Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/CLSR

Computer Law & Security Review

Prohibited artificial intelligence practices in the proposed EU artificial intelligence act (AIA)

Rostam J. Neuwirth[1]

Faculty of Law, University of Macau, Avenida da Universidade, Taipa, Macao 999078, China

ARTICLE INFO

Keywords:
Law
AIA
Artificial intelligence
Prohibited AI practices
Subliminal AI systems
Vulnerabilities
Social scoring systems
'Real-Time' remote biometric identification systems

ABSTRACT

Artificial intelligence (AI) now forms a more and more important part of human lives. After years focussed on the development of AI, the initial hype about its many expected benefits has gradually given way to rising ethical concerns about its inherent risks and dangers. Efforts to confront and contain the most serious risks related to AI have now prompted a number of legislative or regulatory proposals at the national, regional and global level. One of the most comprehensive regulatory initiatives is the European Union's proposal for an Artificial Intelligence Act (AIA), which was released in April 2021 with a view towards establishing a legal framework for trustworthy AI. To this end, the draft AIA pursues a proportionate horizontal and risk-based regulatory approach to AI, broadly classifying AI into the categories of unacceptable risks, high risks and low or minimal risks. The unacceptable risks are those that are deemed to contravene European Union values, and therefore, they are considered to be 'prohibited AI practices' by Article 5 AIA. The prohibited AI practices are classified into four categories, namely 1) AI systems deploying subliminal techniques; 2) AI practices exploiting vulnerabilities; 3) social scoring systems; and 4) 'real-time' remote biometric identification systems. The proposed regulatory approach, however, appears problematic given the four categories' inherent interrelatedness and the numerous possibilities for their mutual combination and entwinement. It is also problematic from the perspective of the human mind, as each of the four categories alone allows for the manipulation of human thought and behaviour, thereby endangering freedom of thought and other fundamental rights. In the context of the proposed AIA, both aspects give rise to unknown and unsolved conundrums that create difficult regulatory challenges that raise the necessity to also look at the wider implications of these technologies for the entire legal system. As these conundrums often find their expression in paradoxes and oxymora, this article calls for a wider interdisciplinary debate and advocates a different regulatory strategy using these concepts to transcend the limitations inherent in dualistic or dichotomous modes of legal thinking.

E-mail address: rjn@um.edu.mo

[1] The author expresses his gratitude for the constructive comments and helpful suggestions received during the peer review process of the Computer Law and Security Review and acknowledges the financial support provided by the University of Macau [MYRG2015-00222-FLL].

https://doi.org/10.1016/j.clsr.2023.105798
2212-473X/

Introduction

Forbidden fruits are the sweetest.[1]

Several regulatory proposals targeting artificial intelligence (AI) have recently been formulated and are currently being simultaneously processed at national, regional and global levels. This trend indicates that in parallel to the race for the development of AI, there is also now a worldwide race for its regulation.[2] The regulatory competition illustrates the concurrence of the growing ethical concerns about the impact of AI on societies, the environment, ecosystems and human lives (including the human mind) that are slowly replacing the initial hype about the many expected benefits of AI.[3] It also confirms earlier assessments according to which the use of ethical principles alone may not suffice to contain the risks that arise from the development, deployment and governance of AI.[4] It is noteworthy that AI has been found to challenge 'the identity and autonomy of both individuals and nations' because it offers increased accessibility to knowledge.[5]

While the risks and ethical concerns must be taken seriously, they should also not let the pendulum swing in the opposite direction to an 'AI-as-existential threat narrative'.[6] Instead, a different and more nuanced approach must be sought that allows the reaping of maximum benefits while at the same time avoiding the hazardous pitfalls of AI.[7] In this respect, the vagueness of concepts and definitions of AI, in particular, should be duly avoided.[8] Such an approach could be described as paradoxical given the frequent conceptual encounters with rhetorical figures of speech known as paradoxes and oxymora. In this context, it is worth mentioning that on 21 April 2021, the European Union released its proposal for an EU Artificial Intelligence Act (AIA).[9] The AIA pursues a horizontal regulatory approach to AI that provides rules for all kinds of AI, rather than a vertical approach that focuses only on one specific aspect of AI. Behind the purpose of the AIA stands the EU's aspiration to be 'a global leader in the development of secure, trustworthy and ethical Artificial Intelligence'.[10]

In a global comparison, the planned EU AIA to date constitutes the most comprehensive single regulatory effort to address both the opportunities of AI and the ethical concerns over the various risks and dangers related to it. The AIA follows a risk-based approach to address the ethical concerns, broadly classifying AI risk into three levels, namely 1) unacceptable risk, 2) high risk, and 3) low or minimal risk. The category of unacceptable risks is to be understood to comprise 'all those AI systems whose use is considered unacceptable as contravening Union values, for instance, by violating fundamental rights'.[11] Consequently, the AIA also proposes that these AI practices should be prohibited, as set out in Title II, Art. 5 AIA 'Prohibited AI Practices'.

Broadly speaking, the list of prohibited AI practices laid down in Article 5 AIA covers four distinct kinds of AI. The first of these is subliminal AI practices, which are those that have 'a significant potential to manipulate persons through subliminal techniques beyond their consciousness'. The second set are practices that 'exploit vulnerabilities of specific vulnerable groups such as children or persons with disabilities'. For both these subliminal and exploitative AI practices, it is required that they are used 'in order to materially distort' a person's behaviour 'in a manner that causes or is likely to cause that person or another person physical or psychological harm'.[12] The third set comprises so-called 'social scoring systems', which are systems used by 'public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics'. Finally, the fourth category consists of biometric AI systems that involve the 'use of "real-time" remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement'.[13]

Since the publication of the draft AIA and during the ongoing legislative process, the initiative was widely welcomed; however, several shortcomings and problems have already been outlined regarding the AIA in general and specific norms in particular. On the positive side, the risk-based approach that plans to restrict the use of AI systems only when they are likely to pose high risks to fundamental rights and safety was welcomed.[14] On the negative side, the Draft AIA has been described as a 'lengthy, sometimes opaque document' or a patchwork lacking consideration for its interaction with other laws and regulations as well as its future enforcement.[15] More specifically, the AI definition was criticised as being too narrow and the harm requirement in the context of manipulative practices was qualified as entailing a range of problematic loopholes and as framed too narrowly.[16] Other specific criticisms include the absence of the treatment of liability for damages,[17] the need for a stricter approach to biometric surveillance systems[18] or a call for a more profound and interdisciplinary approach to the regulation of subliminal AI systems based on the many links between law and neuroscience.[19]

[1] March (1910), p. 432.
[2] Smuha (2022).
[3] UNESCO (2021), Recital 1, and OECD (2019).
[4] Mittelstadt (2019), p. 501; Munn (2022), p. 1.
[5] Bryson (2018), p. 1.
[6] Galanos (2019), p. 421.
[7] Brundage, Bryson (2016), p. 1.
[8] Nordström (2022), p. 1706.
[9] European Commission, *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, COM (2021) 206 final (21 April 2021) [AIA].
[10] European Council, Special meeting of the European Council (1 and 2 October 2020) – Conclusions, EUCO 13/20 (2 October 2020) para. 13.
[11] Explanatory Memorandum of the AIA, EU AIA, *supra* note 9, p. 12 (pt. 5.2.2).
[12] Art. 5(1) lit. a) and b) AIA, *supra* note 9.
[13] Art. 5(1) lit. d) EU AIA, *supra* note 9.
[14] Ebers et al. (2021), p. 589.
[15] Mökander et al. (2022), p. 752; Veale and Zuiderveen Borgesius (2021), p. 26.
[16] Bryson (2022); Veale and Zuiderveen Borgesius (2021), pp. 4–5; Smuha et al. (2021), p. 21.
[17] Raposo (2022a), p. 108.
[18] Barkane (2022), p. 147.
[19] Neuwirth (2023), p. 2.

Against this backdrop, this article discusses the four categories of prohibited AI practices, which will be abbreviated as subliminal practices, exploitative practices, social scoring systems and real-time remote biometric identification systems. Each of these alone poses serious regulatory challenges that are aggravated by their intrinsic mutual connections as part of a general trend towards convergence of products, industries and technologies. It is argued that their inherent interrelatedness also challenges the overall equilibrium and integrity of a legal system and, therefore, warrants a regulatory approach that is at the same time specific and comprehensive. Finally, they also display a complex nature as captured by the important debate on technology neutrality, namely whether technologies are neutral or not.

The article advocates a mode of legal reasoning that transcends the traditional mode of dualistic reasoning and binary logic by better connecting the present legislative process of the AIA with the general debates about technology governance. These debates often and increasingly feature profound paradoxes and surprising oxymora. The term 'artificial intelligence' itself has been qualified as an oxymoron.[20] Paradoxes of technology include the Amara Paradox, which states that in the regulation of a new technology, there is a trend to overestimate its effect in the short run and to underestimate the effect in the long run.[21] This hype has also been expressed via related terms qualified either as an oxymoron, such as 'cyborg', or a paradox, such as 'technological singularity'.[22] A similar problem has been identified as the Collingridge dilemma, which holds that it is difficult to predict the consequences of a technology in its early phase, but once undesirable consequences are discovered, it will be difficult to change the initial regulatory path chosen.[23] For this reason a legal debate should be initiated long before a technology has been put in use.[24] Likewise, a paradox was found for the question of whether to strictly regulate digital technologies or not to regulate them at all.[25] A general regulatory paradox has also been identified, according to which a regulatory initiative can 'achieve an end precisely opposite to the one intended'.[26] Against the backdrop of these paradoxes and oxymora, the article sets out to explore and lay the foundations for a more balanced, dynamic and holistic approach to be used for the regulation of AI in general and prohibited AI practices in particular.

Prohibited AI practices in the EU artificial intelligence act

The first mention of 'prohibited AI practices' is in the Preamble of the proposed AIA, which explains the wider context and underlying reasons for the existence of this category:

> (15) Aside from the many beneficial uses of artificial intelligence, that technology can also be misused and provide novel and powerful tools for manipulative, exploitative and social control practices. Such practices are particularly harmful and *should be prohibited* because they contradict Union values of respect for human dignity, freedom, equality, democracy and the rule of law and Union fundamental rights, including the right to non-discrimination, data protection and privacy and the rights of the child.
>
> (16) The placing on the market, putting into service or use of certain AI systems intended to distort human behaviour, whereby physical or psychological harms are likely to occur, *should be forbidden*.[27]

The implementation of these considerations is concretised in Title II by way of Article 5 AIA that roughly distinguishes the four categories of prohibited AI practices. Concerning these four, the first three categories (in order) are 'subliminal or manipulative practices', 'practices exploiting vulnerabilities' and 'social scoring systems', all of which can be regarded as prohibited in their entirety. The fourth category concerns the use of '"real-time" remote biometric identification systems in publicly accessible spaces', which is also prohibited except for specific law enforcement purposes. Despite their interrelatedness and overlap in the use of underlying technologies, the current analysis follows the order in which they are listed in the AIA.

Subliminal AI systems

The first category of prohibited AI practices contains so-called subliminal AI practices, and this raises several complex, closely intertwined and interdisciplinary questions, including the legal qualification of subliminal AI techniques, the scope of harm, the sensory threshold of subliminal as opposed to supraliminal perception for each of the individual senses and their mutual interplay, the dangers related to existing and future AI techniques, and the enforcement of the provision.[28]

To begin with, the AIA does not define the term 'subliminal', but this term generally refers to perception below the threshold of awareness.[29] While the first scientific experiments on the effects of subliminal visual stimulation on dreams were carried out more than a hundred years ago,[30] it was in the 1950s that the issue received wider attention in connection with subliminal advertising, popularly known as the popcorn and Coca-Cola experiment. It is remarkable that even back then, the author of the book *The Hidden Persuaders* was warning about the danger of subliminal techniques for the manipulation not only of consumers but also of elections.[31] Most of all, he concluded that the most serious offence related to subliminal manipulation lies in the possibility of invading the privacy of our minds.[32] Other historical precedents involving subliminal messages include subliminal

[20] Gidley (2017), p. 99; Svensson (2021), p. 8.
[21] Davenport (2018), p. 7 (citing Roy Amara).
[22] Siivonen (1996); Gozman, Liebenau, Ferris (2019); Eden (2012); Nicolescu (2016), p. 43.
[23] Collingridge (1980), p. 11.
[24] Eisenberger (2017), p. 149.
[25] Duque, Zuluaga Torres (2020), p. 8.
[26] Sunstein (1990), p. 407.
[27] Recitals 15 and 16 EU AIA, *supra* note 9 [*Italics* added].
[28] Neuwirth (2023).
[29] Klein (1966), p. 726.
[30] Pötzl (1960).
[31] Packard (2007).
[32] *Ibid*, p. 240.

messages encoded in background music in shops to prevent shoplifting, as well as legal actions concerning whether the suicides of teenagers were caused by masked subliminal lyrics embedded in rock and heavy metal music.[33] It was noted that in the US, 'the public's concern with the use of subliminals was fervent at one point, interest in the subject eventually subsided' and never led to a concrete legislative act.[34] It was even argued that declining interest in the matter was due to a conspiracy of advertisers and news outlets.[35]

In view of these uncertainties surrounding past experiments, it is essential for the present regulatory proposal to use a sound scientific basis for the formulation and enforcement of the planned prohibition of subliminal AI systems. Lessons from the handling of various past subliminal experiments, such as those involving popcorn and Coca-Cola, must be learned to prevent a sound scientific debate from being devalued by regarding it as a hoax, myth or conspiracy theory.[36] This is important, especially in the context of subliminal techniques that – due to their 'hidden' character – may give rise to wild speculations and conspiracies, such as the government using subliminal messages on TV to influence the public.[37] Although it was observed that numerous conspiracies have turned out to be true,[38] the term 'conspiracy theory' can also be given a pejorative meaning to halt serious questions, silence criticism or to prevent sound scientific or legal inquiries into events or issues that remain unsolved.[39] Yet it was also found that the label of certain debates as conspiracy theories 'does not reduce people's belief in them'.[40] The label may also lead to a greater acceptance of such intrusions in a person's cognitive liberty, as the fear of being branded as a 'conspiracy theorist' leads to a greater acceptance of such risks. A similar problem arises in the context of nudging, which has been defined as 'any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives'.[41] The example of subliminal messages for nudging has been used to ponder whether nudging people, even when used for beneficial purposes, still impinges upon a person's freedom of choice.[42] Simply put, it begs the question of whether there is a moral difference between the use of nudges as opposed to subliminal messages.[43] In view of this complexity and highlighting the limitations of purely binary approaches to such problems, it is also no surprise that it has been argued that the phrase 'conspiracy theory' is an oxymoron.[44]

Overall, these questions show and add support to the need for a broad and interdisciplinary debate based on sound scientific findings. This is particularly true since the initial experiments with subliminal perception conducted by Otto Pötzl in the early 20th century have now been widely confirmed and their application expanded.[45] Hence, the question of subliminal perception is no longer merely of interest to marketing. It has also been applied, for instance, to questions of cybersecurity.[46] A concrete example of cybersecurity concerns resulting from subliminal and supraliminal messages is provided by the so-called 'social engineering attacks', such as phishing, that refer to means of 'manipulating people into performing actions or divulging confidential information'.[47] Particularly in connection with new emerging technologies, the effectiveness of subliminal cues was confirmed for 'fundamental aspects of metacognition such as higher cognitive and emotional metaabilities, affective and behavioural regulation, and academic achievement'.[48] These new technologies and their creative combination can be expected to open many more fields for their application.[49]

Therefore, the old controversy about the effectiveness of subliminal stimuli in terms of their effects on the human mind and behaviour should now be regarded as settled by scientific studies confirming that subliminal stimuli are indeed effective. Especially the claim that the popcorn and Coca-Cola experiment was a hoax has been debunked, as two studies have demonstrated that subliminal priming of a brand name of drink does affect the participant's choice for the primed brand.[50] It was also established that preferences towards unfamiliar drink brands may be influenced through subliminal conditioning.[51]

These experiments confirm the important legal precedent for the planned prohibition of subliminal AI set by members of both the Council of Europe and the European Union, as well as other jurisdictions[52] that decided in 1989 to prohibit all forms of subliminal advertising.[53] Since then, it has also been confirmed that subconscious communication that takes the form of subliminal messages is often far more effective than seeking to directly address, provoke and cause the desired effect. The reason given is that in the realm of subliminal perception, the mind cannot oppose subliminal messages because it does not know about them and, thus, does not recognise them.[54] Overall, the many findings have been well summarised in the sense that 'the literature and evidence supporting subliminal information theory is robust' and that subliminal information is not only processed and retained but also '*acted upon*'.[55] At the same time, a recent study showed that subliminal methods can also be used to forget unwanted memories.[56] While this can be useful for dealing with traumatic experiences, it

[33] Goodkin and Phillips (1981), p. 1078; and Dee (1994), p. 3.
[34] Blen (1992), p. 877.
[35] Key (1976), p 8.
[36] Trank (1978); Weinberger, Hardaway (1990); Pratkanis (2007).
[37] Douglas, van Prooijen, Sutton (2022), p. 585.
[38] Sunstein (2014), p. 4.
[39] Neuwirth (2021), p. 842.
[40] Douglas, van Prooijen, Sutton (2022), p. 585.
[41] Thaler and Sunstein (2008), p. 6.
[42] Hausman, Welch (2010), p. 131.
[43] Bovens (2008), p. 207.
[44] Singleton (2008), p. 31.
[45] Pötzl (1917); Pötzl (1960).
[46] Zhu, Carpenter, Zeng (2022).
[47] Ilangakoon, Abeywardena (2018).
[48] Drigas, Mitsea, Skianis (2022), p. 175.
[49] Elgendi (2018).
[50] Karremans, Stroebe, Claus (2006), p. 792.
[51] Amd, Passarelli (2020), p. 1.
[52] Neuwirth (2023), pp. 46 et seq.
[53] Art. 13(2) European Convention on Transfrontier Television, European Treaty Series – No. 132 (5 May 1989) and Art. 10(3) European Economic Communities, Council Directive 89/552/EEC (Television Without Frontiers Directive), OJ L 298/23–30 (17 October 1989).
[54] Tanasic (2021), pp. 143, 146.
[55] Taylor (2007), p. 32 [*Italics* added].
[56] Zhu, Anderson, Wang (2022).

can also constitute negative forms of brainwashing with respect to identity theft. Other commentators in the field of neuromarketing have concluded in relation to the effectiveness of subliminal priming that it is no longer a question of *whether* but of *how* it influences consumer behaviour.[57] Perhaps summarising the current state of the scientific debate, it was stated that 'theoretical studies and experiments demonstrated that independently of a person's confidence about "object's attributes" it is possible to change his preferences, attitudes, and impression; formation can be developed "outside" of an individual's beliefs'.[58] More broadly still, a similar summary was expressed in another way in a statement about artificial intelligence that it now allows for a move from 'programming computers to programming people'.[59]

These findings about subliminal manipulation raise more serious questions, especially about neuroscientific findings or related disciplines that hold that a larger portion of cognitive functions is carried out unconsciously rather than consciously. Leonard Mlodinow reported that only 5% of our entire cognitive functions are conscious, while the remaining 95% are beyond our awareness but still exert an important influence on our lives.[60] Another source puts the percentage of awareness even lower, with only 2% reserved for conscious awareness and the remaining 98% for unconscious functions.[61] It is likely that a larger portion of cognitive functions is handled unconsciously but that the percentage is dynamic and varies depending on the type of functions being carried out, since in combination, they 'each play distinct and complementary roles'.[62] Most importantly but not surprisingly, it has also been shown that the conscious or unconscious processing of stimuli varies between individuals.[63] In short, this means that a larger portion of cognitive functions is carried out subliminally or unconsciously. If this is true, it would signify that subliminal forms of manipulation can alter a larger part of those functions that determine both our thoughts and actions.

The various studies and scientific findings can thus be taken as scientific support for the planned prohibition in Article 5(1) lit. a) AIA, which in detail reads as follows:

The following artificial intelligence practices shall be prohibited:

> (a) the placing on the market, putting into service or use of an AI system that *deploys subliminal techniques beyond a person's consciousness* in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm [...].[64]

However, the concrete formulation chosen prompts several questions. First, it raises the issue of what constitutes an AI system that deploys subliminal techniques. In this regard, several technologies that make use of or can make use of such techniques already exist. In a concrete example of a relatively simple subliminal technique, videos of hotel rooms were embedded with a smiling face emoji as a subliminal message, and this was found to have a significant effect on consumer selection of hotels.[65] In another experiment, subliminal flashes of national flags showed changes in political attitudes and behaviour.[66] Although these technologies deployed subliminal techniques, it does not mean that they automatically qualify as AI. They will likely have to be combined with other technologies applying machine learning approaches or other criteria falling within the definition of AI (Art. 3 AIA).

There already exist complex and more intrusive forms of subliminal techniques in the form of 'mind reading', 'dream hacking' and 'brain spyware' technologies. For instance, there are reports that the commercial use of dream incubation (i.e., the presentation of stimuli before or during sleep to affect dream content) is being tested in numerous marketing studies to find 'new ways to alter and drive purchasing behaviour through sleep and dream hacking'.[67] Equally, the effectiveness of brain spyware (i.e., 'a software intentionally designed to detect private information') in combination with brain-computer interfaces has already been demonstrated in the lab and may be used to 'deduce private information such as banking information, recognized faces, PIN numbers, location of residency and month of birth from the brains of BCI users by showing them visual stimuli and using a machine learning model to detect familiar information based on the brain's response'.[68] In combination, these different 'mind reading' technologies show real potential to become effective tools of commercial manipulation and to be used to get access to personal data as well as private thoughts 'without the consent of the individual affected or even the awareness that such information is being taken'.[69]

Closely related to or synonymous with subliminal AI systems are so-called 'dark patterns'.[70] These are defined as 'deceptive elements that are intentionally crafted to make the users do actions that they wouldn't do otherwise'.[71] These dark patterns exist in various forms and allow for the manipulation of a user interface by 'embedding subliminal stimuli to influence and encourage users to purchase more products (unethical marketing)'.[72]

Ultimately, many more manipulative subliminal and supraliminal AI practices can be conceived of which will be complemented by the development of new technologies related to the sensory organs (e.g., electronic noses, tongues and skins) that are deemed useful for both the development of robotics and AI.[73] The senses and subliminal stimuli are also said to play an important role in virtual and augmented reality systems, which have also been found to have a strong in-

[57] Chartrand and Fitzsimons (2011), p. 3.
[58] Madan, Rosca, Bucovicean (2021), p. 301.
[59] Helbing et al. (2019), p. 76.
[60] Mlodinow (2012), p. 34.
[61] Wehling (2016), p. 43.
[62] Ramey, Yonelinas, Henderson (2019), p. 71.
[63] Capa, Bouquet (2018), p. 1.
[64] [*Italics* added].
[65] Hsu and Chen (2020), p. 200.
[66] Ferguson et al. (2009), p. 75.
[67] Horowitz, Stickgold, Zadra (2021).
[68] Martinovic et al. (2012), p. 145; Steinhagen, Kettani (2020), p. 78.
[69] UNESCO (2020), p. 15.
[70] European Commission (2022), p. 83.
[71] Cara (2019), p. 105.
[72] Albarrak, Metatla, Roudaut (2021), p. 1.
[73] Shih et al. (2020); Tan, Xu (2020).

fluence not only on the purchasing decision-making process but also on political propaganda and society as a whole.[74] Furthermore, their creative combination with related technologies, including but not limited to brain–computer interfaces (BCIs), functional magnetic resonance imaging (fMRI), big data, the Internet of Things, blockchain, robotics, eye-tracking, and other technologies, tools and applications, likely results in new opportunities for manipulative practices.[75] In sum, subliminal AI techniques fall into the broad category of the manipulation of the mind and behaviour, which has been found to pose serious risks for individuals and communities and 'can cause direct harm, violate the target's autonomy and treat persons as things not ends in themselves'.[76]

Another important aspect concerns the exact line of distinction between subliminal and supraliminal perception, that is, perception below and perception above the level of awareness. These phrases should also be clarified in their meaning with regard to the different sensory organs involved, as multisensory or synaesthetic perception is the rule in humans.[77] The need to look at and study the senses in their combined effects is also supported by the rapid development of an AI-driven creation of a multisensory augmented and virtual reality.[78]

The use of the term 'subliminal' has been criticised because manipulation often combines both levels of perception.[79] For instance, certain subliminal techniques, such as rapidly flashed images or backward-masked music, could be combined with manipulative techniques operating at the supraliminal level, such as fake news or deep fakes. The latter have also been identified as powerful means to manipulate a person's mind and behaviour.[80] However, given its important function as a legal precedent in the regulation of advertising, it could be argued that it would be better to replace 'subliminal' with the term 'transliminal'.[81] This idea is also supported by the finding that no absolute thresholds for subliminal perception exist.[82] The main reason is that 'thresholds for a given stimulus vary both *intra* and *interpersonally*'.[83]

The prohibition of subliminal practices has been equally criticised for requiring 'manipulative intention on the part of the person or entity that develops, launches in the market or professionally uses these AI systems' since this remains difficult to prove.[84]

Finally, there remains the important question of the scope of harm, which is limited to 'physical or psychological harm' in the current draft proposal. This scope was considered to be 'under-protective', and therefore, should be widened to include other significant and potentially more serious harms, such as 'financial and economic harms, cultural harms, harms of recognition and autonomy harms, but also collective and societal harms'.[85] Moreover, novel forms of 'cumulative harms', that is, harms that accumulate over time as more units are consumed, should also be included, especially as they can be reinforced by their impact on individuals' environments with hyper-personalization, engagement and 'dwell' metrics and their impact on children.[86] For these reasons, the present provision prohibiting subliminal AI techniques represents an important step towards the regulation of the potential risks of AI and must be considered as a priority in the EU AIA as well as in the context of attempts at the global level, such as the UNESCO Recommendation.[87]

Practices exploiting vulnerabilities

The second category of prohibited AI practices concerns those that exploit 'any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm'. The principal motive behind this provision is 'vulnerability', which is not exhaustively defined but is merely illustrated by the examples of specific vulnerable groups, such as children or persons with disabilities.

Some interpretative help can be received from the Unfair Commercial Practices Directive (UPD), which contains a similar provision in which vulnerability is understood as 'physical infirmity, age or credulity'.[88] In this regard, the proposed protected characteristics have been criticised as being too limited and unjustified because they exclude other characteristics, such as race, sex, religion or ethnicity, and this is the reason there was a suggestion to expand them to all the characteristics protected under EU equality law as laid down in Article 21 of the EU Charter on Fundamental Rights.[89]

While age may constitute a relatively easy standard to assess, other states of mind or body are more difficult to establish and are not fixed. The reason is that every person potentially or actually has a weak point, or '*locus minoris resistentiae*',[90] at least at some point in time, which makes everyone particularly vulnerable. As in the case of subliminal AI, it is the combination of different technologies and big data that makes it easier to detect and exploit vulnerabilities. For instance, the use of wearable health devices has been found capable of posing threats in this regard as the health data collected may be used by companies for commercial purposes or targeted advertising.[91] So-called intensified data marketing practices have already been identified that combine and rely on manifold technologies and apps, including Radio Frequency Identification (RFID) devices to track movements, real-time target advertising and personal recommendations in politi-

[74] Quattelbaum et al. (2022), p. 43; Henz (2022), p. 4.
[75] Neuwirth (2023), pp. 25–26 and 94.
[76] Uuk (2022).
[77] Sagiv, Dean, Bailes (2009), p. 294.
[78] Neuwirth (2023), pp. 62–68.
[79] Uuk (2022).
[80] Neuwirth (2022a), p. 838.
[81] Neuwirth (2023), p. 88.
[82] Rathus (2020), p. 64.
[83] Smith and McCulloch (2012), p. 551 [*Italics* added].
[84] Raposo (2022a), p. 6.
[85] Smuha et al. (2021), p. 21 [footnote omitted]; Ebers et al. (2021), p. 592; Uuk (2022).
[86] Veale and Zuiderveen Borgesius (2021), p. 102.
[87] UNESCO (2021).
[88] Art. 5 UPD, Directive 2005/29/EC (Unfair Commercial Practices Directive), OJ L 149/22–39 (11 June 2005).
[89] Smuha et al. (2021), p. 22.
[90] Schiavo et al. (2014), p. 553.
[91] Somani et al. (2021), p. 7.

cal advertising, built-in search engine optimisation based on cloud database records, the embedded location-based service by Google Maps and local weather information, instant payment options for commerce and shopping apps (e.g., Google Wallet), biometric data services with voice activations and fingerprint technologies, and medical data tracking and health profile apps.[92]

These technologies and, notably, their combinations have created the conditions that can make it instantly and ubiquitously possible to exploit individual vulnerabilities, even if they are only temporary, such as through an ice cream ad displayed when one's sugar level has dropped or a comedy movie when one feels sad. In this respect, it is also important to consider the possible links that exist between subliminal AI techniques (lit. a) and practices exploiting vulnerabilities (lit. b). For instance, it has been argued that 'the increased predictive power of online AI systems, when contained with extensive data about the online subject (user), their behaviour, and their preferences, create an environment in which our natural psychometric differences can become material vulnerabilities'.[93]

In this case, there are numerous instances of discriminatory AI systems that can pose and have already been found to pose serious threats to individuals, groups and society as a whole based on the exploitation of vulnerabilities. One serious problem is found in AI systems themselves becoming – contrary to the common belief in their neutrality – sources of discriminatory treatment.[94] The discrimination or bias by digital technologies has been identified as a serious problem because of a growing number of decisions being 'delegated to systems increasingly based on artificial intelligence (AI) techniques such as machine learning'.[95]

The problem of discrimination is closely tied to the debate on technology neutrality, which regards technology as 'morally and politically neutral'.[96] This approach has its limitations for several reasons. First, the perception of technologies as 'neutral tools' means that they 'can be used well or poorly, for good, evil, or something in between', which often impedes a deeper enquiry into the consequences or 'whether a given device might have been designed and built in such a way that it produces a set of consequences logically and temporally prior to any of its professed uses'.[97] A second problem is reflected in the statement of 'something in between', namely that even the consequences of different usages can change and do not necessarily follow a dualistic mode of thinking in that the choice of one necessarily excludes the other (*expressio unius est exclusio alterius*).[98] The same fallacy also applies to the standard juxtaposition of regulation and technology, which is a more complex relationship that also should not be regarded as a unilateral or unilinear one given that even regulation can be regarded as a kind of technology.[99]

For good reasons, algorithms as a fundamental component of software have been described with the statement, 'not neutral, impartial expressions of knowledge, their work is not impassive and apolitical'.[100] This is why it was recommended to not only think critically about their nature but to also consider their work, effects and power.

Additionally, a potential for AI decision-making processes 'to fail to reflect the lived experiences of individuals' has been identified, which poses a threat to the protection of human diversity.[101] Paradoxically, this means that these apparently intelligent systems can and do 'artificially' create vulnerable groups or individuals. For this reason, it is submitted that children or people with disabilities should not be considered as different or isolated cases per se, as this could result in undue discrimination. Instead, these cases should serve as benchmarks for how AI systems can be safe, trustworthy and fair for everyone.

For instance, dangers related to exploitation have not just been found for vulnerable groups because everyone is potentially vulnerable to manipulation and exploitation depending on a variety of factors other than characteristics related to personality, such as the quantity, intensity, frequency and quality of exposure to a wide range of stimuli.[102] Thus, exploitation affects not only vulnerable groups but also consumers in the commercial realm and voters in the political realm. The possibility of using subliminal or 'subthreshold effects', both in vision and sound, for the purpose of political indoctrination without a person being conscious of any influence was already mentioned by Packard in 1957.[103] In 1960, the Simulmatics Corporation, a US data firm, used a wide range of what-if simulations and algorithms to target voters and consumers and even claimed credit for the successful election of President John F. Kennedy.[104] Their practices were thus very similar to those revealed by the Cambridge Analytica scandal in 2018, which targeted voters based on a combination of Facebook user data and psychographic profiling algorithms.[105] This scandal illustrated how 'psychographic' profiling, meaning the collection of data online to create personality profiles of voters, can be used to influence the outcomes of elections by targeting 'voters with individually tailored content, which is adjusted in real time to reflect the debate that develops around critical electoral issues'.[106] The same effects on both consumer and voter behaviour have been found to exist in the context of search engines and recommender systems.[107] Moreover, the same AI systems and technologies can be used not just to optimise the manipulation of election results but also to incite hatred, violence, insurrection or even wars and war crimes.[108]

Overall, there exists a considerable overlap between the prohibition of subliminal AI techniques and the prohibition

[92] Park and Skoric (2017), p. 74.
[93] Franklin et al. (2022), p. 2.
[94] Tischbirek (2020).
[95] Ferrer et al. (2021), p. 72.
[96] Miller (2021), p. 53.
[97] Winner (1980), p. 125.
[98] Neuwirth (2023), p. 93.
[99] Wiener (2004), pp. 484, 495.
[100] Kitchin (2017), p. 18.
[101] Krupiy (2020), p. 1.
[102] Dixon and Henley (1986); Randolph-Seng and Maher (2009).
[103] Packard (2007), p. 62.
[104] Lepore (2020), pp. 1–2, 321.
[105] Hu (2020), p. 1.
[106] Risso (2017), pp. 75–76.
[107] Epstein and Robertson (2015), p. E4512.
[108] Lauer (2021), p. 400.

of the exploitation of vulnerable groups. The combination of various subliminal and supraliminal practices of both a coercive and deceptive nature with psychographic profiling, microtargeting, big data and so on was explored and confirmed to constitute a form of psychological operations ('psy-ops').[109] Even in the initial press release for the proposed AIA, the Commissioner outlined the dangers of subliminal manipulation by giving the example of 'the case of a toy that uses voice assistance to manipulate a child into doing something dangerous'.[110] Sadly, only a few months later, a real case was reported in which Alexa told a child to plug a phone charger about halfway into a wall outlet and then 'touch a penny to the exposed prongs'.[111] Fortunately, the child did not follow the instructions and avoided an electric shock. This is merely one of many dangers, both present and future, given that the Internet of Toys depicts a scenario in which 'toys not only relate one-on-one to children but are wirelessly connected to other toys and/or database data'; this is already upon us.[112] For instance, there are numerous apps and games that have been identified that 'spy on kids' and collect data and personal information.[113] As another example, a '(pseudo-)intelligent doll' was reported to have collected personal data about a child in Germany in 2017, prompting the regulator to recommend the destruction of the doll.[114]

In sum, there are numerous dangers, related not only to the exploitation of vulnerable groups but also to actual situations in which people can become or even be deliberately made to be via supra- and subliminal stimuli vulnerable to the effects of AI systems. For these reasons, there is a need in this context as well as in the context of subliminal AI techniques to extend, mutatis mutandis, the harm requirement.

Social scoring systems

The third category of prohibited AI practices concerns so-called 'social scoring systems', otherwise known as 'social credit systems'. Neither of these concepts is defined in the AIA, but the latter, in particular, is best known from related experiments that were primarily conducted in China but also in India, the world's two most populated countries.[115] In China, the system relies on the comprehensive collection and expansive use of personal data, and the formulated objective is "to assess the trustworthiness of Chinese citizens in keeping their promises and complying with legal rules, moral norms, and professional and ethical standards'.[116] However, it would be false to assume that the problem of social scoring is actually limited to just India and China. Other countries also use such scoring systems, including the United Kingdom, which has established a scoring system in the form of a digital identity and that attributes a trust framework applicable to organisations wanting to provide or consume digital identity products and services.[117] In the US, for example, credit scoring algorithms are used by independent credit bureaus to assess one's credit worthiness based on education, employment status or type of residence.[118]

On a general level, personal characteristics are collected from social networks and are widely used 'to rate people in fields ranging from insurance premiums, to hiring decisions and employment chances, to social security benefits'.[119] For quite some time already, everyone has been living in a scored society as everyone's personal data has been collected to be used for automated decisions at some point. The key problem is that such scoring systems 'mine datasets containing inaccurate and biased information provided by people' and there is nothing unbiased about them because, far from eliminating existing discriminatory practices, they instead risk systematizing them in hidden ways'.[120] Such discriminatory results have been recorded for various characteristics ranging from gender, race and ethnicity to wider characteristics that include income, names, location and lifestyle.[121] In addition, these systems have been found to not only pose a threat to equality but also to undermine democracy as a whole, for instance, by scoring individual voters.[122]

These real dangers are recognised by the AIA in Recital 17, which states that 'AI systems providing social scoring of natural persons for general purpose by public authorities or on their behalf may lead to discriminatory outcomes and the exclusion of certain groups'. To prevent these effects of social scoring systems, the AIA enshrines the prohibition of 'the placing on the market, putting into service or use of AI systems by public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics'. As a further requirement, the use of such scoring systems must lead to the 'detrimental or unfavourable treatment of certain natural persons or whole groups' in different conditions.

The most important feature of this category in the AIA is that the prohibition in its original version is applicable only to public authorities or those acting on their behalf. This restriction appears problematic, as there are also many private entities making use of scoring systems or the data that are exchanged between public and private entities. Moreover, ratings of people based on algorithmic assessments of personal characteristics taken from social networks are regularly applied in a wide range of fields, from insurance premiums to hiring decisions and employment chances to social security benefits.[123] They can also give rise to discrimination based on a variety of factors, which is why there is considerable overlap with the prohibition of practices exploiting vulnerabilities. The issue of rating can also apply to the labour market, particularly in the so-called gig economy or platform economy

[109] Bakir (2020), p. 1.
[110] European Commission (2021).
[111] Doe (2021).
[112] Holloway and Green (2016).
[113] Fowler (2022).
[114] van den Hoven van Genderen (2017), p. 347.
[115] Schroeder (2022).
[116] Chen and Cheung (2017), p. 356.
[117] United Kingdom Government Digital Service (2021).
[118] Banasiewicz (2021), p. 225.
[119] Packin (2021), p. 487.
[120] Keats Citron and Pasquale (2014), pp. 4-5 and 13.
[121] Selbst (2017), pp. 120-123 and Criado and Such (2019), pp. 87-89.
[122] O'Neil (2016), p. 196.
[123] Packin (2021), p. 487.

in which across different platforms 'various technologies of algorithmic tracking, tracing and rating of the labour of the individual crowdworker are in place'.[124]

For now, it has been indicated that the final version of the AIA may see an extension of the prohibition of social scoring systems to private entities as well.[125] As a practical problem, it is also difficult to detect and distinguish the proper use of scoring systems from their possible abuse. In this regard, the dualistic or dichotomous conception of the law poses another obstacle, one that is manifest in the form of a binary understanding of instruments of reward and punishment as well as the general dichotomy of good versus bad. The reason is that from a dichotomous perspective, manipulation originally appears as a neutral term because most recommendations or prohibitions tend to 'manipulate' or change behaviour in a certain way deemed desirable in legally binding or non-binding ways. The difficulty also surfaces in the debate over nudging, which has been defined as 'any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives'.[126] The same has been found for 'raw data', which – since data that is collected can be used or abused – has been called an oxymoron and a bad idea.[127] Similarly, the social credit system in China is not a recent phenomenon. Many antecedents have been explored and the system's historical roots traced back to personnel archives for officials during imperial times as early as 1045–771 BCE.[128]

The underlying problem with a scoring system is not per se that critical decisions are made on the basis of the data but that decisions are made 'on the basis of data analysed algorithmically: that is, in calculations coded in computer software', the operations of which are mostly kept secret.[129] Social scoring systems are also related to the question of neutrality of more than just technology, including artefacts or human-made objects in general. In the case of social scoring systems, it is mainly about how data and information are being used. However, the ways in which the data are collected, stored, processed and used is what makes the difference in the age of big data. The main difference between big data and data per se has been said to be that it 'contains greater variety arriving in increasing volumes and with ever-higher velocity'.[130] By aiding the automation of data analysis, AI also plays a crucial role in transforming data into both financial value and other forms of value. The present situation has been compared to that of a crossroads at which 'big data, artificial intelligence, cybernetics and behavioural economics are shaping our society—for better or worse'.[131] This quote reflects not only the increasing complexity and interwovenness of the underlying technologies but also their growing speed of change and overall oxymoronic character that defies their classification as good or bad.

Like statistics, data have been found to never be neutral.[132] Social scoring based on algorithmic decision making exemplifies many of the serious risks related to the growing use of automated decision making. For these reasons, social scoring systems are inextricably tied to the collection of big data. The same systems also share many commonalities with other attempts to influence and perhaps even engineer human behaviour, such as nudging.[133] Ratings can also be applied to corporations as well as to other fields, such as environmental performance, social responsibility and corporate governance (ESG) objectives.[134] In this context, too, it is the combination with AI and other technologies that opens a whole range of new applications for big data and AI, and more particularly, for ratings and scoring systems.

Often, a new technology seems to aggravate problems and even create new ones rather than to mitigate or solve old ones. For this reason, the AIA rightly proposes the prohibition of those practices that use data for the evaluation of natural persons based on their social behaviour as well as their known or predicted personality characteristics. As with the first two categories of prohibited AI practices, the potential for the severity of the dangers to increase in the future is real given that the underlying technologies are becoming faster, more sophisticated and more frequently intertwined or combined. While keeping in mind the potential dangers for first-step fallacies whereby limited early success does not constitute a valid basis for predicting the ultimate success of a project or technology,[135] it is still necessary to constantly and closely monitor the development of these technologies. In regulatory terms, however, the closer mutual technological connections and practical synergies that can be derived from the different categories listed as 'prohibited practices' should be addressed in a more coherent way in the planned AIA. It should be considered that a potential connection also exists between subliminal techniques and scoring systems, as the former can alter a person's thoughts and behaviour without that person's awareness, while the latter will then attach certain possibilities or restrictions to that person. Therefore, as an underlying regulatory challenge, the traditional legal mode of dichotomous thinking and framing of problems needs to be rethought in favour of greater coherence and consistency. In addition, the present debate should be broadly concerned with formulating a methodology and laying down criteria indicating which future methods are permissible to achieve what ends.

'Real-Time' remote biometric identification systems

The fourth and last category of prohibited AI practices relates to so-called '"real-time" remote biometric identification systems'. Biometry generally refers to different methods of identifying people using physiological features. Probably the best-known system that uses biometric data is facial recognition technology (FRT), which has already raised concerns under the GDPR, the EU Charter of Fundamental Rights and the Law En-

[124] Altenried (2020), p. 149.
[125] Kazim et al. (2022), p. 4.
[126] Thaler and Sunstein (2008), p. 6.
[127] Gitelman and Jackson (2013), pp. 2–3.
[128] Jiang (2020), p. 93.
[129] Pasquale (2015), pp. 21–22.
[130] Durovic and Lech (2020), p. 156.
[131] Helbing (2018), p. 7.
[132] Huff (1954); Gitelman and Jackson (2013), pp. 2–3.
[133] Österle (2021), p. 241.
[134] Zheng, Khurram, Chen (2022).
[135] Dreyfus (2012), p. 87.

forcement Directive.[136] However, there exist many more biometric systems, including systems that use other features, such as fingerprints, irises, retinas, hand geometry, voices, signatures, palm-print features and ears.[137] As illustrated by the case of FRT, there are doubts about their accuracy and security.[138] There are various forms of the digital manipulation of faces, such as deepfakes, that can be used to mislead identification systems and which, consequently, 'may lead to false decisions and thus decrease the reliability of the decision system'.[139] These dangers have been experimentally confirmed, for instance by adversarial attacks aimed to mislead 'well-trained face recognition models by applying makeup effect to face images'.[140] In another experiment, frames of eyeglasses with adversarial perturbations were used to deceive a facial recognition system and make it recognise the faces as persons different from those wearing them.[141] In short and paradoxically, the greater sophistication of surveillance technology and biometric identification systems also brings new risks based on ways to circumvent them.

Based on these problems, unimodal biometric systems, that is, those using only one biometric feature, are now widely deemed insufficient in terms of accuracy and security. With no single biometric feature appearing as entirely safe and accurate, it is now common to combine features into 'multimodal biometric systems' or 'multi-biometrics systems', such as voice and face recognition.[142] This is similar to the now common use of two-factor authentication for accessing bank or email accounts that often combines a password with a code generated by a security token.[143] Multi-biometric systems often combine several modes, like faces, fingerprints, palm prints, ear specimens and eye patterns (e.g., iris and retina) or combine modes with behavioural biometric traits, such as gait, signature, typing speed/patterns or voice patterns.[144] Thus, behavioural biometrics means to identify persons based on their behaviour or 'on the way they provide information to the authentication system'.[145] Such behavioural biometric systems can include various methods for the verification of a user, such as voice recognition, signature recognition, keystroke dynamics, graphical authentication systems, mouse dynamics, gait (or the way a person walks), smile or lip movements and even odour as well as biological signals, such as those measured by the electrocardiogram (recording of the beating of the heart) and the electroencephalogram (recording of the activity of the brain).[146] The use of odour as a biometric source, for instance, also confirms the increasing need for a joint study of the senses, whereas the use of the electrocardiogram and the electroencephalogram indicates difficulties in distinguishing behavioural from physical and physiological features. The trend appears to point in the direction of combining the different tools of behavioural biometrics with physical and physiological features into multimodal biometrics to optimise the verification or identification of a person.

In the context of the AIA, it is noteworthy that the term 'biometric data' is already defined in a way that includes not only the physical and physiological but also the behavioural characteristics of a natural person.[147] Behavioural biometrics have also been used to add additional security layers for devices and websites and to protect against identity theft, for instance, through the use of keystroke, touchscreen and computer mouse dynamics.[148] Unfortunately, the same adversarial attacks as were applied to facial recognition technology can be carried out in the case of mouse computer dynamics, resulting in the failure to identify a user.[149] The same attacks can also be applied to gait.[150] Practically, it appears that every mode and combined mode of biometric recognition systems that can be used can also be attacked.

In sum, it appears to be a pattern that with every new technology or combination that adds an extra layer of protection, new ways of circumventing it also become possible. This is not the end of the story, as the same technology used to 'poison' or attack a system (e.g., adversarial attacks) can then be adapted for use as a protection against the threat (e.g., adversarial machine learning) and so on and so forth.[151] This is a problem akin to the 'poison paradox' formulated centuries ago by Paracelsus for the field of medicine, but which can also be used for chemicals and likely also technologies.[152] In accordance with the paradox, biometrics, albeit generally useful to verify or identify a person, can also lead to abuse, discrimination and more severe dangers. The reason is that an important shift has taken place in biometrics. This shift has transformed biometrics from a technology purely for identifying people by looking at 'who you are' into a technology that can answer the question of 'how you are'.[153] Again, it appears to be a double-edged sword, as this not only opens a range of opportunities for the use of biometrics, for example, in predicting criminal behaviour, but also raises numerous risks. The risks that have been identified include discrimination, stigmatisation and unwanted confrontation (i.e., in the form of a violation of the right of privacy to not be informed about something undesirable like a potential disease) as well as de-individualisation and stereotyping.[154]

Biometric data also allow for profiling, which has been defined as 'the process of "discovering" correlations between data in databases that can be used to identify and represent a subject and/or the application of profiles (sets of correlated data) to individuate and represent a subject or to identify a subject as a member of a group or category'.[155] The initial im-

136 Raposo (2022b), p. 2.
137 de Luis-García et al. (2003), p. 2539.
138 Raposo (2022c), p. 13.
139 Ibsen et al. (2022), p. 27.
140 Zhu, Lu, Chiang (2019), p. 2516.
141 Ren et al. (2020), p. 350.
142 de Luis-García et al. (2003), p. 2552.
143 De Cristofaro et al. (2014), p. 1.
144 Ambekar, Kakarwal, Khedgikar (2020), p. 9.
145 Revett (2008), p. 1.
146 Revett (2008), pp. 2–12.
147 Art. 3 (33) AIA, *supra* note 9.
148 Moskovitch et al. (2009).
149 Tan et al. (2019), p. 1.
150 Guo et al. (2020), p. 221.
151 Biggio et al. (2015).
152 Neuwirth (2023), p. 85; Timbrell (2005), pp. 2, 250.
153 Schumacher (2012), p. 215.
154 Schumacher (2012), p. 222; Schermer (2013), p. 137.
155 Schreurs et al. (2008), p. 241.

petus for research on behavioural profiling was said to have been linked not to authentication but to fraud detection.[156] Biometrics is thus also a process that confirms that raw data that is available in abundance is indeed an oxymoron because even anonymised data can be used to identify a person. This is possible because of the evolution of re-identification systems that connect different data that can then be traced back to a single subject.[157] For example, it has been shown that it is possible to identify 1% of the Swedish population based only on publicly available data on age, occupation, municipality and gender.[158]

Thus, biometrics is a rapidly evolving field which – with the assistance of AI – allows for a wider use of data today and tomorrow. As is common with any technology, there are already many related risks, such as cybersecurity risks, because the 'same biometric methods that an organization employs for greater security can be turned against them by intruders'.[159] Biometrics also allows for 'emotion recognition', which refers to the use of data to identify the emotional state of a person.[160] Such systems can be used, for instance, in law to understand the intentions of a suspect, in monitoring for risk prevention and for smart health care.[161] On the other hand, they also open the doors to new forms of manipulative practices operating both sub- and supraliminally.

As another example, multimodal forms of biometrics and their combination with other technologies may challenge a traditional understanding and definition of identity.[162] Many technologies in use have already added new aspects of identity, for instance, in the form of a digital identity, which has been said to raise legal, philosophical and other issues.[163] These challenges may not only raise questions about the uniqueness of a person but also extend to questions about the dichotomy of offline and online identity and, related to that, to possible multiple identities. One of the reasons is that the digital realm opens new ways for representing oneself. This was reflected in the Dolly Parton Challenge, which illustrated how people exhibited different presentations of self in different social media, such as Tinder, Facebook, Instagram or LinkedIn.[164] On the other hand, these possibilities also create new risks, such as greater vulnerability, stigmatisation or commodification of the self.[165] In a complex digital environment, multiple digital user identities also increase the need for their secure management, the convergence of which is feared to create additional security and privacy issues.[166] The new possibilities also support the need to complement the regulatory process with parallel research and, notably, interdisciplinary research into the underlying issues of self, identity and related dichotomies.

Overall, the present situation has been compared to an interconnected society in which 'aspects of someone's personal and social life, professional affiliations, hobbies, and interests become part of a public profile'.[167] Yet, recent advancements in machine learning and deep learning have been said to pose new threats to user privacy because of the new opportunities they create 'to extract new knowledge from the publicly available data'.[168] Paradoxically, while research is being carried out to enhance the methods for biometrics verification and the identification of persons, technologies are also being developed that allow for the anonymisation and de-identification of biometric data to protect user privacy.[169] The same paradox applies to the rise of the importance of the Internet of Things, illustrated by the connections amongst numerous devices, such as smartphones, wearables, robots, autonomous vehicles or drones. Here it is the massive interconnection of these different devices that, on the one hand, makes the accuracy and security of user authentication critically important,[170] while on the other, hand, it also increases the possibilities for the collection of biometric data to identify persons and intrude in their privacy in the first place.

The AIA addresses these concerns and counters the possible risks by qualifying certain biometric systems, namely the 'the use of "real-time" remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement' as prohibited AI practices.[171] In contrast to the terms used in the other categories of prohibited practices, the term '"real-time" remote biometric identification system' is defined by the AIA as 'a remote biometric identification system whereby the capturing of biometric data, the comparison and the identification all occur without a significant delay'.[172] The same definition clarifies that 'real time' not only includes 'instant identification, but also limited short delays in order to avoid circumvention'.[173]

Thus, the keywords for the prohibition of these biometric identification systems are that they are operated 'remotely', in 'real time' (including short delays) and in 'publicly accessible places' as well as 'for the purpose of law enforcement'. A specific example of a system falling within this definition would be the use of a large-scale CCTV network coupled with facial recognition software.[174] However, the relevant prohibition in the AIA is narrowed down to the context of law enforcement, which is also regulated in other EU acts, namely the EU Law Enforcement Directive.[175] At the same time, the

156 Clarke (2011), p. 130.
157 Hildebrand (2006), p. 552; Bolognini and Bistolfi (2017), p. 176.
158 Moreton and Jaramillo (2021), p. 277.
159 Barton et al. (2005), p. 26.
160 Yannopoulos, Andronikou, Varvarigou (2008), p. 91.
161 Mehta, Haque Siddiqui, Haque Siddiqui (2019), p. 3.
162 Rannenberg, Royer, Deuker (2009), p. 1.
163 Allison et al. (2005), p. 325.
164 Shulman (2022), p. 30.
165 Shulman (2022), p. 36.
166 Agbinya, Islam (2020), pp. 259–260.
167 Shopon (2021), p. 470.
168 Shopon (2021), p. 470.
169 Jeong et al. (2020).
170 Liang (2020), p. 9128.
171 Art. 5(1) lit. d) AIA, *supra* note 9.
172 Art. 3(37) AIA, *supra* note 9.
173 *Ibid*.
174 Veale and Zuiderveen Borgesius (2021), p. 101.
175 Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ L 119/89.

provision also features three broad exceptions to the prohibition, namely when the said system is used for:

(i) the targeted search for specific potential victims of crime, including missing children;
(ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack;
(iii) the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA62 and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State.[176]

These three exceptions to the proposed prohibition are rather wide and have been criticised for that reason. Arguments have also been made for a stricter approach, 'given that remote biometric identification, where AI may contribute to unprecedented developments, presents extremely high risks of deep and non-democratic intrusion into individuals' private lives'.[177] The narrow scope and limitation to the context of law enforcement has also been criticised because this 'appears to permit the use of such systems for other purposes'.[178] Other commentators have highlighted the intrusive nature of some relevant systems and have called for stronger protection against AI-enabled manipulation, social scoring and the use of biometric identification systems in general, while singling out the need to 'prohibit the use of emotion recognition systems and the use of live remote biometric categorisation systems by law enforcement and by other public actors with coercive powers, or private actors acting on their behalf'.[179]

In this last category of prohibited AI practices listed in Art. 5 AIA, several serious concerns and concrete risks related to remote biometric recognition were outlined:

> The increasing use of AI biometric surveillance systems raises serious concerns as to fundamental rights. Remote biometric recognition is linked to deep interference with the right to privacy, including people's autonomy, their right to establish details of their identity and psychological integrity. It negatively impacts freedom of expression, association and freedom of movement. Remote biometric identification and predictive tools may lead to discrimination, violate the values of equality and justice due to biased data sets and errors as well as undermine the rights to liberty and to a fair trial.
>
> Moreover, biometric recognition systems, including those used for social scoring and predictive policing, may enable mass surveillance. The clearest distinction between AI systems in authoritarian countries and AI systems in democratic countries is the use of facial recognition for mass surveillance.[180]

These serious ethical and legal concerns are thus not exclusive to remote biometric identification systems but are also known in the context of the previous categories of prohibited AI practices. In addition, the dilemmas and paradoxes caused by dualistic or dichotomous thinking encountered in the case of other technologies are applicable here. This leads to another fallacy, which – despite its serious implications – has humorously been called the 'Humpty Dumpty fallacy' in the context of the EU Commission's choice to dichotomise high-risk and non-high-risk AI systems.[181] The fallacy illustrates the limitations inherent in dichotomies and has been restated as: 'Just because the Commission exhaustively enumerates high-risk AI systems does not mean the residual category displays non-high-risk'.[182]

To conclude the detailed overview of the single categories qualified as prohibited AI practices by Art. 5 AIA, it is important to stress that there are inherent deficiencies in the present legislative regulatory framework, both materially and institutionally, as well as the underlying modes of thinking, reasoning and logic. Together, they call for a new conceptual approach that also considers the growing convergence of AI systems and related technologies.

The convergence of AI systems and other technologies

Based on the dichotomy of high-risk versus non-high-risk AI systems, the proposed AIA lists the four different categories of prohibited AI practices. Each of these covers different aspects, usages and purposes of the same or at least related underlying technologies. They not only overlap in the technologies but also in their application to different media. They have in common other profound elements that manifest in the general trend of convergence, a faster pace of innovation and change and, notably, several unsolved regulatory conundrums that are often framed as oxymora and paradoxes. For instance, the technological convergence is often met by a regulatory divergence or general fragmentation of law.[183] Paradoxes also clash with the tendency towards dualistic thinking and the framing of complex problems by means of dichotomies. This was particularly exemplified for the current time of digital transformation in which paradoxes display contradictions between interdependent elements that often suggest the use of a both/and approach instead of an either/or one.[184] For this reason, it has been noted in the field of regulation that strict prohibitions or requirements as regulatory instruments for technology and innovation have their limitations.[185]

Eventually, these questions all relate to the unsolved conundrum of whether all technology is neutral and dual

[176] Art. 5(1) lit. d) AIA, *supra* note 9.
[177] European Data Protection Supervisor (2021).
[178] Gill-Pedro (2021), p. x.
[179] Smuha et al. (2021), p. ii.
[180] Barkane (2022), p. 150 [citations omitted].
[181] De Cooman (2022), p. 50.
[182] De Cooman (2022), p. 50.
[183] Neuwirth (2015), pp. 3, 26.
[184] Danneels, Viaene (2022), p. 484.
[185] Eisenberger (2016), pp. 156-157.

use.[186] The same issues arose in the context of the (at least for now) conceptual difficulties in distinguishing between humans and machines as captured by the notions of cyborgs or technological singularity.[187] Formulated in positive terms, it must also lead to better ways of combining substantive and procedural law as well as the adoption and enforcement of law. This last point finds its expression in the necessity to ex ante ponder ways of monitoring compliance with and enforcement of the proposed AIA. Therefore, the ongoing legislative process must duly include a debate about the competent regulator(s) for the oversight of AI, such as a possible 'European Agency for AI'.[188]

In the end, it needs to be critically asked whether these conundrums have deeper causes that must be investigated at the level of the human mind by notably aiming to understand the strong tendency across times and cultures to apply dualistic thinking.[189] These questions may render it necessary to rethink existing fundamental rights in the context of AI and novel scientific findings or even to create new digital rights in the future.[190] It is likely that the answers to these questions will be found in the way humans think, and this emerges in our language use, because AI itself, cyborgs, raw data and machine learning have all been defined as oxymora or contradictions in terms.[191] These linguistic trends highlight the need to search for a new legal logic and to rethink the basic premises of the traditional forms of dualistic legal reasoning and binary logic.[192] Already in 1924, John Dewey had suggested a social and intellectual need for the infiltration into law 'of a more experimental and flexible logic'.[193] Due to new and mostly disruptive technologies like AI, this intellectual need may have transformed into a practical necessity. In practice, steps regarding the implementation and enforcement of AI laws must go beyond dichotomies of unacceptable or no risks, legal prohibitions or requirements and punishment or reward as well as public or private entities. Instead, a search must be made for novel and more holistic ways of technological governance. This must also apply to the four listed categories of prohibited AI practices, which provide sufficient evidence that their mutual overlap, entwinement and complementarity reinforce their individual effects. In this regard, an important lesson can be learned by way of analogy with common ownership theory about a party holding minority equity shares in multiple competing firms,[194] according to which even small stakes of control in different AI systems can, when combined, be sufficient to create dangerous levels of power in the control of people's minds, behaviours and lives. Such a scenario must be avoided in all aspects of life to make sure that the conditions guaranteeing the privacy of thoughts and autonomy of actions are maintained in the future.

Conclusion

> *The Old Law restrains the hand, but the New Law controls the mind.*[195]

The European Union's proposal for an Artificial Intelligence Act, released in April 2021, marks a first important step in the direction of the regulation of AI in a comprehensive way and on the basis of a legally binding instrument. From the perspective of the complex, convergent and often contradictory characteristics inherent in the distinct technologies underlying AI, the horizontal and risk-based approach chosen by the EU is to be applauded. The same applies to the planned prohibition of certain AI practices enshrined in Article 5 AIA, which broadly covers the four categories of subliminal AI systems, AI practices exploiting vulnerabilities, social scoring systems and 'real-time' biometric identification systems.

In practice, every person is already often but unknowingly and simultaneously confronted with each of these categories through a complex combination of techniques, applications and instruments. For instance, in one single day, a person is likely exposed to personalised advertisements using subliminal techniques addressing one or all their senses for both commercial and political purposes as well as subject to the known manipulative effects of search engines and recommender systems. The same person is profiled based on a collection of all kinds of data, including biometric data, derived from Internet activities, wearable health and other Internet of Things devices, which can then be used for scoring when applying for a job, a loan or an educational or health institution, to mention but a few.

The overall complexity of the various sources that inform and shape a person's thoughts and behaviour have been expressed by the notion of the 'glass human being' (or 'vitreous human'). This term was originally coined in the German language to summarise the dangers related to an excessive and encroaching collection of personal data as well as the storage and coupling of personal data from public and private bodies.[196] But it is no longer just a literary metaphor for a dystopian scenario; it is also part of today's reality. Various technologies and their applications, such as AI, data mining, machine learning, data integration and fundamental algorithms as well as optimisation techniques and web mining methodology, are combined with the goal 'to know it all'.[197] Knowledge is often equated with power and knowing it all means omniscience leading to omnipotence or all-encompassing power. Such omnipotent omniscience was described by Simon Laplace using a hypothetical Demon as a being able to predict future events.[198]

However, the power to predict the future also manifests itself in another way, namely in the power to manipulate and control.[199] It has already been shown that synthetic media, broadly defined as the artificial manipulation, modification

[186] Miller (2021), p. 53 and Pustovit and Williams (2010), p. 18.
[187] Siivonen (1996), Gozman, Liebenau, Ferris (2019), Eden (2012), and Nicolescu (2016), p. 43.
[188] Stahl et al. (2022).
[189] Webb (2020), p. 702.
[190] Bublitz (2014) and Custers (2022).
[191] Neuwirth (2020).
[192] Jeutner (2017), Glenn and Smith (2017) and Neuwirth (2018).
[193] Dewey (1924), p. 26.
[194] Schwalbe (2018), p. 596.
[195] Aquinas (1915), p. 18.
[196] Graf v. Westphalen (1983).
[197] Chakraborti (2009).
[198] Laplace (1995), p. 2.
[199] Neuwirth (2022b), p. 43.

and production of information through a wide spectrum of communication media from audio-video deepfakes to text-based chatbots, allows for the creation of digital puppets that can be manipulated by a human puppet master's movements or speech.[200] Thus, the next dystopian scenario could well be one of a puppet human being whose 'private thought' has become an oxymoron, a scenario in which human beings are remotely controlled and even programmed.[201] Yet, this power of control vested in AI has been said to also lead to a paradox of control, according to which 'the very technologies and practices we develop in order to enhance our control end up undermining it'.[202]

For these reasons, the protection of the cognitive freedoms of present and future generations from intrusions is one of the major regulatory challenges related to AI. Each of the prohibited AI practices already separately poses serious dangers to everyone, and their growing mutual entwinement may aggravate the dangers in the future. In attempts to meet these dangers, mere prohibitions and other existing legal tools may not suffice, on the one hand, to contain the multiple risks and dangers caused by AI, while, on the other hand, allowing good use to be made of AI's benefits and opportunities. Instead, efforts should focus on a complete revamp of the legal system by better coordinating the introduction of incremental changes introduced in special legal fields. As indicated by St. Aquinas' prediction of the new law controlling the mind, future laws and legal systems should also be based on a renewed understanding of the workings of the human mind in line with findings formulated by neuroscience and related fields.[203] In these multiple tasks, creative efforts using paradoxes and oxymora can help show the way to reconceptualise the meaning of law, to future-proof it and to redefine law in its relationship to other fields and to work towards a coherent, inclusive strategy for the collective design of the future.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

[200] Canham, Tuthill (2022), p. 50.
[201] Neuwirth (2023), p. 121.
[202] Di Nucci (2022), p. ix.
[203] Bublitz (2014).