by the creators of this concept to model the cognitive functions of a human being. Although the hopes of scientists to discover the mysteries of the mind as a result of such activities have

---

<sup>7</sup> On loneliness and depression among social media users writes Manfred Spitzer (2016). “Anti-social” face of social media was discussed in the book by Siva Vaidhyanathan (2018).

<sup>8</sup> Facebook also assumes that we publish “real” data about ourselves like our place of work, residence, age, etc. One may wonder what the reason for it may be and whether it is for commercial purposes, a targeted identification of potential customers for advertisers.

been disappointing (Schank, 2020), the tasks they perform stimulate their anthropomorphization, which is not always justified. Research comparing the understanding of AI by laymen and experts show that algorithms are a key dimension of the AI category for the latter (Fortuna and Gorbaniuk, 2022). Laymen identify AI primarily as embodied algorithmic forms in the shape of humanoid robots. Experts' opinions also differ from common beliefs about artificial systems, consistent with pop culture narratives<sup>9</sup>. For example, Marta Kwiatkowska from the Department of Information Science at the University of Oxford and a member of the Royal Society bluntly stated in an interview that artificial intelligence,

is still very stupid. It can hardly be called intelligence; it is more of an idea created by the media. This supposed intelligence is in no way equal to that of a human. It can only classify human input and make decisions based on it. It cannot cope with the context in which the system operates.<sup>10</sup>

Despite the limitations of AI, it ought to be indicated that the tasks it carries out are impressive. For example, Youyou et al. (2015) compared the accuracy of personality assessments made by people and by the artificial systems. They found that the assessments made by a rational agent who analyzed the respondents' reactions in the form of liking Facebook posts were more accurate than those made by their friends using a personality questionnaire. The algorithm's ratings also had a higher agreement with the judges' ratings and better accuracy in predicting such aspects as drug use, political attitudes, and physical health.

Research on AI acceptance and adoption emphasizes the need to distinguish between two types of tasks performed by artificial entities (Inbar et al., 2010). The implementation of objective tasks requires the use of rules and logic, such as making calculations or sorting data. However, in order to perform subjective tasks it is necessary to be guided by emotions and intuition. The research results clearly indicate that the level of trust towards algorithms significantly decreases when they are intended to perform subjective tasks. When ranking the tasks, Castelo et al. (2019) stated that the most trusted algorithms were the ones used for navigation, data analysis, event planning, and weather forecasting, while the least trusted ones were used for predicting funny jokes, hiring and firing employees, recommending a partner, and writing press articles. A negative attitude towards artificial systems, the so-called algorithm aversion (Dietvorst et al., 2015), is observed where artificial systems perform tasks that are usually performed by people. It is strongly noticeable in the domains of art, marketing (Davenport et al., 2020), and medicine. In the last, respondents prefer a medical provider that offers a diagnostic or treatment service (Longoni et al., 2019).

Based on the results of research on the adoption of digital technologies, Fortuna (2021) distinguished the following conditions that, in the opinion of the respondents, should be met for the level of acceptance towards algorithms to increase:

- considering events that are extremely rare;
- elimination of irregularities by learning from mistakes and taking responsibility for them;
- the ability to see a mistake before making a final decision;

---

<sup>9</sup> A clear example is the widespread attribution of artistic abilities to the Ai-Da robot in the media.

<sup>10</sup> *Sztuczna inteligencja jest jeszcze bardzo głupia*, <https://tygodnik.tvp.pl/46340221/sztuczna-inteligencja-jest-jeszcze-bardzo-glupia> [5.02.2022].

- having knowledge of what information should be taken into account in the decision-making process;
- using procedural, overt, and tacit knowledge (similar to experts);
- cognitive flexibility enabling the transfer of knowledge from one field to another;
- the ability to “read between the lines”;
- relying on evidence in making conclusions;
- proper “reading” and understanding of emotional expressions (e.g., patients);
- respecting the uniqueness of people and the ability to conduct in-depth psychological analyses;
- considering qualitative data, not just parameters (e.g., test results);
- compliance with ethical principles.

Improper delegation of tasks (techno-empowerment) to artificial systems (due to their lack of competence) may lead to human machine trans role conflict (HMTRC), which in the consumer’s perception affects the negative assessment of managers, lowers the sense of security, and arouses the intention to boycott the institution. Such results have been obtained in research conducted in Poland in museum studies. Consumers revealed more negative attitudes when they learned that robots (a strong HMTRC) would act as guides and animators at the museum. Attitudes were more positive when participants were informed that artificial systems would deal with statistical data analysis and ticket sales (weak HMTRC). Of course, the accurate use of algorithms can be very beneficial for humans. Examples include social and police robots or systems adapted to be worn by humans (wearable robots), for example in the form of a suit (e.g., Mollii Suit<sup>11</sup>) that allow people suffering from Parkinson’s disease to move more freely.

Algorithmic systems are developed every day. In Germany 78 000 patents that included the use of AI were registered in 2018, which means that over the course of a decade their number has more than tripled (in 2008 there were about 23 000)<sup>12</sup>. Google for Startups data from 2020 shows that the number of startups in Poland has doubled over the past five years and about 60 percent of them are companies from the IT industry. In turn, the market of social robots is growing seven times faster than the market of industrial robots. At the end of 2017 it reached US\$5.4 billion and is expected to increase to US\$14.9 billion in 2023 (Business Wire 2017). Significant changes will be related to the introduction of augmented reality systems. Lenses with a built-in screen and camera lenses are already available on the market. Overcoming the technical difficulties related to powering these artifacts will allow each consumer to see advertisements in which the product or service is offered to them by a character bearing their face. The option of eliminating advertisements is also possible because the algorithms that have access to data about us can independently compose an offer. They can also use a courier to send us products that we did not order but that potentially may suit our tastes

---

<sup>11</sup> <https://www.hobbsrehabilitation.co.uk/mollii-suit.htm> [14.02.2022].

<sup>12</sup> Data from the German research company IPlytics (Bettman and Oksanowicz, 2021).

(Diamandis and Kotler, 2020). Huge changes are also associated with the digitization of workplaces. In the post-humanist perspective, the functioning of organizations that not only delegate various types of tasks to algorithms but also assign them the role of a supervisor or even of a CEO has been considered for years (Gladden, 2016).

## IN THE IMAGE AND LIKENESS OF AI

Is AI really an imitation of human intelligence? Algorithms prepared by teams of engineers cannot be as malleable as the human mind because we do not know everything about it. There is no consensus or certainty about how the mind works, what consciousness and the unconscious are. We use intuition that has been neglected by research in psychology and philosophy.<sup>13</sup> The mere introduction of deep learning in the development of AI<sup>14</sup> has nothing to do with the maturation of the mind and human functioning, where apart from rational operations other crucial elements are developed, e.g., understanding of other aspects of reality, social world, habits, sympathies, etc. People are not perfectly rational, they are able to hold contradictory views, change their minds under the influence of emotions, etc. The languages we use are also far from being universal and they are also changeable creations.<sup>15</sup> Also, the symbols we use are not culturally universal. As shown by Richard E. Nisbett in “The Geography of Thought” (2003), people differ in their approach to the world: They construct different systems of values and intellectual interpretations of the world, of nature, and social reality. Moreover, symbols used by people change throughout history, sometimes their meanings are complex or self-contradictory.<sup>16</sup> How does an algorithm deal with such irrational or ambiguous thinking? Can we say that AI is shaped like human intelligence? If so, which human being would be its model?

As humans, not only are we task-oriented, but we are also sentient beings. However, problems still arise with designing the whole area of understanding emotions. It is worth mentioning that Marvin Minsky, guru of the cyber-construction world, attempted to reduce various human states and reactions to intellectual operations. Minsky stated:

Although this book is called *The Emotion Machine*, it will argue that emotional states are not especially different from the processes that we call “thinking;” instead, emotions are certain ways

---

<sup>13</sup> In some publications, intuition is presented as being typical of mythical-magical or “pre-theoretical” thinking (Copp, 2018; Svedholm and Lindeman, 2013). “Intuition is a type of mental activity that unconsciously communicates perceptions – it is the perception of possibilities. The contents of intuition are given, they are not spoken, they suddenly present themselves as a ready whole. [...] Intuition is an irrational (in the sense - non-rational) function. The rationalization of the content of intuition is possible ex post” (Motycka, 2005, 221).

<sup>14</sup> See (Ho, 2019; LeCun et al., 2015; Liang et al., 2021).

<sup>15</sup> “Language has its own history. Each of us has our own language. Two people who share their lives have their own language. There is no problem with one language common to all, there is only a wonder of the fact that although each of us has a different language, we can understand each other across the borders of individuals, nations and times” (Gadamer, 2000, pp. 50–51).

<sup>16</sup> This was the case, for example, “in medieval symbolism one thing could have two opposite meanings depending on the context in which it was seen” (Eco, 2004, p. 121).

to think that we use to increase our resourcefulness—that is, when our passions don't grow till they handicap us—and this variety of ways to think must be such a substantial part of what we call “intelligence” that perhaps we should call it “resourcefulness.” And this applies not only to emotional states but also to all of our mental activities” (Minsky, 2006).<sup>17</sup>

Algorithms that control social media content are not always able to understand context; they block content that is either ironic or historical, such as quotes from old literary works, and they sometimes block angry reactions resulting from suffering. This is because before AI was fully prepared to understand the world of human emotions, it was used for contextual reading of private and public statements as a guardian of the ethical order in the virtual world (see Martin, 2019). The unfinished machine has been harnessed to control such delicate matters as interpersonal relationships and contexts of expression. Let us look at some examples.

#### CASE I.

##### A REMINDER ABOUT MANIPULATION AND EMOTIONS. EMOTIONS AND TRUTH. MANIPULATIONS OF EMOTIONS OR EXPRESSION ON FACEBOOK?

We will cite an example of the infamous manipulation perpetrated by researchers associated with Facebook (now ∞Meta). Over the course of a week in 2012, almost 700 000 users were presented with the manipulated entries of the friends they followed (Kramer et al. 2014):

The experiment manipulated the extent to which people ( $N = 689,003$ ) were exposed to emotional expressions in their News Feed. This tested whether exposure to emotions led people to change their own posting behaviors, in particular whether exposure to emotional content led people to post content that was consistent with the exposure—thereby testing whether exposure to verbal affective expressions leads to similar verbal expressions, a form of emotional contagion. People who viewed Facebook in English were qualified for selection into the experiment. Two parallel experiments were conducted for positive and negative emotion: one in which exposure to friends' positive emotional content in their News Feed was reduced, and one in which exposure to negative emotional content in their News Feed was reduced.

Criticisms were focused on the ethical aspect of the research: tracked Facebook users were unaware of the experiment; they did not agree to participate. The size of the manipulated cohort was also shocking. The authors of the study wrote about “emotional contagion:” “results indicate that emotions expressed by others on Facebook influence our own emotions, constituting experimental evidence for massive-scale contagion via social networks.”<sup>18</sup> Adam Kramer et al. (2014), who published the study, following media critique, also published the relevant explanations.<sup>19</sup> The question of why Kramer's team decided to undertake such an experiment may be a good starting point for new sociological research and philosophical considerations.

---

<sup>17</sup> <https://web.media.mit.edu/~minsky/Introduction.html> [4.02.2022].

<sup>18</sup> <https://www.pnas.org/content/111/24/8788> [4.02.2022].

<sup>19</sup> The first criticism was done by the editor (Verma, 2014).

However, one can have not only ethical but also methodological doubts. Words, however, do not reflect all emotions, and the icons or comments posted on Facebook do not necessarily reflect the feelings of Internet users. After all, the expression of emotions, rather than the emotions experienced, was studied. In order to conclude that emotions on Facebook are “emotionally contagious,” a much more complex study than a study of posted content would be needed. Apart from individual blocks in expressing emotions in the real world, in the virtual world, people may not react because they do not want to leave a trace. It is also not known what the motivation for using a particular icon or comment was. In addition, the response spectrum offered by Facebook is very narrow, not to mention the available reactions on Instagram or TikTok, which only include likes and hearts, when it comes to icons; it is possible to add a verbal-iconic comment, those also have to be chosen from a limited number of options. The spectrum of our emotions is much richer than that.

The future of the web and the future of social media are shaped with the participation of AI. The features of this actor/factor, in creating Internet reality, are sometimes forgotten. The algorithm has no feelings but people have a whole range of them. The truth of feelings is not welcome online. Both anger and violence appear in human experience, but it is happy and smiling content that works well in the case of social media. Let us remind ourselves that “we are a very sad generation with very smiling photos.” Content deemed “offensive” is not shown, based on political correctness and a code of courtesy in Western culture. Users may report a profile as ignoble, indecent, full of offensive content, etc. (various possible reasons for reporting can be presented).

AI censors certain words and phrases in social media, although this does not constitute the entire context. As Hans-Georg Gadamer noted:

There is no statement that can be captured solely for the content it contains if one wants to capture it in its truth. Each statement is motivated. Each utterance has assumptions that it does not utter [...] each utterance has its own situational horizon and is directed to someone (Gadamer, 2000, pp. 46, 49).

## CASE II.

### INSTAGRAM, A COMBINATION OF VISUAL AND VERBAL CONTENT

On Instagram, AI prevents scandalous and discriminatory content by tracking down unwanted content through hashtags assigned to the posts. The unaccepted content is blocked. The question therefore arises: is the poison of anarchy averted? Some Instagram hashtags (indicated by ###) are blacklisted, and posts using them are immediately censored. This blacklist remains flexible. During training, social media specialists pay attention to check ### in special search engines that can sometimes be surprising. One can expect that the content with hashtags such as #porn or #pedofile should be blocked by AI, but the same fate may be met by posts with quite innocent hashtags, such as #exotic. While #exotic is “banned” (see Fig. 2b) and its content cannot be viewed at all, #exoticdance is allowed by the algorithm, although the content there can be erotically provocative (see Fig. 2c). Using this ### with a photo showing exotic landscapes (n.b., exotic from the point of view of Europeans, e.g., palm trees, atolls, and lagoons) may result in not showing this post even to people who follow the account. However,

at times, some Instagrammers try to get around these restrictions, e.g., the neutral #newyork generates photos of naked ladies offering their company either directly or more discreetly;<sup>20</sup> they pose in only their underwear and in a provocative manner, sometimes even NYC does not appear in the background as the photos are taken in the kitchen or in their own bedrooms. The algorithm will catch this content only when other users report it. AI sees no relationship between images and ###. It does not understand the sense that the images create in combination with the caption and the link. Users, either out of ignorance, or on purpose, apply frequently viewed ###, such as #nature, to present themselves, their products and services, etc. One can understand that a naked person may be understood as part of nature, although it is not as easy to consider a car that takes up 70% of a photo a piece of nature (see Fig. 2a).

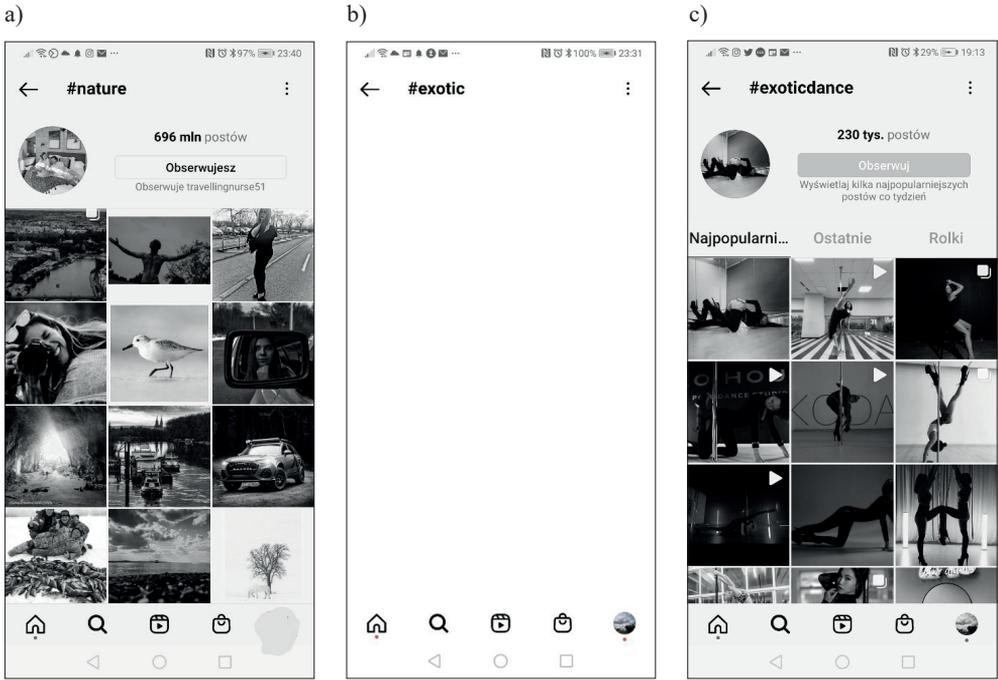


Figure 2. Screenshots showing some of the content of the selected ###: a) #nature; b) #exotic; c) #exoticdance

### THE NETWORK ETHICS. UNDER THE SIGN OF UNICORN?

Artificial intelligence is a crucial factor in shaping Facebook content. How does technology detect violations?

<sup>20</sup> Directly – phone number provided in the post or in the comment; more discreetly – a person requests a private message, which Instagram allows (“DM”).

Our technology proactively detects and removes the vast majority of violating content before anyone reports it. We remove millions of violating posts and accounts every day on Facebook and Instagram. Most of this happens automatically, with technology working behind the scenes to remove violating content – often before anyone sees it. Other times, our technology will detect potentially violating content but send it to review teams to check and take action on it. This work is never finished. People will keep trying to evade our technology, so we need to keep improving.<sup>21</sup>

Social and political meaning is sometimes attributed by AI to content that does not contain such meaning. Sometimes it is impossible to send an advertisement, the meaning of which will be interpreted in its own way by the algorithm, which is the first and main decision maker. When looking for false or socially harmful information in private posts, sponsored, or other content, AI may act overzealously. Once again, there are also gaps in understanding the context of such statements, which results in blocking “innocent” content.

### CASE III.

#### FACEBOOK. LITERATURE AND IMAGE. COMPLEX CONTENT

From October 18 to December 15, 2021, posts related to the project called “With Sienkiewicz by coffee (Letters)” were presented on the WSPAK Foundation’s Facebook page. Their goal was popularizing audio recordings of Henryk Sienkiewicz’s private letters on the Internet. Eight out of those posts served as advertisements, with a total reach of 106 729 users. The mere insertion of posts did not have any unpleasant repercussions, although problems started when trying to set a few of them as advertisements. We will mention two such cases here.

The first of the blocked posts informed about the writing workshop of Henryk Sienkiewicz, his understanding of historical literature and the Old Polish language. There was no contemporary content or veiled political message there. It was illustrated by postcards from the nineteenth and twentieth centuries showing stereotypical images of the nobility at that time. What did the algorithm not like? Could the automatic translation of the content of the post<sup>22</sup> into English cause this confusion, or did the images feel too aggressive or inappropriate (Fig. 3)?

<sup>21</sup> <https://transparency.fb.com/pl-pl/enforcement/detecting-violations/technology-detects-violations/> [4.02.2022].

<sup>22</sup> Letter from Henryk Sienkiewicz to Mściśław Godlewski 14 [September 1, 1880]

How Sienkiewicz prepared himself to write the Trilogy? 😊 We can assume that a good school for a writer was to prepare “Kroniki Szlacheckie” 👉 written for “Niwa”. Later on they were published as the novel “Niewola tatarska” 📖 Here is what the writer’s workshop looks like: “...I have read a lot of things from the 16<sup>th</sup> century and beyond. Therefore, I have mastered the language, and by the way, I try to avoid affectation and various various archaic filler-words like “wždy” and “przedsię”, which are used to patch up ignorance” 😞😞😞 According to Sienkiewicz: “Ancient language relies mostly on volubility, which has almost the seriousness of Latin 🙌, not on scattering it with wild archaisms. If you want, I’ll use so many of them that no one will understand what’s going on [...]. 😊 In order to write well, it is important to know not only the Polish authors from the 16<sup>th</sup> and 17<sup>th</sup> centuries, but also the Latin ones, and their translations made in the past 📚📚📚, because our ancestors were educated based on those texts and they were taken over by their spirit. You also need to be able to recreate the domain of concepts specific to those people, which is already a matter of intuition” 🍯🍯🍯 How do you like the writer’s approach to historical topics? 😊😊 Well, we are impressed! ❤️ (The post of November 8, 2021: <https://www.facebook.com/fundacjaWSPAK/photos/pcb.3004874973110446/3004869436444333/> [4.02.2022].

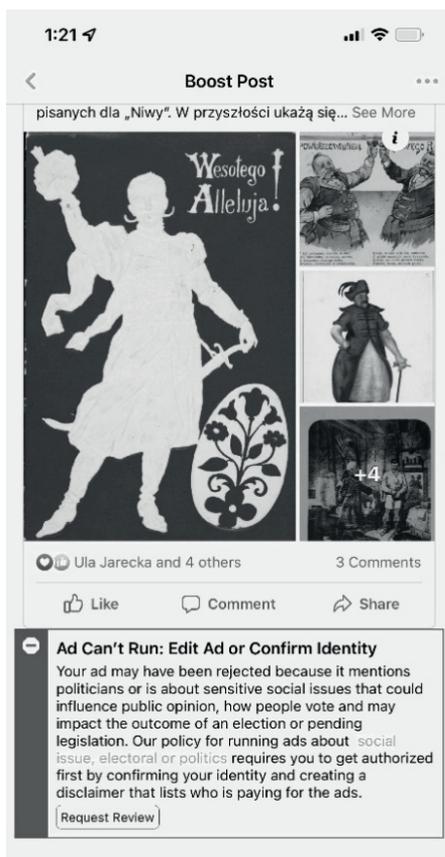


Figure 3. Screenshot from the website of the WSPAK Foundation in November 2021



Figure 4. A still frame from an advertising film for the project “With Sienkiewicz by coffee (Letters)”

Another post blocked by the algorithm was an advertising film based on one of Henryk Sienkiewicz's official correspondences<sup>23</sup>. This letter was also published in a form of English used in the nineteenth century. It contained a historical description of New York, which the future Polish Nobel Prize winner did not like at all.

The advertisement was not permitted in this case either and general information was sent to the Foundation informing that the post violated Facebook's rules. After much correspondence and explanations, both ads were permitted to be launched. However, the Foundation did not get an explanation as to what the algorithm did not like. One can only guess what violated political correctness in these cases. Could it have been a picture showing bags of rubbish lying on the street and rats running around them<sup>24</sup> or perhaps a joyful face of an African American person dancing and singing in Times Square (Fig. 4). The clip was recorded on a casual evening at a place where people are habitually entertained. It was not a social demonstration, there were no politically incorrect slogans in the background, and no subliminal information was used. Perhaps the text-based content of the post raised alarms but the content there was completely neutral<sup>25</sup>. The post contained the usual formula indicating the source financing the project, which was the Polish Ministry of Culture and National Heritage. Could it convey a wrong social or political element? The reason for the algorithm's decision remains a mystery.

When the ads were unblocked, their audience reach was much smaller than in other cases. As for the film, it was nine times smaller than in the rest of the project's commercials<sup>26</sup>. This has not been explained, it can be guessed that an unlocked post is already considered suspicious and is being cautiously or reluctantly displayed by AI to network users. Facebook takes pride in its results of social control in the "Recent Trends" report: "In addition to new categories and ongoing improvements in reducing prevalence, we saw steady progress across many problem areas"<sup>27</sup>. More and more violations were detected.

<sup>23</sup> <https://fb.watch/b2MsOYm5dY/> [17.01.2022].

<sup>24</sup> A few months later, similar images were posted on Instagram: New York for tourists and New York as residents see it. It was a set of photos with garbage bags, corks, etc. [https://www.instagram.com/reel/CZXX4CEoiwV/?utm\\_medium=copy\\_link&fbclid=IwAR2UY0rxia5D2CKYcg18S](https://www.instagram.com/reel/CZXX4CEoiwV/?utm_medium=copy_link&fbclid=IwAR2UY0rxia5D2CKYcg18S) [10.02.2022].

<sup>25</sup> "Henryk Sienkiewicz. A man – a riddle. 😊 Rich personality, combining contrasting traits, paradoxical predictions! 🏠 A traveler and homebody, a writer of serious novels and the author of humorous letters 📖📚 A man sensitive to female charms, but also a slightly malicious observer of female weaknesses. 🌸🌺 Unsure of himself, though sometimes even a conceited writer. We can do this endlessly... And we write about it, because we will soon finish the presentation of Henryk Sienkiewicz's letters in the audio version 🎧🎧 And we have a movie surprise for you! 😊 And a riddle 😊 What city does our pen master describe here? Where these "stray" pigs wander among the omnibuses in the streets 🐷🐷🐷 Where "the shop windows are admirably awesome and rich, but tastelessly decorated"... What is this city? 😊 Does anyone already know?". <https://fb.watch/b2MsOYm5dY/> [17.01.2022].

<sup>26</sup> Data referring to these materials is presented thanks to the courtesy of the Board President of WSPAK Foundation, Marcin Steczkowski.

<sup>27</sup> "On Facebook in Q2 we took action on: 6.2 million pieces of organized hate content [...], 16.8 million pieces of suicide and self-injury content [...], 34.1 million pieces of violent and graphic content [...]. On Instagram in Q2 we took action on: 367,000 of organized hate content [...], 3 million pieces of suicide and self-injury content [...], 7.6 million pieces of violent and graphic content, compared to 5.5 million pieces in Q1 2021". The last example is the most "effective". <https://about.fb.com/news/2021/08/community-standards-enforcement-report-q2-2021/> [4.02.2022].

However, company ∞Meta is aware that the context for any information put in the social media could cause some problems for AI in adequate interpretation. Some examples with pictures are presented on the information pages.<sup>28</sup>

This does not mean that actual violations of the rules have been detected and only inappropriate content has been blocked. The question remains: is social media dependent on AI control forever?

## OPTIMUM: IN SEARCH OF A COMMON PATH

What might the social world be like if designed by computer scientists and political correctness, shaped and controlled by AI?<sup>29</sup> Could it be a world without envy and violent emotions, full of rationality? Would the layer of lack of truth disappear on the Internet and would emotions be limited to just a few icons? According to one of the designers of artificial intelligence, Marvin Minsky, emotions do not differ much from thinking processes, in fact emotions are a variation of cognition. Such variant-based processing that must also be included in the algorithm is, in Minsky's opinion, a good equivalent of emotions.<sup>30</sup> An algorithm is a very simplifying formula, rigid, and it behaves in restrictive and repetitive ways as well. It sticks to principles that it does not fully understand. Years ago, Andy Clark provocatively argued that we are affected by cyborgization, regardless of our physical state, which refers to our way of thinking, using tools, technologies and applying them in our activities. All of this is responsible for the cyborgization of humans (Clark, 2005, pp. 71–72). From a slightly different perspective, in his critique of capitalism, Erich Fromm also warned about a future in which humans may become slaves and robots (1955, pp. 96, 312–313).

*Does AI create a utopian reality?* A unicorn reality. The unicorn frequents the land of happiness. Wayward and elusive, only innocence can lure him. The unicorn's horn has magical properties: it is an antidote to poison. As a mythical creature, it appears in a world that meets its conditions. And this is what the web of the future is supposed to look like: it should include no poisoned words, no false information (fake news), and no verbal or visual violence. Some 20<sup>th</sup> century thinkers, such as Erich Fromm, dreamt about shaping reality, although it may be unattainable:

Man can protect himself from the consequences of his own madness only by creating a sane society which conforms with the needs of man, needs which are rooted in very conditions of his existence. A society in which man relates to man lovingly, in which he is rooted in bonds of brotherliness and solidarity, rather than in the ties of blood and soil; a society which gives him the possibility

---

<sup>28</sup> Explanation of the AI problems in detection violations because of context: "Technology can find and remove the same content over and over. But it's a big challenge to get a machine to understand nuances in word choice or how small differences may change the context". <https://transparency.fb.com/enforcement/detecting-violations/how-enforcement-technology-works/> [4.02.2022]. However, some efforts were undertaken to improve the platforms: <https://about.fb.com/news/2021/03/more-control-and-context-in-news-feed/> [4.02.2022].

<sup>29</sup> On methods of controlling media content by AI, e.g. (Bratu, 2017; MacAvaney et al., 2019).

<sup>30</sup> <https://web.media.mit.edu/~minsky/Introduction.html> [9.02.2022].

of transcending nature by creating rather than destroying, in which everyone gains a sense of self by experiencing himself as the subject of his powers rather than conformity, in which a system of orientation and devotion exists without man's needing to distort reality and to worship idols (Fromm, 1955, p. 314).

*Is there a common path?* When talking about the future of social media, it should be considered which age groups use which specific media (Head et al., 2019; Metzler and Scheithauer, 2015), and how those in these groups use social media in general, or more specifically, how they use applications and other smartphone capabilities. Dependence on media and devices may arise in many groups. According to Manfred Spitzer, it is possible to patiently change the alternative lifestyle:

Without being available around the clock; (2) without a thousand friends we don't know whose messages reach us only when a computer wants it to; (3) And without being controlled by a device that can do many things, but also tell us what to do day and night and also spy on us better than intelligence agencies around the world have ever been able to do<sup>31</sup> (Spitzer, 2016, p. 358).

In turn, Howard Gardner wrote about 5 types of "minds of the future" that will be helpful in shaping it, so that society could develop dynamically and in interesting ways. It is not about the mind or style of thinking, but rather an attitude towards the world that is based on certain skills: "Rather than being distinct computational capabilities, they are better thought as a broad uses of the mind that can cultivate at school, in professions, or at the workplace." (Gardner, 2009, p. 4). It is also quite a utopian educational proposition to prepare future generations to cope with an increasingly technologically complex world.<sup>32</sup> The first way of thinking about action is "*the disciplined mind* has mastered at least one way of thinking" (Gardner, 2008, p. 3; more pp. 21–44, 154). "*The synthesizing mind* takes information from disparate sources, understands and evaluates that information objectively, and puts it together in ways that make sense to the synthesizer and also to other person" (Gardner, 2008, p. 3; more pp. 45–76, 155). The next attitude described by Gardner that would be desirable in the future is the creating mind: "Building on discipline and synthesis, *the creating mind* breaks new ground. It puts forth new ideas, poses unfamiliar questions, conjures up fresh ways of thinking, arrives at unexpected answers" (Gardner, 2008, p. 3; more pp. 77–101, 156).

Facebook and Instagram algorithms seem to follow the principle behind another attitude distinguished by Gardner:

*The respectful mind* notes and welcomes differences between human individuals and between human groups, tries to understand these 'others', and seeks to work effectively with them. In a world where we are all interlinked, intolerance or disrespect is no longer a viable option (Gardner, 2008, p. 3; more pp. 103–125, 157).

---

<sup>31</sup> English translation: <https://www.dasmili.eu/art/cyberkrank-wie-das-digitalisierte-leben-unsere-gesundheit-ruiniert/> [10.02.2022].

<sup>32</sup> "...With these 'minds', as I refer to them, a person will be well equipped to deal with what is expected, as well as what cannot be anticipated; without these minds, a person will be at the mercy of forces that he or she can't understand, let alone control" (Gardner, 2008, p. 2).

Metaverse algorithms track violations of etiquette, including the socially harmful behavior already mentioned. Therefore, they can be helpful in shaping the above attitude. However, the last of the “minds” proposed by Gardner is, “the ethical mind ponders the nature of one’s work and the needs and desires of society in which one lives” (Gardner, 2008, p. 3; more pp. 127–151, 158). This may turn out to be hard to achieve, not only on social media, but in the real world in general.

In conclusion, it is not known whether AI-shaped media will allow us to freely express opinions and feelings in the future. For now, the algorithm cannot read images and texts in context, which, for example, ∞Meta knows very well.<sup>33</sup> It is not just about knowing an extremely elaborate library; an algorithm may have huge amounts of content in its memory, yet it is not able to make comparisons on its own. It is also necessary to be able to ask a question related to this content. The human mind itself asks questions (Gadamer, 2000, pp. 46–48), and AI is not very good at it just yet.

One may wonder what the consequences will be of the two shortcomings that are important in the design of Artificial Intelligence. First, it is a simplified understanding of emotions, as seen by Minsky, who worked for many years in creating AI. According to Minsky, emotions are actually intellectual operations, which is a great simplification, and the consequences of such “emotional sterility” of AI are unknown. If we take into account that women are almost absent in the area of designing AI,<sup>34</sup> this excludes a large part of humanity from that process. Does this signify a new form of patriarchy? Well, at least in social media. Not only is gender imbalance written in the creation of AI, Joy Buolamwini, founder of the Algorithmic Justice League,<sup>35</sup> in her research, pointed to racial bias in the process of AI design.<sup>36</sup> Moreover, it’s good to remember that the creation of AI is based in the English language. Human languages are always rooted in culture. In this case, in a part of Western culture. Therefore, one can ask about culture bias throughout the process.<sup>37</sup> Could the future wave of globalization, induced by the online activities of societies and the popularization of the Internet, be more harmful than beneficial to cultural diversity?

---

<sup>33</sup> <https://transparency.fb.com/enforcement/detecting-violations/how-enforcement-technology-works/> [4.02.2022].

<sup>34</sup> As Toby Walsh noted, only about 10% of AI designers are women (Walsh, 2018, p. 161): “Gender imbalance is detrimental to the process of AI development. For this reason, there are unasked questions and unanswered problems.” (Walsh, 2018, p. 162).

<sup>35</sup> <https://www.ajl.org/> [5.04.2022].

<sup>36</sup> Buolamwini’s research on facial recognition technologies, and her activism for the development of unbiased technologies, lead to some serious decisions for AI design corporations such as IBM: “The Algorithmic Justice League commends this decision as a first move forward towards company-side responsibility to promote equitable and accountable AI. This is a welcome recognition that facial recognition technology, especially as deployed by police, has been used to undermine human rights, and to harm Black people specifically, as well as Indigenous people and other People of Color (BIPOC)” [5.04.2022].

<sup>37</sup> Analogue technologies, such as the steam engine or the sewing machine, seemed to be neutral, and easily applicable in any society. Digital technologies, especially used for designing AI, seem not to be neutral. The authors are far from seeing the process as a case of “cultural imperialism;” we would like to show some potential spheres of exclusion or serious changes. One can ask, what happens to complex native languages such as Thai or Chinese, when the online environment has a multilingual character? (on the Thai language: Troyer R. (2012). English in the Thai linguistic Netscape. *World Englishes*, 31, 1, 93–112).

## REFERENCES

- Bettman, D., Oksanowicz, P. (2021). *Technologiczne magnolie. Gdy większość z nas uwierzy, że dzięki technologiom zmienimy świat na lepsze*. Warszawa: Wydawnictwo Naukowe PWN.
- Bratu, S. (2017). The inexorable shift towards an increasingly hostile cyberspace environment: The adverse social impact of online trolling behavior. *Contemporary Readings in Law and Social Justice*, 9, 2, 88–94.
- Brougham D., Haar, J. (2018). Smart technology, Artificial Intelligence, robotics, and algorithms (STARA): Employees' perceptions of our future workplace. *Journal of Management & Organization*, 24, 2, 239–257.
- Business Wire. (2017). *\$5.43 Billion Consumer Robot Market 2017 – Industry Trends, Opportunities and Forecasts to 2023*. Retrieved from: Business Wire, <https://www.business-wire.com/news/home/20171219005736/en/5.43-Billion-Consumer-Robot-Market-2017> [2.02.2022].
- Castelo, N., Bos, M.W., Lehmann, D.R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, 56, 5: 809–825.
- Clark, A. (2003). *Natural-born Cyborgs: Minds, technologies, and the future of human intelligence*. New York: Oxford University Press.
- Copp D. (2020). Just too different: normative properties and natural properties. *Philosophy Studies*, 177, 263–286.
- Davenport, T., Guha, A., Grewal, D., Bressgott, T. (2020). How Artificial Intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48, 1, 24–42.
- Diamandis, P.H., Kotler, S. (2020). *The future is faster than you think. How converging technologies are transforming business, industries and our lives*. New York: Simon & Schuster.
- Dietvorst, B.J., Simmons, J.P., Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144, 1, 114–126.
- Duignan, P., Gann, L.H. (1995). *Political correctness. A critique*. Redwood City, CA: Stanford University Press.
- Dunbar, R. (2010). *How many friends foes one person need?: Dunbar's number and other evolutionary quirks*. Cambridge, MA: Harvard University Press.
- Eco, U. (2004). *On beauty*. Translated from the Italian by A. McEwen. London: Seeker & Warburg.
- Fortuna, P. (2021). *Optimum. Idea cyberpsychologii pozytywnej*. Warszawa: Wydawnictwo Naukowe PWN.
- Fortuna, P., Gorbaniuk, O. (2022). What is behind the buzzword for experts and laymen: Representation of “Artificial Intelligence” in the IT-professionals' and non-professionals' minds. *Europe's Journal of Psychology*, 18(2), 217–218.
- Fromm, E. (1955). *The sane society*. Greenwich: Fawcett Publications, Inc.
- Gadamer, H.-G. (2000), *Rozum, słowo, dzieje. Szkice wybrane*. Translated by M. Łukasiewicz, K. Michalski. Warszawa: Wydawnictwo Naukowe PWN.
- Gardner, H. (2008). *Five minds for the future*. Boston, MA: Harvard Business Press.

- Germann, M. (2018). *Delegated investing and the role of investment outcomes. Quality, and Trust*. Retrieved from: Universitätsbibliothek Mannheim, [https://madoc.bib.uni-mannheim.de/47040/1/Germann\\_M\\_Dissertation.pdf](https://madoc.bib.uni-mannheim.de/47040/1/Germann_M_Dissertation.pdf) [4.03.2021].
- Gladden, M.E. (2016). *Posthuman management*. Indianapolis: Synthypnion Press.
- Head, A.J., Fister, B., MacMillan, M. (2020). *Information literacy in the age of algorithms. Study report*. Harvard: Knight Foundation-Harvard Graduate School.
- Hjorth, L., Hinton, S. (2019). *Understanding social media*. London: Sage.
- Ho, A. (2019). Deep ethical learning: taking the interplay of human and Artificial Intelligence seriously. *Hastings Center Report*, 49, 1, 36–39.
- Inbar, Y., Cone, J., Gilovich, T. (2010). People's intuitions about intuitive insight and intuitive choice. *Journal of Personality and Social Psychology*, 99, 2, 232–247.
- Kirsten, M. (2019). Algorithms raise questions about ethics and accountability. *MIS Quarterly Executive*, 6, 129–142.
- Kramer, A.D.I., Guillory, J.E., Hancock, J.T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *PNAS*, June 17, 2014, 111 (24), 8788–8790; first published June 2, 2014.
- LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.
- Lem, S. (1988). *Filozofia przypadku*. Kraków: Wydawnictwo Literackie.
- Levinson, P. (2013). *New media*. Boston: Pearson.
- Liang, X., Straub, J. (2021). Deceptive online content detection using only message characteristics and a machine learning trained expert system. *Sensors*, 21, 7083.
- Longoni, C., Bonezzi, A., Morewedge, C.K. (2019). Resistance to medical Artificial Intelligence. *Journal of Consumer Research*, 46, 629–650.
- Lovink, G. (2016). *Social media abyss. Critical Internet cultures and the force of negation*. Cambridge: Polity.
- MacAvaney, S., Yao, H-R., Yang, E., Russell, K., Goharian, N., Frieder, O. (2019). Hate speech detection: Challenges and solutions. *PLoS ONE*, 14, 8, e0221152.
- Martin, K. (2019). Designing ethical algorithms. *MIS Quarterly Executive*, 18, 2, 129–142.
- Metzler, A., Scheithauer, H. (2015). Adolescent self-presentation on Facebook and its impact on self-esteem. *International Journal of Developmental Science*, 3–4, 135–145.
- Minsky, M. (1996). *The Emotion Machine* [draft]. Retrieved from: MIT Media Lab, <https://web.media.mit.edu/~minsky/Introduction.html> [4.02.2022].
- Motycka, A. (2005). *Rozum i intuicja w nauce. Zbiór rozpraw i szkiców filozoficznych*. Warszawa: Eneteia. Wydawnictwo Psychologii i Kultury.
- Nisbett, R.E. (2003). *The geography of thought: how Asians and Westerners think differently... and why*. New York: Free Press.
- Schank, R. (2020). What is AI? *Journal of Artificial General Intelligence*, 11, 2, 89–90.
- Spitzer, M. (2016). *Cyberchoroby. Jak cyfrowe życie rujnuje nasze zdrowie*. Translated from German M. Guzowska. Słupsk: Wydawnictwo Dobra Literatura.
- Svedholm, A.M., Lindeman, M. (2013). The separate roles of the reflective mind and involuntary inhibitory control in gatekeeping paranormal beliefs and the underlying intuitive confusions. *British Journal of Psychology*, 104, 303–319.

- Thomé, S., Härenstam, A., Hagberg, M. (2011). Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults – a prospective cohort study. *BMC Public Health*, 11, 66. Doi: 10.1186/1471-2458-11-66.
- Troyer, R. (2012). English in the Thai linguistic netscape. *World Englishes*, 31, 1, 93–112.
- Vaidhyanathan, S. (2018). *Antisocial media. How Facebook disconnects us and undermines democracy*. New York: Oxford University Press.
- Verma, M. (2014). Editorial expression of concern: experimental evidence of massive-scale emotional contagion through social networks. *PNAS*, July 22, 111(29), 10779 [first published July 3, 2014].
- Walsh, T. (2017). *It's Alive!: Artificial Intelligence from the logic piano to killer robots*. Collingwood: Schwartz.
- Youyou, W., Kosinski, M., Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112, 4, 1036–1040.

Submitted: 15.02.2022

Reviewed: 1.04.2022

Revised: 8.04.2022

Accepted: 10.04.2022

Published online: 30.06.2022