Article

Practices manipulatives of artificial intelligence and consumer protection

Ana Felicitas Muñoz Pérez

Full professor of Commercial Law, Private Law Department, at the Universidad Rey Juan Carlos (URJC). Master's degree in Corporate Legal Consultancy from IE. Her main lines of research are in the field of company law and the securities market, and she currently devotes particular attention to digital innovation in the commercial field, coordinating the Research group in digitalization and business law of the Rey Juan Carlos University 'DYDEM-URJC' (https://gestion2.urjc.es/pdi/grupos-investigacion/dydem). His latest monograph is on 'DAOS and the rest of controlling the algorithm'. E-mail: anafelicitas.munoz@urjc.es

Received 21 January 2024, Accepted 02 Febrary 2024

KEYWORDS:

Digitalisation, Artificial Intelligence (AI), Consumer Law, Data privacy

ABSTRACT:

Artificial intelligence (AI), together with other technologies, is enabling the targeting of consumption through massive data analysis and the development of the ability to predict reactions through so-called 'influence machines'. These tools pose new risks, as they may exploit psychological biases, information gaps or personal needs and manipulate decisions, raising ethical and legal questions. The research focuses in particular on 'subliminal, manipulative or deceptive' methods prohibited under the AI Regulation and so-called 'dark pattern' designs that employ similar practices. The study explores a holistic approach to the legal provisions covering manipulative and deceptive techniques in the AI Regulation and other EU rules, although only a case-by-case approach could conclude an accurate assessment of the impact on consumers' privacy and freedom of choice.

PALABRAS CLAVES:

Digitalización, Inteligencia Artificial (IA), Derecho del consumidor, Protección de datos.

RESUMEN:

La inteligencia artificial (IA), junto con otras tecnologías, permite orientar el consumo mediante el análisis masivo de datos y el desarrollo de la capacidad de predecir las reacciones a través de las denominadas «máquinas de influencia». Estos instrumentos plantean nuevos riesgos, ya que pueden llegar a explotar sesgos psicológicos, lagunas de información o necesidades personales y manipular las decisiones, lo que plantea cuestiones éticas y legales. La investigación se centra, en particular, en los métodos «subliminales, manipuladores o engañosos» prohibidos en el Reglamento de IA y en los denominados diseños de «patrones oscuros» que emplean prácticas similares. El estudio explora un planteamiento holístico de las disposiciones legales que contemplan las técnicas manipuladoras y engañosas en el reglamento de inteligencia artificial y otras normas comunitarias, aunque solo un enfoque caso por caso podría concluir una evaluación certera de la incidencia en la privacidad y la libertad de elección de los consumidores.

MOTS CLES :

Numérisation, intelligence artificielle (IA), droit de la consommation, confidentialité des données

RESUME :

L'intelligence artificielle (IA), associée à d'autres technologies, permet de cibler la consommation grâce à l'analyse massive de données et au développement de la capacité à prédire les réactions par le biais de ce que l'on appelle les “machines à influencer”. Ces outils présentent de nouveaux risques, car ils peuvent exploiter des biais psychologiques, des lacunes en matière d'information ou des besoins personnels et manipuler des décisions, ce qui soulève des questions éthiques et juridiques. La recherche se concentre en particulier sur les méthodes « subliminales, manipulatrices ou trompeuses » interdites par le règlement sur l'IA et sur les modèles dits « sombres » qui recourent à des pratiques similaires. L'étude explore une approche holistique des dispositions juridiques couvrant les techniques de manipulation et de tromperie dans le règlement sur l'IA et d'autres règles de l'UE, bien que seule une approche au cas par cas puisse conclure à une évaluation précise de l'impact sur la vie privée et la liberté de choix des consommateurs.

CONTENTS:

1 INTRODUCTION

Artificial intelligence has provided technologies capable of influencing consumers by being able to predict their reactions. Machine learning technology analyses the correlation between consumer data - e.g., purchases, sites visited, social media "likes" - anticipating the responses that consumers may offer to advertisements from entrepreneurs. The so called "influence machines" based in systems of AI rise new threaten to consumer, some of them has be here considered.

The ability of this systems to anticipate consumer response can be used to manipulate decisions, as responses could be based on irrational aspects of their psychology, lack of information or a situation of need. It must be ensured that the development and deployment of AI tools takes place in a socio-technical framework – inclusive of technologies, human skills, organizational structures, and norms – that avoid such risk and where individual interests and the social good are preserved and enhanced.

This situation of risk has been warned by the doctrine (SARTOR a)2020)[1] and the law is incorporating new definitions to catch this behavior and avoid harm consumer. The AI act has defined as unacceptable risks those behavior that are deemed to contravene Union values, and they are therefore considered as "prohibited AI practices" by Article 5 AIA. These definitions must be combined in the traditional concepts included in other statutory provisions in order to harmonize the legal treatment. The restrictions come out in other normative framework, diverse sectorial regulations, such as unfair law, consumer law and data protection regulations.

We focus on this paper a general approach to the context of the recent AI regulation related to AI systems deploying "subliminal, manipulative, or deceptive" techniques and the partial equivalent concept is known as "dark patters", which encompasses certain tactics that the AI Regulation does not express included, but mostly are partial coincident with de definition of "Subliminal, manipulative, or deceptive techniques", all of them forbidden for being deceptive or manipulative. Nevertheless, the application of the rules will depend on a case-by-case assessment, considering the specific nature of the technique used and its impact on users.

2 THE SHIFT FROM THE CONTEXT-BASED TO TARGETED ADVERTISING AND THE RISK OF "INFLUENCE MACHINES".

The last decade has seen the evolution of the 'information economy', with the emergence of monopolies - 'digital empires' - that control data and services and where the massive collection of personal data is used to develop tactics to influence consumer behavior. As a result, a new data-driven economy is emerging that challenges traditional moral and ethical approaches in a new world where information is power. Any research on data exploitation helps to explore possible ways forward, advocating for a fair and balanced digital future (KISHOR MAHAPATRA, R. (2024)).

The so-called Web 2.0. reflected the emergence of a new model of disintermediated production and distribution (SARTOR, G, 2020). Parties had unlimited access to information, in a global marketplace. In the consumer sphere, the Internet was supposed to reinforce the market power of the consumer in relation to traders, enhancing the selection opportunities of the former and thus leaving the behaviour of the latter to the discipline of the market. Today, the emergence of the new socio-technical infrastructure has unbalanced this trend.

[1] Giovanni SARTOR, "New aspects and challenges in consumer protection, IPOL | Policy Department for Economic, Scientific and Quality of Life Policies, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/648790/IPOL_STU(2020)648790_EN.pdf, provided a document published by the Policy Department for Economic, Scientific and Quality of Life Policies at the request of the committee on the Internal Market and Consumer Protection (IMCO).

A set of monopoly intermediaries dominates access to information on the Internet - search engine services, content sharing platforms, e-commerce, cloud services, online payments... -. Information from provider to consumer now also flows from consumer to provider. Service providers have created a business model that is based on selling to 'advertiser' clients the data they have previously collected from user clients in exchange for the offer of their services (CATELLS (2001).

Mass data capture leads to the economic model called 'surveillance capitalism' which is characterized by the commodification of human experience. Under this model the analyzed and recorded data of users is used to build opportunities to anticipate and influence their behaviour. As the Cambridge Analytica case clearly exemplified, data collected from social networks, merged with data from other sources, can be used to understand people's character, interests and political opinions, and consequently target them with messages aimed at changing their behaviours (CHAN (2019)[2]. The so-called "surveillance society" and the economic model named "surveillance capitalism" comes out the risks to consumer privacy and autonomy (CRACKED LAB (2017).

I this realm arise the "influence machines", mechanisms of "influence on citizens" and "on consumers" (MAGNANI, E. (2020). In the "surveillance capitalism" online advertising shifts from innocuous automatic and personalized messages to individual consumers based on information about them to the activity of using this information for the purpose of targeting consumers. Context-based advertising -where users see ads that match the page they are visiting or the search they have just performed-, is shifting to targeted advertising -that aims to influence consumption, creating the risk of manipulation-.

Under this title a new "personal data economy" is emerging, where all kinds of personal data are collected and exchanged, with their value being their potential use to anticipate and modify the behaviour of individuals, consumers in particular Giovanni SARTOR, 2020. AI causes a new case of information asymmetric between marketers and consumers[3]. There are the multiple players involved in the consumer tracking industry in which online platforms play a key role, but which also includes retail providers, media and publishing industries, telecommunications and telecommunications services. All of them coordinated create models characterized by the commercialization of "human experience" which it converts the behaviour of the users into recorded and analysed behaviour, which brings the enterprises transforming it into marketable opportunities to anticipate and influence the future consume and causing risk of manipulation and other unfair practices.

It is said that IA allows the move from "programming computers to programming people" (HELBING et. Al (2019)4. The convergence of AI with various related technologies, such as brain–computer interfaces, functional magnetic resonance imaging, robotics and big data, already allows for "mind reading" or "dream hacking" through brain spyware, as well as other

---

[2] As it was explained CHAN, R., "The Cambridge Analytica whistleblower explains how the firm used Facebook data to sway elections" https://www.bbc.com/mundo/noticias-43472797- Cambridge Analytica targeted users that were "more prone to impulsive anger or conspiratorial thinking than average citizens" by creating Facebook groups, sharing articles, and advertising. According to news published by The New York Times and The Observer, private data was then used to psychologically manipulate voters in the 2016 US elections, where Donald Trump was elected president

[3] The former know their products and services better than the consumer does, and can find out much more about consumers than consumers know about them, MARWALA, T., HURWITZ, E., (2015) "Artificial Intelligence and Asymmetric Information Theory"

[4] HELBING D ET AL. (2019) "Will Democracy Survive Big Data and Artificial Intelligence?: Essays on the Dark and Light Sides of the Digital Revolution, Towards Digital Enlightenment" (pp.73-98), relates that the field of artificial intelligence is, indeed, making breathtaking advances and it will change fundamentally the way in which we organize the economy and society. "Singapore is seen as a perfect example of a data-controlled society. What started as a program to protect its citizens from terrorism has ended up influencing economic and immigration policy, the property market and school curricula. China is taking a similar route. Recently, Baidu, the Chinese equivalent of Google, invited the military to take part in the China Brain Project. It involves running so-called deep learning algorithms over the search engine data collected about its users. Beyond this, a kind of social control is also planned. According to recent reports, every Chinese citizen will receive a so-called "Citizen Score", which will determine under what conditions they may get loans, jobs, or travel visa to other countries. This kind of individual monitoring would include people's Internet surfing and the behavior of their social contacts (see "Spotlight on China")".

practices that intrude on cognition and the right to freedom of thought. The literature warned that future innovations will enhance the possibilities for manipulating thoughts and behaviour, and they threaten to cause serious harm to individuals as well as to society (NEUWIRTH, (2022).

AI systems have developed the capacity to use instruments of manipulation, which is caused by using the irrational aspects of consumers' psychology, leading them to unconsciously replace "rational consumption" with "irrational choices". Consumers are driven to buy goods they do not need, to overspend, to engage in risky financial transactions or to give in to their weaknesses -e.g., gambling or drug addiction-. The "radical behaviourist approach" relies on cognitive and behavioural patterns that often operate on automatic, almost instinctive levels, and can be manipulated instrumentally, giving way to the named "dark pattent".

Additionally, even, AI systems can learn and cause other unfair practices, i.e. "price discrimination" finding out the reserve price of individual consumers, that means the highest price they can accept. The "reserve price" of particular consumer, -i.e. the highest price they may accept, also name consumer's willingness to pay (WTP)-, empower the traders due they may be able to push onto the consumer the highest offer that the latter may accept, eroding consumer surplus (BAR-GILL, O. (2019)5.

In this situation industry claims that regulators should seriously explore under the concept of "surveillance capitalism" and "personal data economy" the new "market law personalized" that reveal certain tactics of the sellers that imposes changing behaviours or drives to the imposition of personal treatment breaching the uniform behaviour in the market - for instance, targeted marketing, personalized pricing…-. This may not only affect consumer welfare, but also the efficiency of markets, which is in fact based on the assumption of general commercial conditions and the absence of price discrimination.

It is beyond the scope of this study to determine all the general features upon the treatment for consumer derived from the so-called "surveillance capitalism". The research only focus a general approach to the case of consumer deception facilitated by new technologies.

3 THE FORBIDDEN MANIPULATIVE PRACTICES IN THE IA ACT

3.1 FORBIDDEN PRACTICES

According the definitions incorporated in the recent enacted Artificial Intelligence Act[6], an "artificial intelligence system" (AI system) is defined as: A machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how

[5] The author states that "Sellers are increasingly utilizing big data and sophisticated algorithms to price discriminate among customers. Indeed, we are approaching a world in which each consumer will be charged a personalized price for a personalized product or service", and wonder if this type of price discrimination good or bad. "Sellers are interested in the consumer's willingness to pay (WTP) for their goods or services: they maximize profits by charging a price that is as close as possible to the consumer's WTP. This WTP is a function of consumer preferences on the one hand and consumer (mis)perceptions on the other hand. When algorithmic price discrimination targets preferences, it harms consumers but increases efficiency. Legal intervention may be needed when price discrimination targets misperceptions, specifically demand-inflating misperceptions, it hurts consumers even more and might also reduce efficiency.

[6] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevanceRegulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance.

to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

The IA act, prohibit create an IA system to develop prohibited practices and pointed out which one must be considered forbidden. Recitals provide context about how the articles of the IA Act should be interpreted or implemented. The 29th recital allude to the manipulative or exploitative artificial intelligence enabled practices defined as prohibited to protect consumers. Accordingly, "AI-enabled manipulative techniques can be used to persuade persons to engage in unwanted behaviors, or to deceive them by nudging them into decisions in a way that subverts and impairs their autonomy, decision-making and free choices"[7].

The art. 5 concrete that must be considered as prohibit practices, the one that drives to deploying "subliminal, manipulative, or deceptive techniques" targeted to distort behavior and impair informed decision-making, causing significant harm, and exploiting vulnerabilities related to age, disability, or socio-economic circumstances with the same effect of distort behavior, causing significant harm.

Nevertheless, *common and legitimate commercial practices*, for example in the field of advertising, that comply with the applicable law should not, in themselves, be regarded as constituting harmful manipulative AI-enabled practices encompassed in this frame.

3.2 Legal concern on the frame of Prohibitions

The question that arises in the AI regulatory framework is whether the best way to combine the protection of consumer interests and the support of technology and innovation (NEUWIRTH, (2022).

The AI regulation forbidden the tactic of abuse that consist in the practices on the control on the people's minds, behaviors and lives, that raises the impact in the violation of fundamental rights of privacy and freedom.

The increasing convergence between existing and future technologies, applications and tools increase the risk of manipulation.

Also, if those technologies operating with these different categories of prohibited practices against consumers ends up in the hands of one or a few dominant entities, whether public or private (SCHAWALBE, 2018)[8].

On central issue raised around this regulation is the best perspective to address ethical treatment by questioning whether it should be approached from the perspective of human behavior or starting from the machine design approach. And on this question, the constant blurring of the lines of distinction between humans and machines captured by the notions of cyborgs ("cybernetic organisms") or "technological singularity" should be underlined (CORTINA, 2024).

In the background, it is being questioned whether the technology is technology neutral or dual use (NEUWIRTH, (2022).

[7] The 29th IA Regulation

[8] It is obvious, that the total profit in a market is maximized when firms do not compete with each other but form a cartel and behave collectively like a monopolist, if there is widespread common ownership, the overall intensity of competition would be reduced, prices were higher and consumer welfare would decrease SCHWALBE, U., "Common Ownership and Competition" – The Current State of the Debate, Journal of European Competition Law & Practice, Volume 9, Issue 9, November 2018, Pages 596–603, https://doi.org/10.1093/jeclap/lpy050 warns because this trend the concern caused by horizontal shareholdings of institutional investors.

3.3 THE LEGAL DEFINITION OF "SUBLIMINAL" "MANIPULATIVE", OR "DECEPTIVE" TECHNIQUES.

As I have previously said, the IA act leaves the scope of term undefined but the art. 5 prohibits AI systems from using "subliminal techniques" that influence decisions or actions in ways likely to "cause significant harm".

Generally, "subliminal" AI practices refer to strategies with perception below the threshold of consciousness. "Subliminal" is defined to "not recognized or understood by the conscious mind, but still having an influence on it" (CAMBRIDGE DICTIONARY). The application of the standard definition of "subliminal techniques" would render the AI Act's prohibition irrelevant to most forms of problematic manipulative influence in practice. Related to AI refers to those practices that have a significant potential to manipulate people through techniques beyond their consciousness and to ensure that the AI Act is legally workable and to reduce regulatory uncertainty, a different, the regulation act states a explicit definition targeted to addresses the underlying societal concerns about manipulation without overburdening service providers (NEUWIRTH, R., 2022).

Subliminal AI practices, raises several questions, such as 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 associated with existing and future AI techniques, and the enforcement of the provision (NEUWIRTH, R., 2022).

The research of "subliminal AI techniques" belongs to the field of the general psychoanalytical techniques employed by the advertising industry. The term "subliminal" has faced critique, as manipulative techniques often engage both subliminal and supraliminal levels. For instance, subtle methods like briefly flashed images or masked sounds might be used in conjunction with more overt manipulations, such as fake news or deep fakes—both identified as powerful tools for influencing behavior. Some propose using "transluminal" instead of "subliminal" to reflect that perception lacks absolute thresholds, as individual sensitivity varies.

This technique affects the autonomous choices in ways that people are not consciously aware of or cannot control or resist. 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 (STONE, 2023)[9]. The members of the Council of Europe and the European Union, as well as other jurisdictions, decided in 1989 to prohibit all forms of subliminal advertising[10].

Scientifically, the old controversy about the effectiveness of subliminal stimuli in terms of their effects on the human mind and behavior requires currently further clarification. The IA regulation raises the question of what constitutes an AI system that deploys "subliminal techniques" that must be considered manipulative. Particularly regarding the multisensory human organs involved in the new AI subliminal techniques the recital 29Th mention as examples of "subliminal AI techniques" that enables the IA to both, "machine brain interface", commonly named dark pattern, and the "virtual reality". The rapid advancement of AI-driven multisensory augmented and virtual reality further emphasizes the need to study sensory integration.

[9] In 1957, James Vicary claimed to have exposed moviegoers to brief messages about popcorn and Coca-Cola, stating that it increased sales. He later admitted that the experiment was fraudulent and that such subliminal advertising was ineffective. (STONE, (2023) James Vicary's Popcorn Experiment in 1957 | Impact & Issues, https://study.com/academy/lesson/james-vicary-experiment-lesson-quiz.).

[10] 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).

Certainly, IA allows more sophisticated techniques like “mind reading” and “brain spyware” -software intentionally designed to detect private information-, “dream hacking” or “dream incubation” -stimuli before or during sleep to affect dream content- which combines visual stimuli and machine learning model to detect familiar information base on brain response. This different mind readings technologies combined can become effective tools for commercial manipulation and be used to get access to personal data as well as private thoughts “without consent of the individual affected or even the awareness that such information is being taken (UNESCO 2020).

The most serious offence related to subliminal manipulation lies in the possibility of “invading the privacy of our minds”, -i.e. audio, image, video stimuli that persons cannot perceived, messages include subliminal 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-.

Overly strict definitions should be rejected as inefficient, as a narrow definition would exclude many of the cases that justify the prohibition and risks missing many cases of manipulation. The IA Act should include a broader definition with practical considerations. Currently, the doctrine seeks to include in the IA Act a broad definition that aims to capture most problematic cases of AI-driven influence, while remaining focused enough for practical enforcement.

The 'mechanisms of influence' forbidden are those with lack of clarity and unaware for consumers and influence on the decisions or beliefs of consumers so deep that are capable to change their decisions and beliefs. The following enumeration clarifies the features of the content:

- The premise of the AI Act is that AI-driven influence can be problematic due to lack of awareness. Subliminal techniques are defined as methods that aim to influence behavior in a way that the individual is likely to remain unaware of the 'attempt to influence'. Consumers need to be informed about when they are being targeted, how these techniques exploit psychological biases, and what the impact may be.
- Many users are unaware that websites customize content based on personal information to drive engagement or sales, and that such customization can nudge consumers towards certain actions without our awareness
- The undisclosed and not transparent techniques impacts and influence on the decisions or beliefs of consumers and changes their decisions and beliefs. The technic in marketing named nudge easily turns into a base dark patterns shove (as CDS Marketing, Dark Patterns – When a nudge becomes a shove, https://blog.cds.co.uk/dark-patterns-when-a-nudge-becomes-a-shove. Companies uses nudges and dark patterns to avoid or disclose these methods).
- If the prohibition on subliminal AI techniques also requires proving intentional manipulation by the developer, remains challenging. The IA systems so-called “dark patters” are defined as “deceptive elements that are intentionally crafted to make the users do actions that they wouldn´t do otherwise. They are identified as unethical marketing cause usually are embedding subliminal stimuli to influence and encourage users to purchase more products (BEUC (2022)[11]. But the recital 29th

[11] BEUC in “Dark patterns” and the EU consumer law acquis Recommendations for better enforcement and reform” states that “Consumers increasingly face “dark patterns” because on-line user interfaces are designed to deceive consumers to take decisions that are not in their favor but serve companies’ commercial interest. The use of unfair practices to distort consumers’ economic behavior is not new, but it takes a new important dimension as a result of the massive collection of data and the use of technology to build consumer profiles and anticipate consumer behavior. EU consumer law already has partial capacity to address these situations, but it is currently not sufficiently enforced. In addition, EU law must be updated to tackle these unfair practices and ensure consumers are not harmed by misleading user interfaces and data personalization techniques”.

clarifies that are prohibits the AI systems with the "objective" to or de "effect" of materially distorting human behavior.

In this light, the provision against subliminal AI techniques is an essential step toward managing AI risks and must be prioritized in the EU's AI Act (AIA) and in broader initiatives, such as the UNESCO Recommendation.

3.4 EFFECTS, "CAPABLE TO DISTORTION OF THE BEHAVIOR OF CONSUMER IN HARM MANNER".

They are prohibiting AI systems with the "objective" to or de "effect" of materially distorting human behavior whereby significant harms. I.e. subliminal AI practices which "materially distort" a person's behavior in a way that causes or is likely to cause physical or psychological harm to that person or to another person.

It means that the AI Act does not outright prohibit subliminal techniques; it restricts only those likely to cause "significant harm" the technologies that materially distort their behavior in a significantly harmful manner. Consequently, the main regulatory burden falls on assessing whether a system raises significant harm risks—a due diligence step already common in user protection- and determine the definition enough to capture problematic cases of manipulation but without overburdening regulators or companies (BERMÚDEZ,J.P., RUNE NYRUP, R., SEBASTIAN DETERDING,S., RAFAEL A. CALVO, R., (2024).

The current AI Act limits harm scope to "physical or psychological harm," which seems to be too narrow. Critics recommend broadening it to include financial, economic, cultural, and autonomy-related harms, as well as collective or societal harms. Furthermore, it is essential to recognize "cumulative harms"—those that intensify with repeated exposure, affecting environments through hyper-personalization and engagement metrics, especially impacting children.

It is particularly dangerous the placing on the market, putting into service or use of certain AI systems with the aim or effect of significantly distorting human behavior, where significant harm is likely to occur, having a sufficiently serious adverse effect on physical or mental health or on financial interests.

3.5 THE EXCEPTION OF THE COMMON AND LEGITIMATE COMMERCIAL PRACTICES.

In any case, it is not necessary for the provider or the deployer to have the intention to cause significant harm, provided that such harm results from the manipulative or exploitative AI-enabled practices. They are both forbidden with the objective to or the effect or materially distorting human behavior. This statement makes questionable the scope of the forbidden practice where the distortion results from factors external to the AI system which are outside the control of the provider or the deployer, namely factors that may not be reasonably foreseeable and therefore not possible for the provider or the deployer of the AI system to mitigate

The AI Act prohibition will not affect lawful practices in the context of medical treatment such as psychological treatment of a mental disease or physical rehabilitation, when those practices are carried out in accordance with the applicable law and medical standards, for example explicit consent of the individuals or their legal representatives.

4 THE OVERLAPPING CONCEPT OF "DARK PATTERNS" IN THE CONSUMER EU LAW

4.1 General approach

The manipulative practices upon users leading them into making unintended actions, often for the benefit of the company, are named "Dark patterns" and they are commonly created thought deceptive user interface designs. The tactics can range from subtle to outright manipulative, and they're becoming increasingly common in digital products and services (LUGURI, STRAHILEVITZ (2021))[12]. Users exposed to mild dark patterns are twice as likely to sign up for a dubious service, while aggressive patterns led to nearly four times the sign-up rate, though they also caused consumer backlash. Less educated users are more vulnerable to mild dark patterns. Dark patterns -like hidden information, trick questions, and obstructions- most effectively misled consumers, while other strategies -like urgency messaging- are less impactful (LUGURI, STRAHILEVITZ (2021)). Notably, dark patterns led users to ignore service costs, making decision architecture rather than price the main driver. Therefore, legal reforms, as many dark patterns may violate trade laws and consumer consent under them could be voidable.

Traditional protection related to these deceptive tactics comes under the horizontal EU consumer law acquis. The behavior is regulated mainly under the following general framework: Unfair Commercial Practices Directive[13], the Consumer Rights Directive[14] and, the Unfair Contract Terms Directive[15].

4.2 The application of Unfair Law (UCPD)

Unfair Commercial Practices Directive is capable of addressing many of dark pattern's practices (BEUC (2022)[16]. The prohibitions for AI practices are complementary to the provisions contained in Directive 2005/29/EC of the European Parliament and of the Council[17], unfair commercial practices leading to economic or financial harms to consumers are prohibited under all circumstances, irrespective of whether they are put in place through AI systems or otherwise. It is possible to apply the "duty of professional diligence" to counter "dark patterns" according to the art. 5 UCPD. Additionally, dark patterns can be

---

[12] LUGURI, STRAHILEVITZ (2021) "Shining a Light on Dark Patterns" Journal of Legal Analysis, Volume 13, Issue 1, Pages 43–109, https://doi.org/10.1093/jla/laaa006, explains that dark patterns are user interfaces whose designers knowingly confuse users, make it difficult for users to express their actual preferences, or manipulate users into taking certain actions. They typically prompt users to rely on System 1 decision-making rather than more deliberate System 2 processes, exploiting cognitive biases like framing effects, the sunk cost fallacy, and anchoring. The goal of most dark patterns is to manipulate the consumer into doing something that is inconsistent with her preferences, in contrast to marketing efforts that are designed to alter those preferences.

[13] Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair businessto-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive').

[14] Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council (Consumer Rights Directive)

[15] Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts

[16] BEUC (2022 "dark patterns" and the EU consumer law acquis", includes "Recommendations for the European Commission regulatory intervention", pág 2, recommend the Review the Unfair Commercial Practices Directive (UCPD) to include: • Introduce new benchmarks for the assessment of what constitutes (un)fairness in commercial communications to update the "average" consumer concept, including how digital asymmetry impacts consumers' transactional decisions. • Introduce a new rule alleviating the burden of proof for plaintiffs and enforcement authorities. • The establishment of an obligation for business to respect the general principle of "fairness by design". • Certain dark patterns should be banned in the annex of the UCPD such as the practice of "confirm-shaming" which implies using language and emotion (e.g., shaming) to steer users into or away from - or make them feel guilty about - making a specific choice or from refraining to take action.

[17] Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive')

contemplated as "misleading actions" or "misleading omissions" under Articles 6 and 7 UCPD. Dark patterns can be considered forms of "aggression" under Articles 8 and 9 of the UCPD. Lastly, Anex I of the UCPD lists prohibitions relevant to dark patterns, such as practice 6 on "bait and switch", where consumers are drawn in by a product but then redirected to a different one via deceptive interfaces, and practice 7, which pressures consumers into quick decisions (e.g., false scarcity claims on booking platforms). Additionally, the Annex could be expanded to include tactics like "confirm shaming", which uses guilt or shame to manipulate users' decisions. The IA reinforce the chance to catch the emotional analysis of users and increase this risk of the development of this tactic.

4.3 THE CONSUMER REGULATION (CDR)

The Consumer Regulation Directive (CDR) indicates that in distance contracts, the information needs to be provided in a "way appropriate to the means of distance communication used in plain and intelligible language". This transparency requirement has been extrapolated from the Unfair Contract Terms Directive and requires consumers to be able to understand the information that is being provided and the consequences of entering a contract. Through the design of the interface, the trader can breach this transparency requirement by hiding some material information and deviating the consumer's attention to other elements. The prohibition of Article 22 of the CRD on pre-ticked boxes is also relevant since the very essence of the prohibition was to prevent traders from taking advantage of consumers' status quo biases. While complying with the prohibition of pre-ticking boxes implying additional payments, companies can still push consumers to un-tick boxes via a misleading user interface, for example, by placing the button to go to the next page (e.g., in a booking site) right next to the ticked box or in a very similar shape and color or by making it look like ticking the box is required in order to proceed. Furthermore, except for the right of withdrawal, the CRD does not regulate the conditions for contract termination. In this regard, to prevent consumers from facing difficulties to terminate contracts it could be amended to include an obligation to have a contract cancellation button.

4.4 UNFAIR CONTRACT TERMS (UCT)

The UCTD in the Section 3 under art. 5 discusses the assessment of the transparency and (un)fairness of contract terms as significant tool in addressing dark patterns, particularly through its transparency requirement, which mandates that terms be presented in "plain intelligible language"- i.e."the clarity of the visual presentation, including font size, the fact of whether a contract is structured in a logical way and whether important stipulations are given the prominence they deserve and are not hidden amongst other provisions". The Commission's guidance highlights the need to assess factors like visual clarity, logical structure, and prominence of important terms. This interpretation makes the UCTD useful for examining whether dark patterns obscure terms unfairly, impacting transparency. Transparency goes beyond the contractual terms as it relates to the way consumers are informed, but also how it allows them to make a conscious and informed transactional decision and therefore makes the UCTD a relevant instrument to scrutinize whether through dark patterns suppliers are not rendering terms and conditions unfair, through lacking transparency (BEUC, 2023, pág. 10). According, transparency requires more than contract terms being formally and grammatically intelligible and implies that consumers must be able to evaluate the economic consequences of a contract term or contract[18].

[18] For example, (COMMISSION NOTICE Guidance on the interpretation and application of Council Directive 93/13/EEC of 5 April 1993 C(2019) 5325 final, pág 20 and 26 ss) "Opinion of Advocate General Hogan of 15 May 2019 in Case C-621/17 Kiss, paragraph 41" stated "on unfair contract terms in consumer contracts" that contract terms" whose impact can only be understood when reading them jointly, should not be presented in such a way that their joint impact is obscured, e.g. through placing them in different parts of the contract". The Court has ruled that that requirement, also repeated in Article 5 thereof, cannot be reduced merely to their being

Many dark patterns exploit consumer decision-making by manipulating user interfaces. Through techniques like A/B testing, traders can predict consumers' likely choices and use this knowledge to influence them[19]. For example, making contract cancellation difficult discourages consumers from unsubscribing, as seen in the Norwegian Consumer Council's Complaint against Amazon for complicating Amazon Prime cancellations. Similarly, UK research found that insurance companies impose "loyalty penalties" by making it harder for long-term consumers to cancel, leading to higher premiums. Privacy settings also frequently use "click fatigue," requiring so many steps for privacy-friendly options that consumers give up, as noted in Facebook and Google's privacy designs.

4.5 Consumer protection in the Data act (GDPR)

The activity of Artificial intelligence-based consumer data processing named "profiling" is relevant due must respect the main objectives of consumer protection law: "protection of the weaker party", the "autonomy of consumer" and the "non-discrimination practices". First, the "protection of the weaker party" could be challenge using artificial intelligence by suppliers/traders and service providers, due it may introduce additional imbalances and asymmetries of information between the supply side and the demand side. Second, the manipulative use of big data and artificial intelligence may limit "consumer independence", so the use of big data and artificial intelligence may impact in the ability to make properly informed decisions based on their reasoned preferences which is challenged by the possibility of influencing consumer choices, possibly without consumers being aware of such influence. Consumers may be "hyper-guided" by targeted advertising and adaptive manipulative design toward choices they are likely to regret. Third, automated decisions may disadvantage certain individuals and groups, without any acceptable justification creating "discrimination practices". The discrimination, as it enables new and more refined and systematic forms of stereotyping and differentiation. AI systems may present challenges for anti-discrimination law, as individuals may be unfairly subjected to differential treatment according to criteria that do not directly coincide with the discriminations prohibited by EU law.

Additionally, AI and big data systems are required to address potential flaws to ensure that consumers' privacy is protected. Controllers engaging in AI-based processing should endorse the values of the GDPR and adopt a responsible and risk-oriented approach. This can be done in ways that are compatible with the available technology and economic profitability -or the sustainable achievement of public interests, in the case of processing by public authorities-. However, given the complexity of the matter and the gaps, vagueness and ambiguities present in the GDPR, controllers should not be left alone in this exercise. It is said that Data protection authorities need to actively engage in a dialogue with all stakeholders, including controllers, processors, and civil society, to develop appropriate responses, based on shared values and effective technologies. Consistent application of data

formally and grammatically intelligible, but that, to the contrary, since the system of protection introduced by Directive 93/13 is based on the idea that the consumer is in a position of weakness vis-à-vis the seller or supplier, in particular as regards his level of knowledge, that requirement of plain and intelligible drafting of contractual terms and, therefore, the requirement of transparency laid down by the directive must be understood in a broad sense. According to the Court, transparency requires more than contract terms being formally and grammatically intelligible and implies that consumers must be able to evaluate the economic consequences of a contract term or contract

[19] Multivariate (A/B/n) experiments are Quickly test multiple feature variations with multivariate experiments. The test removes silos and complexity from the experimentation process, allowing the trader team to easily set up and run valid experiments and make improvements based on outcomes. The technic "Context-aware targeting" allows gain total control of your releases. "Targeting Segments" make it easier to group similar targets into a cohort and create feature flag rules for it. The trader can even import large user segments from product analytics software providers—and sync changes to those segments between providers.

protection principles, when combined with the ability to efficiently use AI technology, can contribute to the success of AI applications, by generating trust and preventing risks[20].

4.6 THE IMPACT ON CONSUMER OF THE DIGITAL SERVICES ACT – DSA- AND THE DIGITAL MARKET ACT

Dark patterns were a key topic in the legislative discussions around the Digital Markets Act (DMA)[21] and the Digital Services Act (DSA)[22].

In their respective positions, both the European Parliament and the Council chose to address dark patterns specifically within the anti-circumvention provisions related to gatekeepers' obligations outlined in Articles 5 and 6 of the DMA. The restriction applies in relation to compliance with the DMA's obligations. In this context, the Unfair Commercial Practices Directive (UCPD) remains fully applicable to any dark patterns employed by traders classified as 'gatekeepers' under the DMA.

The Regulation "Digital Services Act" ensures that all digital services provided by big digital platforms are engines subjects to measures to protect the rights of their users, ensure safety and curb the dissemination of illegal or inappropriate content. The Regulation requires online platforms to consider the impact of their services on issues such as fair elections, public safety, the mental and physical well-being of users or gender-based violence (BEUC 2022)[23].

The Digital Services Act (DSA) also addresses dark patterns within intermediation services. The Council mandates that online marketplaces and very large online platforms, particularly in the use of recommender systems, must avoid designs, structures, or functionalities that undermine user autonomy, decision-making, or choice (as outlined in Council amendments to Articles 24b and 29 of the Commission's text). The European Parliament goes further, proposing a general ban on manipulative interface designs for all intermediary service providers (Article 13a). The DSA maintains the core "safe harbor" liability protections from the E-Commerce Directive (ECD) for information society service providers but updates it by incorporating certain criteria established by the Court of Justice. Notably, the DSA's framework diverges from the previous concept that intermediaries who exert minimal control bear less liability, instead clarifying an active duty of care in moderation practices (PÉREZ BENITEZ, (2024). The Regulation also provides for solutions specific to the Digital Markets Directive, such as the establishment of notification mechanisms that allow any person to report the presence of illegal content, and imposes specific risk assessment and detection obligations on the major platforms.

5 REFERENCES

BAR-GILL, O. (2019) "Algorithmic Price Discrimination When Demand Is a Function of Both Preferences and (Mis)perceptions," University of Chicago Law Review: Vol. 86: Iss. 2, Article 12.
Available at: https://chicagounbound.uchicago.edu/uclrev/vol86/iss2/12

[20] BEUC (2022) "Dark patterns" and the EU consumer law acquis Recommendations for better enforcement and reform.

[21] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)

[22] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)

[23] The Digital Services Act – DSA (https://www.aepd.es/documento/el-reglamento-de-servicios-digitales-dsa-explicado.pdf) require certain restriction so that anyone can analyze their potential risks of on line platforms. The platform must include, among other things, information on the content of the ad and who paid for it, especially when it is directed at minors

BERMÚDEZ,J.P., RUNE NYRUP, R., SEBASTIAN DETERDING,S., RAFAEL A. CALVO, R., "The AI Act needs a practical definition of 'subliminal techniques" (2024) https://www.euractiv.com/section/artificial-intelligence/opinion/the-ai-act-needs-a-practical-definition-of-subliminal-techniques/

BEUC, (2022) "DARK PATTERNS" AND THE EU CONSUMER LAW ACQUIS. Recommendations for better enforcement and reform

CASTELLS (2001), The internet galaxy.

COMMISSION NOTICE Guidance on the interpretation and application of Council Directive 93/13/EEC of 5 April 1993 C(2019) 5325 final

CDS Marketing, Dark Patterns – When a nudge becomes a shove, https://blog.cds.co.uk/dark-patterns-when-a-nudge-becomes-a-shove

CHAN, R., "The Cambridge Analytica whistleblower explains how the firm used Facebook data to sway elections" https://www.bbc.com/mundo/noticias-43472797

CORTINA, 2024 ¿Etica o ideologia de la inteligencia artificial?

CRACKED LAB. Corporate Surveillance in Everyday Life. Report and web publication on how companies collect, analyze, and use personal data. June 2017

EDPS, EUROPEAN DATA PROTECTION SUPERVISOR (2018) Summary of the Opinion of the European Data Protection Supervisor on the legislative package "A New Deal for Consumers".

HELBING D ET AL. (2019) "Will Democracy Survive Big Data and Artificial Intelligence?: Essays on the Dark and Light Sides of the Digital Revolution, Towards Digital Enlightenment" (pp.73-98)

KISHOR MAHAPATRA, R. (2024), Digital Empires: The Rise of Colonies Through Data Collected by the Internet.

KLEIN, (1966) A Comprehensive Etymological Dictionary of the English Language. Elsevier, Amsterdam.

LUGURI, STRAHILEVITZ (2021) "Shining a Light on Dark Patterns" Journal of Legal Analysis, Volume 13, Issue 1, Pages 43–109, https://doi.org/10.1093/jla/laaa006

MAGNANI, E. (2020). Reseña de The age of surveillance capitalism (Hachette Book Group, 2019) de Shoshana Zuboff. Revista Hipertextos, 8 (14), pp. 165-171. DOI: https://doi.org/10.24215/23143924e024

MARWALA, T., HURWITZ, E., (2015) "Artificial Intelligence and Asymmetric Information Theory"

NELSON, (2016) The hidden persuaders Then and Now, University of Illinois, Urbana-Champaign.

NEUWIRTH, R., (a) (2022)) "The EU Artificial Intelligence Act: Regulating Subliminal AI Systems" SSRN Electronic Journal, DOI: 10.2139/ssrn.4135848

NEUWIRTH, R., (b)(2022)) -"Prohibited artificial intelligence practices in the proposed EU artificial intelligence act (AIA)", March 2022 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4261569

SARTOR, G ((a)2020) "New aspects and challenges in consumer protection, IPOL | Policy Department for Economic, Scientific and Quality of Life Policies, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/648790/IPOL_STU(2020)648790_EN.pdf;

SARTOR, G ((b)2020) "The impact of the General Data Protection Regulation (GDPR) on artificial intelligence", STUDY Panel for the Future of Science and Technology;

STONE, James Vicary's Popcorn Experiment in 1957 | Impact & Issues, https://study.com/academy/lesson/james-vicary-experiment-lesson-quiz.