Computer Law & Security Review 55 (2024) 106031

Contents lists available at ScienceDirect

Computer Law & Security Review: The International Journal of Technology Law and Practice

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

Who is vulnerable to deceptive design patterns? A transdisciplinary perspective on the multi-dimensional nature of digital vulnerability[1]

Arianna Rossi [a], Rachele Carli [b,d], Marietjie W. Botes [c,e], Angelica Fernandez [d], Anastasia Sergeeva [d], Lorena Sánchez Chamorro [d,f,*]

[a] LIDER Lab, Dirpolis, Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà 33, Pisa, 25156, Italy
[b] AlmaAI, University of Bologna, Via Galliera 4, Bologna, 40121, Italy
[c] Center for Research Evaluation Science and Technology (CREST), Stellenbosch University, Cnr Merriman and Ryneveld Streets, Stellenbosch, South Africa
[d] University of Luxembourg, 2, Avenue de Université, Esch sur Alzette, 4365, Luxembourg
[e] School of Law, Howard College, University of KwaZulu Natal, Mazisi Kunene Road, Glenwood, Durban, South Africa
[f] Human-Media Interaction Group, University of Twente, Drienerlolaan 57522 NB Enschede, The Netherlands

ARTICLE INFO

Keywords:
Digital vulnerability
Deceptive design patterns
Dark patterns
Risk assessment
EU digital strategy

ABSTRACT

In the last few years, there have been growing concerns about the far-reaching influence that digital architectures may exert on individuals and societies. A specific type of digital manipulation is often engineered into the interfaces of digital services through the use of so-called dark patterns, that cause manifold harms against which nobody seems to be immune. However, many areas of law rely on a traditional class-based view according to which certain groups are inherently more vulnerable than others, such as children. Although the undue influence exerted by dark patterns on online decisions can befall anybody, empirical studies show that there are actually certain factors that aggravate the vulnerability of some people by making them more likely to incur in certain manipulation risks engineered in digital services and less resilient to the related harms. But digital vulnerability does not overlap with traditionally protected groups and depends on multifaceted factors. This article contributes to the ongoing discussions on these topics by offering (i) a multidisciplinary mapping of the micro, meso, and macro factors of vulnerability to dark patterns; (ii) a subsequent critical reflection on the feasibility of the risk assessment proposed in three selected EU legal frameworks: the General Data Protection Regulation, the Digital Services Act, and the Artificial Intelligence Act; (iii) and multidisciplinary suggestions to increase resilience towards manipulative designs online.

1. Introduction

Dark patterns or deceptive design patterns[2] are design strategies in user interfaces that influence users' decisions concerning the use of their financial resources, of their personal data, and of their time through manipulation, coercion or deception [1,2]. Dark patterns are a distortion of the UX and UI design strategies that are intended to make the interactions between humans and technologies easier, smoother and more pleasurable. Even when they are not employed with the intentional purpose of manipulation [3–5], dark patterns have harmful effects on users, such as loss of autonomy, privacy invasion, financial losses, cognitive burdens, and discrimination, among the others [1,2]. Even though anybody can fall prey to dark patterns because they exploit cognitive biases [6,7], there are factors that make certain individuals, groups, or communities more vulnerable than others [8]. In fact, vulnerabilities are transversal in human-machine interactions and are

* Corresponding author.
E-mail addresses: lorena.sanchezchamorro@utwente.nl, lorena.sanchezchamorro@uni.lu (L. Sánchez Chamorro).

[1] This article is the result of a fruitful transdisciplinary collaboration and all authors have contributed to the drafting of the first version of the paper. The main authors of each section are: Section 1: A. Rossi and L. Sánchez Chamorro (equally); Section 2: A. Rossi; Section 3: A. Rossi; Section 4: A. Rossi and L. Sánchez Chamorro (equally); Section 5: A. Rossi, R. Carli, M. Botes and A. Fernandez (equally); Section 6: L. Sánchez Chamorro, apart from Sections 6.3.1, 6.3.2, and 6.3.3 whose main author was A. Sergeeva and Section 6.4 whose main author was A. Rossi; Section 7.1: R. Carli; Section 7.2: A. Rossi; Section 7.3: L. Sánchez Chamorro; Section 7.4: A. Rossi and L. Sánchez Chamorro (equally); Section 8: A. Rossi.

[2] Nowadays, "deceptive design patterns" is a preferred term to "dark patterns". In this article, we will employ both interchangeably, alongside manipulative designs and similar wordings.

https://doi.org/10.1016/j.clsr.2024.106031

Available online 4 October 2024

compounded by a set of external and internal fac tors.

Following a general trend in the consumer protection domain that aims at providing a more realistic definition of vulnerable consumers in digital settings [8–10], in the last few years rising concerns have been voiced about the situations where certain people might be more vulnerable than others [8]. There is a pressing need to identify and denounce the specific, complex, and intermingled conditions of vulnerability towards manipulative design in online services, going beyond the traditional class-based legal views that assign a vulnerability status to specific groups, such as children (see e.g., the European Data Protection Board's guidelines on deceptive designs [11]) and that ignores the factors that may render everyone dispositionally vulnerable [10], also due to the way digital architectures are designed [8].

In this article, we intend to identify the sources of vulnerability to dark patterns so that appropriate protections and countermeasures can be devised and effectively applied. We will argue, in line with existing accounts (e.g., [12]), that vulnerability towards dark patterns encompasses two dimensions: (i) the susceptibility to being influenced by the design elements of user interfaces (UIs), and (ii) the severity of the harms and consequences that people may suffer. Such a conceptualization closely follows the notion of risk, which is defined by both the likelihood that a certain event occurs and the severity of the impact it can provoke [13]. In our approach, we seek to show that it is necessary to include a multifaceted notion of vulnerability in risk appraisal methods to enhance their accuracy and efficacy, in data protection as well as in other domains of the law. For instance, both the EU Artificial Intelligence Act (AI Act) and the Digital Services Act (DSA) emphasize the necessity of a proactive technology risk assessment that seeks to develop safe-by-design products and services, while enhancing the accountability of those who conceive, create, and ultimately produce them. Hence, with our analysis of the concrete sources of vulnerability, we intend to make an innovative contribution to the definition of the elements that a sound risk assessment methodology needs to entail.

In order to do so, we will first (Section 2) develop the critiques to the particularistic approach (that identifies certain groups as inherently weaker due to their inner characteristics, such as age) and universalistic approach (that generalizes the risk of being vulnerable to anyone and thereby carries the danger of levelling out relevant differences that would deserve special protection) of vulnerability [14] to dark patterns and explain why both are partial views that cannot appropriately account for this complex phenomenon. We will then (Section 3) illustrate how the consumer protection domain has pioneered the revisiting of the concept of vulnerability in digital settings. We will propose in Section 4 an overview of the harms that dark patterns can pose, while, in Section 5, we will explore how data protection law (i.e., the General Data Protection Regulation (GDPR))[3] and the emerging normative framework designed by the AI Act[4] and the Digital Services Act[5] (DSA), address online manipulative design, vulnerability and the assessment of risks posed by technologies. We will propose in Section 6 a multidisciplinary mapping of the micro, meso, and macro factors that can influence the disposition to harm that people can suffer. Through the discussion of practical examples and empirical results, this conceptualization will show that certain factors may be equal for all while others may disproportionally affect certain people. However, as we will conclude in Section 7, this composite reality adds complexity to the implementation of risk assessment procedures in practice. Conscious of this challenge, this work does not aim to provide definitive conclusions, nor ready-made solutions: rather, it aims to provide the foundational elements that are necessary to evaluate the risks of digital technologies in a more accurate, complex manner.

The contributions of this article are the following:

1. a multidisciplinary mapping of the micro, meso, and macro factors of vulnerability to dark patterns;
2. a critical reflection on the feasibility of the assessment of risk factors of vulnerability to dark patterns, as proposed in the three EU regulations chosen for the analysis (i.e., GDPR, DSA, AI Act);
3. multidisciplinary suggestions to increase resilience towards digital manipulative designs.

The authors have expertise in different domains spanning AI, consumer and data protection law, human-computer interaction, and usable privacy. Their backgrounds range from psychology to social sciences and law. There is also an intersection of the authors' research areas: online manipulation, ethics, and regulation of technologies. The ideas developed in this article result from a fruitful transdisciplinary dialogue across the different domains. The authors position themselves as firm defenders of online users and their everyday interactions with technologies and understand the complexity of their problems as issues that can only be solved with the engagement of all the stakeholders that shape and enable contemporary digital markets.

2. Current conceptualizations of vulnerability to manipulative designs

2.1. A universalistic view: vulnerability is inherent to human psychology

In the current academic scholarship on dark patterns, there seems to be convergence on the idea that every user of technology can be vulnerable to online manipulation because of common cognitive fallacies shared by all human beings. According to certain accounts [15], dark patterns affect our so-called System 1, which is the complex of mental processes that corresponds to our quick, intuitive, automatic mode of thinking and that enables us to efficiently navigate the numerous, complex decisions we are faced with in every moment of life (see dual process theory by [16]). Specifically, the user interface (UI) of a technology can be designed in a way that exploits the vulnerabilities arising from cognitive biases and bounded rationality. Cognitive biases are systematic deviations from the decisions and behaviours that a rational decision-maker would enact [1,6,7,17]; while bounded rationality refers to the limited mental resources to which we can resort to consider all possible courses of action and their consequences when faced with a decision [17].

Both are inherent to human nature and prevent people from acting as perfectly rational agents. As a result, people may overlook, and even consciously ignore, overwhelming information; or select an unfavorable option when presented with many choices, since the cognitive energies and time that can be devoted to each task are limited. For example, studies [18,19] have demonstrated that privacy invasive design patterns (such as defaults) on cookie banners can play on the status quo bias and significantly increase consent rates. Despite their potential negative impact, cognitive biases also have a functional reason to exist, since they can, together with rules of thumb (i.e., heuristics), facilitate meaningful choices that allow human beings to "satisfice" (satisfy + suffice) [20]. In other words, they enable users to take decisions in a relatively quick manner, instead of getting paralyzed by the pondered analysis of the

[3] Regulation (EU) 2016/679 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 and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Published: O.J. L 119, 4.5.2016, p. 1–88.

[4] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised 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) PE/24/2024/REV/1 OJ L, 2024/1689, 12.7.2024, p. 1-144.

[5] 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) PE/30/2022/REV/1 OJ L 277, 27.10.2022, p. 1–102.

overwhelming quantity of information, options, and their consequences that they need to navigate in all aspects of digital life.

Identifying cognitive biases as the primary psychological correlate of vulnerability to online manipulation suggests that vulnerability can be minimized or eradicated by overcoming such biases [21,22], for example by revealing the influence of design elements through increased transparency, or by educating users to recognize manipulative attempts. Such interventions aim at nudging users to engage in more reflective thinking, therefore activating our so-called analytical System 2, as opposed to System 1. This assumption underlies regulatory interventions that mandate information disclosures with the goal of lowering the informational asymmetry between organizations and individuals.

The same assumption also motivates the proposal for cognitive boosts [23,24] aimed at strengthening people's competence to make their own choices and for friction designs [25,26] aimed at making certain actions less easy (i.e., automatic) to perform. Although these are useful interventions in specific settings for specific goals, they cannot be reasonably embedded in the sheer number of decisions that individuals need to take every day on digital public and private services: after all, our less reflective cognitive system carries the benefit of ensuring that we act quickly and with relatively low effort in our everyday life [16]. Further, such rationality-enhancing interventions are likely unable to counter hidden deceptive attempts that act below our awareness [27,28], such as when cookies are installed on a device irrespective of the option selected by the user on the interface (see e.g., [29]). Moreover, they cannot target factors of vulnerability such as e.g., socio-economic differences, as we will detail in Section 6. In conclusion, referring to cognitive biases and the UI design that exploits such biases as the sole elements that can explain the effect of dark patterns on human beings provides a limited view of what is a wider problem space, that hence needs varied solutions.

2.2. A particularistic view: vulnerability depends on inherent conditions of certain groups

A different conceptualization of vulnerability to dark patterns is proposed in the 2023's European Data Protection Board (EDPB) guidelines [11] that provide instructions on the design patterns that are not compliant with the GDPR. The EDPB's guidance document recognizes only certain groups as especially vulnerable to online deceptive designs, thereby reflecting a 'particularistic' (or 'class-based' [8]) view of vulnerability. Following the conceptualization contained in the GDPR's Recital 38, the EDPB recalls that dark patterns may have a particularly severe impact on children, since "they may be less aware of the risks and consequences concerned [sic] their rights to the processing" (p. 10). The guidelines also mention "the elderly, persons who are visually impaired, or not as digitally literate as others" (p. 10), who would be less capable of recognizing deceptive designs and less aware of their susceptibility to being influenced.

The EDPB only mentions vulnerable groups again once in para. 44 (p. 20), in relation to the manipulative design practice called "emotional steering" which can allegedly have a bigger impact on those groups that have a "vulnerable nature as data subjects" who may be lured into excessively disclosing their personal data "due to a lack of understanding". Following the GDPR's provisions which mandate that information disclosures should be understandable to children (Article 12(1)) and official guidelines [30], the EDPB's recommendation against such a dark pattern consists in targeted language, tone, and style to raise the chances of understanding and thereby obviate to the vulnerable people's lack of awareness of the risks and consequences of data processing.

We have two main critiques to the EDPB's conceptualization of vulnerability. First, they frame the problem of dark patterns as merely one of informational asymmetry, lack of comprehension and impediment to rational decision-making, which reflects an imperfect understanding of the reasons why dark patterns work. This, in turn, impacts the solutions that are proposed to meaningfully counter their influence, which mainly aim at increasing transparency of data processing operations. Indeed, transparency requirements risk not only to be shortsighted, but also to have the paradoxical effect of increasing the cognitive burden on users, therefore causing individuals to read and understand the provided information even less (an effect called "the transparency paradox" [31]), let alone use it for more conscious decisions. Second, the EDPB's formulation of vulnerability only refers to specific groups of users who are intrinsically weak, but it ignores the multifaceted factors that can contribute to a state of vulnerability outside those groups, as we will detail in Section 6. In this context, we do not maintain that children and other vulnerable groups do not deserve special protection, we rather claim that the particularistic approach is an insufficient approach.

3. Reframing vulnerability in digital markets

3.1. The reasons why existing approaches are insufficient to account for dark patterns

Taking a universalistic position that ascribes the influence of dark patterns only to common cognitive biases and bounded rationality would minimize the specific susceptibility of certain individuals or groups to dark patterns and the disproportionate effects they may suffer. On the other hand, we are cautious of embracing a particularistic view that identifies certain groups as inherently weaker as if vulnerability was part of their immutable essence, but ignores the multidimensional, dispositional state of vulnerability that anyone may suffer from under various circumstances.

Both approaches provide a non-realistic view of the dynamics to which users of digital technologies are subjected and reflect a partial understanding of the complexity of factors that play a role in such dynamics. Furthermore, the digital environment exacerbates our vulnerability and lowers our resilience. In online settings we tend to act faster than in analog ones, skim rather than read, suffer from shorter attention spans, and more easily trust strangers' recommendations [12]. Moreover, we are less able to process information, while we resort to rules of thumb and underestimate manipulation more often than in offline contexts [8]. This is why all individuals can be effectively manipulated in digital interactions: vulnerability is common to all human beings because it is rooted in our bodily, limited, perfectible being [32]. Having said that, it is paramount to acknowledge the conditions under which certain individuals or groups may be more vulnerable than others. Reconciling both views is necessary to reliably assess and mitigate the risks carried by technologies and consequently design counter measures that have the desired impact, including the application of effective policy instruments and the appropriate attribution of responsibilities to the stakeholders of digital markets.

3.2. Reframing consumer vulnerability in digital markets

In the consumer protection domain, there have been many recent efforts aimed at surpassing categorical notions of vulnerability and at providing a more complex understanding of the conditions under which consumers may be vulnerable. This must be interpreted as an attempt to go beyond the static view of vulnerable consumers conveyed by the Unfair Commercial Practices Directive[6] (UCPD) that describes them as

[6] Directive 2005/29/EC of the European Paliament 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') OJ L 149, 11.6.2005, p. 22–39.

"people particularly vulnerable to the practice or the underlying product because of their mental or physical infirmity, age and credulity" (Article 5(3)) as opposed to the "average" or "reasonable" consumer. For example, Baker, Gentry, and Rittenburg [33] reject the limitation of immutable internal traits to propose a multidimensional, situational model of consumer vulnerability that is brought about by the interaction of personal states (e.g. mood), personal characteristics (e.g. socioeconomic status), and external conditions (e.g. discrimination) and aggravated by the lack of control and experience that consumers may have.

In digital environments, however, vulnerability can be exacerbated to a novel, worrisome level. In their comprehensive essay, Helberger and co-authors [10] propose that "vulnerable consumers are not the exception, they are the rule" (p. 180, [10]). In the digital society, the authors convincingly argue, vulnerability is (i) architectural, as the digital choice architecture we engage with on a daily basis are designed to infer the vulnerabilities of individuals or even to produce them (e.g., consider the practice of hypernudging [34] which becomes very concerning when it becomes personalized manipulation); (ii) relational, as individuals' ties to others increase their vulnerability to influencing factors; and is exacerbated by (iii) a general lack of privacy and the concentration of personal data in the hands of few actors. Hence, there are internal vulnerability drivers as well as external drivers in the digital economy, including choice architecture, that make everyone "(dispositionally) vulnerable under the right conditions" (p. 194, [10]).

Such a nuanced discussion echoes the findings of the European Commission's (EC) report published in 2016 [9], where the vulnerable consumer was defined as "a consumer who, as a result of socio- demographic characteristics, behavioral characteristics, personal situation and market environment is at higher risks of experiencing negative outcomes in the market; has limited ability to maximize their well-being; has difficulty in obtaining or assimilating information; is less able to buy, choose or access suitable products; or is more susceptible to certain marketing practices" (p. 5, [9]). The EC study highlights that certain factors have a more severe impact on vulnerable consumers than others, such as market related drivers (like the inability to read contract terms and conditions because of small print), behavioral drivers (like impulsivity and risk aversion), as well as situational drivers (like finding it difficult to making ends meet) [9]. Building on this, in 2021 the Commission returned to the interpretation of the concept of vulnerable consumer, defining this condition as situational and dynamic [35]. Thus, it was intended to allude to vulnerability as a modulable condition, which is not the exclusive prerogative of some, an interpretation that would inevitably leave everyone else immune.

This reconceptualization of consumer vulnerability in a multidimensional key is so well established in the consumer protection domain that it has even been codified in the recent international standard ISO 22458:2022 "Consumer vulnerability — Requirements and guidelines for the design and delivery of inclusive service" [36] where consumer vulnerability is a "state in which an individual can be placed at risk of harm during their interaction with a service provider due to the presence of personal, situational and market environment factors" and that it can be "permanent, temporary or sporadic, long or short term". Similarly, ISO 31700-1:2023 on "Consumer protection — Privacy by design for consumer goods and services" highlights the influence on the vulnerable state of consumers of "market environment factors" that include "demographic factors, ecological factors, economic factors, socio-cultural factors, political and legal factors, international environments, and technological factors" (p. 6). In 2023, the OECD has dedicated an entire report to consumer vulnerability in the digital age [8], where it emphasizes that "consumer vulnerability online is increasingly systemic [...] even if at times some consumer groups will continue to warrant specific attention" (p. 6). In addition, it remarks that it may be hard to reach general conclusions about certain groups of consumers, such as the elderly, as vulnerability can be context-specific and influenced by a series of factors, while individuals exhibit varying types of behavior.

In conclusion, it is increasingly recognized that a state based view of vulnerability is insufficient to account for digital interactions, which is why this notion is evolving in consumer law regimes around the world [8]. In the EU, this novel conceptualization could also provide an update to the UCPD's notion of vulnerable consumers within the framework of the so-called "fitness check" on digital fairness, carried out by the EC [37]. As we will detail in Section 6, these factors are relevant for understanding vulnerability to dark patterns as well. For instance, people with mental health conditions may be more susceptible to dark patterns in gambling websites because of their tendency to impulsive decisions and their lower self-control.

4. Risk of vulnerability towards manipulative designs and the harms they engender

The idea that vulnerability is not a monolithic concept but a rather complex, stratified one is a useful construct to interpret the various imbalances in human-technology interactions afforded by interface design elements, in particular to correctly identify the risks people are exposed to and the subsequent harms they may incur into. This is necessary to determine and implement measures that avoid or lower such risks and thereby contribute to more responsible, vulnerability aware design.

4.1. Harms caused by deceptive design patterns

Various authors have proposed the categorization of the harms engendered by dark patterns [1,2,38], but they have not linked them to the sources of vulnerabilities, which is our intent. The OECD [2] arranges the harms in three broad categories, namely (i) those that affect consumer autonomy, (ii) those that cause personal consumer detriment and (iii) those that engender structural consumer detriment. The subversion of individual's autonomy and decision-making through a more or less overt influence (e.g. a transparent forced action versus a hidden "sneaking" feature) is a constituent element of dark patterns [2,38]. Individual harms may include financial losses caused by purchasing unneeded or unsuitable products, receiving items and services of unacceptable low or poor value or quality, spending more than intended, searching for less alternatives [12], and can be provoked by deceptive designs such as subscription traps, unfavorable pre-selections, urgency inducing elements, and confirmshaming.

Further individual harms include psychological detriment that depends on emotional distress, such as anxiety [39], cognitive burden due to the spending of unnecessary time and attention [1] (for example, when obstructive dark patterns add unnecessary friction to online processes), including behavioral addictions experienced on social media platforms and videogames.

Then, there are privacy harms, that have been convincingly related to those deceptive designs that are present in consent interactions, exit requests, and user settings [38], even though such harms may be more difficult to identify, quantify, and prove [40], and therefore to compensate, especially if they cannot be robustly linked to concrete financial or material harms or damages. Moreover, data gathering processes may be invisible to users, while their correlated harms may only occur in the distant future [40]. Privacy harms can also cause secondary harms ranging from reputational harms, psychological harms (e.g. embarrassment, anxiety, fear), autonomy harms (e.g. lack of control) and discrimination harms [41]. What makes it worse, is that privacy harms may be impossible to avoid even for informed and careful consumers, as power asymmetries and lock-in effects do not enable users or make it very costly for them to switch between providers [40].

The third category proposed by the OECD concerns the harms that have "a cumulative impact on consumers collectively, even where they [are] imperceptible harms at the individual level" (p.26) [2], namely weaker and distorted competition and loss of consumer trust. In this regard it must be noted that often the harms that consumers endure

online are mostly embedded in micro-transactions which, considered individually, may *per se* not be sufficient enough to motivate a consumer to take action, denounce the malicious deed and report it to the relevant authority or ask for redress, regardless of the fact that the cumulative effect may be quite big or extensive. Moreover, there often needs to be a minimal threshold of materiality, significance, or severity of damage present before one can obtain redress, as is for example, the case with GDPR infringements [38]. Hence, even in the rare cases when users are aware that they have been impacted, it may be too expensive to seek justice for individual dark patterns they encounter in online interactions.

4.2. The two-fold vulnerability to dark patterns' harms: likelihood and severity

There are at least two ways in which vulnerability threats can be experienced: the susceptibility to the influence of dark patterns and the severity of their effect [12]. First, although everyone is dispositionally vulnerable to the influence of manipulative design patterns because of common cognitive biases or market conditions, some people under certain circumstances are more likely to be influenced or deceive., For example, evidence shows that the risk is higher when they have lower educational levels, are older, or are under time pressure [37].

Second, although everyone can be harmed by the use of dark patterns, for someone the detriment may be more severe, for instance because they are less able to recover from a negative experience, like spending more than they can afford, or because they may have less access to remedies [12]. For instance, recent statistics published by the French data protection authority show that most complainants in 2021 were managers and had an elevated level of education, which may be correlated to the digital skills that are needed to submit complaints online [42]. This shows that there are variations in the extent to which people can uphold their rights. Moreover, both primary and secondary harms can exert varying impacts on different people. For example, survivors of domestic abuse may also suffer from physical harm because of the reckless use of privacy-invasivee defaults in certain applications that broadly share one's people position and other personal details by default, often in an invisible manner to the user [43]. The two-fold nature of vulnerability closely recalls the definition of risk which is determined by its likelihood of occurrence and the impact it may have [44].

5. Risk assessment and the conceptualization of vulnerability in the regulation of manipulative designs

5.1. Risk assessment as safeguard in technology development

Being able to determine how harms may disproportionately affect certain individuals or groups is an integral part of the process of evaluation of risk which encompasses the identification of the risks, the evaluation of their impact and the establishment of appropriate mitigation measures. Risk assessment is indeed a central instrument in various European legislative instruments that regulate the development and deployment of technologies, such as those that process personal data, digital services and artificial intelligence (AI) [45].

Assessing when the design of technology becomes problematic is a necessary preventive approach that can address at least two challenges raised by dark patterns. Firstly, designers and developers do not employ manipulative designs only when they want to intentionally deceive users, but also when they have good intentions [5,46]. Secondly, the distinction between illegitimate manipulative patterns and legitimate persuasive design techniques that are commonly adopted to steer individuals' actions towards intended goals is so subtle [4,5] that it raises ethical questions. Designers can often not discern whether what they are doing is persuasive or manipulative [5,47], for instance in the case of emotional design [48] and friction design [25]. The lack of clearcut distinctions makes it hard to regulate interface design with straightforward rules and blacklists. What constitutes a manipulative, illegitimate (and even unlawful) design element may be context dependent and, in some cases, difficult to gauge objectively. This is where a risk-based regulatory approach results more useful than a rule- based one, because it can better adapt to the ever-evolving nature of digital technologies and is therefore more flexible and future proof [49].

In the following sections, we briefly analyze three EU regulations that (i) introduce risk assessment to identify the mitigation measures against the risks entailed by emerging technologies, services or processing operations; (ii) define vulnerability; (iii) constitute relevant legal instruments to contrast dark patterns, as they contain provisions against manipulative digital interface designs. The selection is premised on the General Data Protection Regulation, the Digital Services Act and the AI Act. The analysis intends to show that, first, digital manipulation through the design of technologies is a growing regulatory concern for the EU policy- makers who make attempts to prohibit problematic design practices in favor of fairer ones; second, the three legal instruments reflect that the notion of vulnerability is evolving from a rigid particularistic view into a more nuanced approach, but without necessarily converging; third, it is necessary to com- pose a realistic, layered perspective on the concept of digital vulnerability to be able to assess risks in an accurate manner.

5.2. The GDPR

5.2.1. Vulnerable groups in data protection law

Data protection law is meant to shield people from data practices that may erode their rights related to their personal information and their freedoms or weaken their decision making ability [50], which is what privacy- related dark patterns indeed seek to achieve. As anticipated in Section 2, the GDPR explicitly acknowledges that certain groups deserve strengthened protection, namely children, because of their lower awareness of the risks inherent to data processing and of their rights (Recital 38). Such a class-based perspective is reinforced by official interpretations of the legislation, like in the EDPB's Guidelines on Data Protection by Default and by Design [51] and the already cited Guidelines on Deceptive Design Patterns [11], where there is an emphasis on the necessity to provide "specific protection" to children under 18 and other vulnerable groups. This perspective recalls consumer protection views where children are regarded as vulnerable consumers because of their lack of experience and their lower ability to resist influence [8].

5.2.2. Contextual risk assessment to account for nuances of vulnerability

The lack of awareness and understanding of the consequences of data processing and the existence of legal rights can befall anybody. In fact, as noted by Malgieri and Niklas [14], the GDPR also includes a more nuanced view of vulnerability, since the notion of risk assessment is central to enable the effective protection of the people whose data is processed. If we understand vulnerable people as those exposed to higher risks of damages, then the risk-based approach in the GDPR "can play a significant role in recognizing and conceptualizing the variety of risks (and layers) that can amplify, expose and exploit different vulnerabilities" ([14], p.11). Indeed, the risk analysis carried out by the data controllers needs to consider the "varying likelihood and severity [of the risks] for the rights and freedoms of natural persons" (Article 24), namely how the risks related to data processing may concretely exert different impacts under different circumstances. With the goal of yielding the appropriate measures for the mitigation of the identified risks, such an assessment must be contextual by factoring in the specific nature, scope, context, and purpose of the processing, and continuous by considering the state of the art [51].

Further, the processing of personal data of vulnerable individuals may result in a high-risk activity for the involved individuals. This is one of the conditions under which a Data Protection Impact Assessment

(DPIA) must be carried out (Article 35) to enable the design of suitable, contextual mitigation measures. The guidelines on DPIA [13] contain a non-exhaustive list of specific vulnerable people such as children, employees, and "vulnerable segments of the population" such as mentally ill persons, asylum seekers, the elderly, and patients. Nonetheless, they also include power imbalance as a factor of vulnerability, which may make certain people 'unable to easily consent to, or oppose, the processing of their data, or exercise their rights' (p. 10) [13], for example in the employee-employer relation. In this regard the GDPR is clear, as for example consent for the processing of data where there is a clear imbalance between consentees and the organizations requesting consent should be examined very closely, as consent should be freely given (Recital 43 GDPR).

Power asymmetries in digital markets are widespread, though, and can not only be redressed through enhanced transparency about the processing operations and purposes, since the latter seeks to address *informational* asymmetry. Rather, manipulation can be countered by fairness, the main principle of data protection violated by any deceptive design [11]. Fair data management excludes any processing that is "unjustifiably detrimental, unlawfully discriminatory, unexpected, or misleading to the data subjects" (p. 18-19, [51]) and avoids the exploitation of their needs and vulnerabilities [51], the latter being a constitutive element of manipulative practices [52].

The DPIA is likely to be mandatory also in the case of automated decision- making with legal or similar significant effects (Article 22). Dedicated official guidelines [53] have expanded the notion of vulnerable groups be-yond children recognizing that "[p]rocessing that might have little impact on individuals generally may in fact have a significant effect for certain groups of society, such as minority groups or vulnerable adults" (p. 22). When the effects of automated decisions on individuals are assessed, an important factor will be whether the data controller used "knowledge of the vulnerabilities of the data subjects" in a targeted way, such as people in financial difficulties targeted with adverts for high-interest loans.

5.2.3. Legal provisions against dark patterns

The GDPR does not contain explicit references to online manipulation, although it has introduced behaviorally informed provisions aimed at contrasting the use of deceptive design techniques that affect individuals' data privacy. For instance, the principle of data protection by design and by default enshrined in Article 25 seeks to minimize data collection, storage and use as the default situation. Together with the obligations related to data minimization (Article 5(1)(c)) and purpose limitation (Article 5(1)(b)), Article 25 attempts to shield individuals from function creep and abusive data-hungry practices by countering the status quo bias, according to which people tend to stick with the default option provided to them (i.e., the path of least resistance) (see Section 2.1).

Similarly, the notion of unambiguous consent (Article 4) is paramount to combat default effects. In 2019, the Court of Justice of the European Union in a landmark case[7] has provided the interpretation that pre-ticked boxes cannot signify a legally valid consent, because the user is not actively engaged in the decision. Further, the obligations concerning transparency (Article 12), which also pinpoints the notion of informed consent, should also be interpreted as a tentative to contrast hidden, obscure or misleading data practices. Moreover, as mentioned in Section 2, the GDPR mandates to adapt the language and style of communications to children and, more broadly, to any intended audience [30].

In general, though, it is fairness the foundational principle that is violated by any dark pattern [11] since fairness is meant to ensure that personal information is "not processed in a way that is unjustifiably detrimental, unlawfully discriminatory, unexpected or misleading to the data subject" (p. 12, [51]).

5.2.4. A combination of universalistic and particularistic views

In conclusion, in the GDPR, the concept of vulnerability is both tied to a 'particularistic' approach that explicitly mentions children and other groups, as in need of strengthened protections, and to a 'universalistic' approach that is based on contextual risk management. This is why Malgieri and Niklas [14] fruitfully adopt Luna's idea of 'layered vulnerabilities' [54] and transpose it to the data protection domain, where everyone is deemed vulnerable because of the general "inferiority, dependency, and subjugation of individuals in the context of processing data" (p. 16) [14], but where some are more vulnerable than others, depending on internal and external factors.

5.3. The Digital Services Act

5.3.1. Definition and prohibition of dark patterns

The Digital Services Act (DSA) sets the groundbreaking record of being one of the first EU regulations that explicitly refers to "dark patterns on online interfaces of online platforms", defined in Recital 67 as "practices that materially distort or impair, either on purpose or in effect, the ability of recipients of the service to make au-tonomous and informed choices or decisions". The DSA aims to protect individuals' autonomy from the undue influence of online intermediaries, which is one of the main harms engendered by dark patterns (see Section 4), by prohibiting deception and manipulation in the design, organization, or operation of the online interfaces of platforms (Article 25 (1)). However, the scope of application of these provisions is still debated, as Article 25 (2) clearly excludes the practices already covered by the UCPD and the GDPR, wherein most dark patterns fall under the scope of those two regulations [55]. Article 25(3) grants the power to the Commission to develop guidelines clarifying how Article 25(1) applies to the asymmetric presentation of choices, that is to say assigning (visual, auditory, etc.) prominence of certain options over others and making it more difficult and time-consuming to select certain options; to nagging users with repetitive re-quests which interfere with the user experience; or to making the cancellation of a subscription overly difficult.

5.3.2. Systemic risk assessment of manipulative designs

Although the DSA does not explicitly acknowledge that dark patterns may have layered effects on different users, it does recognize the necessity of contextual risk assessment when the design of widely adopted digital ser-vices can impact the vulnerability of people. Of interest is the introduction in DSA of the obligation for very large online platforms and for very large online search engines to assess systemic risks arising from content moderation, recommender systems, advertising, and other parts of the design of their services. This duty is combined with the obligation of applying reason- able, proportionate, and effective mitigation measures (Article 35) that include the adaptation of the design, features and functioning of the services, including their online interfaces (Article 34 (1)(a)).

5.3.3. Vulnerability to online manipulation

One of the categories of systemic risks identified in DSA concerns the impact of a service on fundamental rights, including the right to privacy and data protection, non-discrimination and consumer protection, and the rights of children, among others. The emphasis is placed on the specific risks that children may incur into, which may depend on the "design of online interfaces which intentionally or unintentionally exploit weaknesses and inexperience of minors or which may cause addictive behavior" (Recital 81) and impairs their health, physical, mental and moral development. This does not merely refer to the impact of dark patterns on children but to that of online manipulation in

[7] C-673/17 - Planet49. Judgment of the Court (Grand Chamber) of 1 October 2019 Bundesverband der Verbraucherzentralen und Verbraucherverbände - Verbraucherzentrale Bundesverband e.V. v Planet49 GmbH (Court of Justice of the European Union October 1, 2019).

general, which includes the harmful effects of online hate speech (Recital 62), advertisement (Recital 95), and disinformation (Recital 104). In line with such a reading, Article 28 introduces an obligation for providers of online platforms to "ensure a high level of privacy, safety, and security of minors".

Although minors may be particularly exposed to the nefarious consequences of addictive and exploitative designs, all adults may be susceptible to them, given that they may be similarly inexperienced or unable to protect themselves from the effects of online manipulation. The DSA has an opening in this respect: Recital 83 acknowledges that an additional systemic risk (identified in Article 34(1)(d)) derives from the "design, functioning or use, including through manipulation, of very large online platforms and of very large online search engines" that can engender serious negative consequences, for instance on a "person's physical and mental well-being, or on gender-based violence". Such risks may also originate from "online interface design that may stimulate behavioural addictions", such as design patterns meant to make users increase the time they spend on a certain service.

Only in relation to targeted advertisement there is a specific reference to individuals' vulnerabilities (Recital 69), because it may negatively impact certain groups and amplify societal harms. This is why profiling for marketing purposes cannot be based on sensitive data (Article 26(3)). Since technological evolution and big data gathering will make it increasingly easier to tailor dark patterns to target specific user characteristics and vulnerabilities [2,8,37], thereby increasing their potential for harm, an anticipatory perspective would caution to foresee and mitigate such developments proactively even beyond advertisement.

5.3.4. Contextual, empirically-based risk assessment

It is difficult to anticipate at this point how the risk assessment should be carried out in practice and how to ensure that it accurately accounts for vulnerabilities. Article 34(2) provides a non-exhaustive list of elements to factor in the risk evaluation. Yet, the focus is on the internal design features of the system, while the other elements that could contribute to the risk (e.g., users' digital literacy) are only vaguely alluded to in Recital 84 [56]. That said, Recital 90 shows an interesting opening to the engagement with impacted stakeholders: it suggests that very large online platforms should conduct their risk assessments, and design their risk mitigation measures "based on the best available information and scientific insights and that they test their assumptions with the groups most impacted by the risks and the measures they take" and "with the involvement of representatives of the recipients of the service, representatives of groups potentially impacted by their services, independent experts and civil society organisations" through "surveys, focus groups, round tables, and other consultation and design methods". Such a participatory assessment must be carried out periodically and "in any event prior to deploying functionalities that are likely to have a critical impact on the risks identified" (Article 34(1)).

5.3.5. An opening towards an empirically-based determination of vulnerabilities for accurate risk assessment

To sum up, the DSA requires very large online platforms and very large search engines to carry out a contextual, systemic risk assessment to avoid all forms of online manipulation, including dark patterns, when an impact on fundamental rights is foreseeable. Albeit the regulation mainly identifies internal risks deriving from the design of the system, it also opens up to the necessity of involving the impacted stakeholders to provide data-informed and scientifically grounded measures for risk appraisal and mitigation. This indicates sensitivity towards the concrete impact that digital services, especially very large ones, may have on various consumers. However, digital services that do not fall under such categories are exempted from such obligations while the scope of application of the prohibitions contained in Article 25 is narrow.

5.4. The Artificial Intelligence Act

5.4.1. Towards a more nuanced understanding of vulnerability

Although the Artificial Intelligence Act (AI Act) Proposal launched in April 2021[8] adopted a class-based interpretation of vulnerability dependent on age or disability status, the final text, approved by the European Parliament in March 2024 aims for a more nuanced view of vulnerability. In line with the GDPR's provisions on automated decision-making, Article 5(1)(b) prohibits AI systems that exploit human vulnerabilities, due to their age or disability, or a specific social or economic situation, to materially distort people's behaviours "in a manner that causes or is reasonably likely to cause [...] significant harm".

Such an approach demonstrates the recognition that even certain situations, and not only a status, can expose people to the manipulative dynamics of AI systems, such as the emotional steering [57] and cuteness [58] that e.g., social robots and conversational agents can leverage to incentivise certain behaviours (e.g., purchases). What is even more worrying is the hyperpersonalized manipulation that AI systems can engender: there are growing concerns that e.g., smart devices can profile their users through the unique interactions that occur between them and then leverage that knowledge to enhance the effectiveness of their influence [59] by targeting specific vulnerabilities. Considering these concerns, the AI Act misses the chance to expand on the notion of vulnerable people. Unlike the broader definition of the AI Act text approved by the European Parliament June 2023[9] that included personality traits and reference to abilities rather than disabilities, the current definition in Article 5(1)(b) replicates a classification based on specific conditions that may be transitory or permanent, but are nevertheless always proper of certain pre-classified groups, thus not easily generalizable to other cases.

5.4.2. The prohibition of subliminal and manipulative techniques

Article 5(1)(a) uses ambiguous wording to prohibit the use of AI systems that deploy "subliminal techniques beyond a person's consciousness" (i.e., stimuli that people cannot perceive, see Recital 29) or "purposefully manipulative or deceptive techniques" which distort people's behaviour by appreciably impairing their ability of making informed choices. Such expressions are not precisely defined within the document, but the AI Act qualifies such techniques by their effect, since they would direct individuals to take decisions that they would have not taken otherwise and that cause or could reasonably cause significant harm. By adding reference to manipulative strategies that can, but must not, be below the level of consciousness to effectively distort autonomy and by recognising that harm can have a multifaceted nature, the latest amendments reflect a more comprehensive and realistic view of the many ways AI systems can unduly influence human beings.

However, technologies that implement AI are often designed in a way that humans can interpret and understand, such as the humanoid traits assigned to artificial agents to enable smooth, natural interactions (e.g., the unnecessary eyes and human-like voice of social robots) [60]. This is where manipulation can be functional: attributing human physical and mental characteristics to nonhuman entities (i.e., anthropomorphism) helps people to interpret the actions of computer and robots and to create mental models [57] that enable meaningful interactions. This and other sorts of "banal deception" are inescapable for the acceptance and integration of AI systems into our everyday lives: it is hence purposefully engineered into devices like social robots and voice

[8] Proposal for a regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain EU legislative acts.

[9] Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

assistants [60,61]. As a consequence, adopting a literal, but plausible, interpretation of Article 5(1)(a) would have the paradoxical effect of characterizing all AI systems as deceptive, and of reducing the assessment of their prohibition only to the severity of any damage produced (or producible). However, how to determine whether a harm is significant is still open to debate. Conversely, since AI applications generally circumvent human rationality or at least act below the awareness level [62], the harms that are inherent to subliminal techniques appear difficult to identify and quantify, thereby threatening the applicability of Article 5(a).

5.4.3. A selective and yet uncertain risk-based approach

To sum up, the AI Act adopts a hard approach to AI technologies that have the purpose or effect of infringing individuals' autonomy and cause significant harm through manipulation or exploitation of their vulnerability. These are prohibited, as they pose an unacceptable risk to the health, safety, and fundamental rights of individuals. Apart from prohibited systems, the AI Act also foresees certain categories of AI systems that are classified as high-risk . Whether an AI system falls under the category of high-risk has been predetermined by the legislator (in Annex I, Annex III, and Article 6) and must be assessed to determine whether the harm they can cause is "significant". This is a challenging aspect, since when the harm cannot be classified as significant, including when it does not materially influence the outcome of decision-making (Article 6(3)), an AI system is not regarded as high-risk and is thus subject to less stringent oversight, since there is only a general obligation to transparency. This implicitly suggests that if the artificial nature of the generated outputs is made clear, the adverse effects of an AI system are automatically reduced [63]. However, given that dynamics and features can influence individuals in subtle ways, the mere awareness of the artificial nature of the system or of the outcome are not sufficient to prevent possible manipulative drifts [64,65].

Therefore, while the subdivision into risk classes seems to be capillary, determining the level of harm, specifically in terms of manipulative power, is still critical. The uncertainty on the assessment of such harms [66] casts doubts on the effectiveness of the AI Act [67–70] and calls for the determination of the actual drivers of vulnerability to digital manipulation.

5.5. Brief conclusions on the three regulatory approaches

The GDPR and the DSA specifically refer to children as a vulnerable category of individuals that require additional protection due to their inexperience, susceptibility to influence, and other internal characteristics regarded as proper of non-adults, which constitutes a particularistic approach. However, both instruments introduce risk assessment as a mandatory measure under certain conditions of potential harm, which suggests a layered approach that takes into account the composite nature of vulnerability in digital environments. The AI Act also attempts to move away from a particularistic approach by prohibiting AI systems that exploit human vulnerabilities that concern personal traits such as age and disability, as well as situational factors that may be temporary. In all three regulations, the evaluation of risk appears to be based on human characteristics that encompass both internal and external drivers of vulnerability, as well as on the design features of the technological system at hand. Following the tendency observable in consumer protection policies (see Section 3.2), the concept of vulnerability to digital technologies starts moving away from a class-based perspective, even though this tradition is still entrenched in the three normative instruments under analysis. Moreover, the GDPR and the AI Act find it relevant to make a distinction, even if implicit, between vulnerable subjects and those that can recur to rationality to shield themselves from manipulative techniques, for example, thanks to increased transparency on the logic involved or on the output of the AI-based decision. Thus, laying down a reasoned, realistic mapping of the various drivers of vulnerability to manipulation exerted by interface design in digital settings becomes necessary for guiding organizations in their risk assessment practices.

6. Factors influencing the risk of vulnerability to manipulative designs

This section will illustrate the elements that must be factored in the evaluation of the threats posed by deceptive design patterns. We will complement vulnerability theories with the scientific perspectives of human-computer interaction and ecological psychology, that maintain that the immediate surroundings of users influence their perception [71,72] and their actions [73]. Their environment impacts their use of technologies, which can lead to vulnerability. For instance, a cookie banner that presents the rejection option with an inconspicuous link instead of a salient button might prevent users from noticing it, especially those with visual impairment, leading them to accept privacy-invasive cookies that weaken the protection of their personal data. We posit that understanding how technologies can trigger vulnerabilities can usefully support risk assessment practices and the identification of the ways people can become resilient to technology-mediated manipulation harms.

To understand the different factors that can drive vulnerability to deceptive designs we leverage the Ecological System Model [74] and ecological theory of affordances [71,72]. The way people perceive, interact, and experience technology depends on their environment in every system, as defined by Bronfenbrenner [74]: micro, meso, macro, exo and chronosystem [72,75–77]. The combination of the elements belonging to such systems increases the likelihood of being affected by manipulative designs as well as their impacts. In the domains of human-computer interaction, computer-mediated technology and computer-supported collaborative work, this framework has been expanded to include the mediated aspect of technology through the interactions or informatics layer [78], or the techno-subsystem [76]. This informatics layer mediates how users understand and interact with technology: deceptive designs are situated in this layer, and therefore, mediate the opportunities for vulnerability through the interaction.

We have taken inspiration in how this model has been applied in HCI and the design of technologies in contexts like design for healthcare [79], personal informatics [78] or socio-digital inequalities [72], and we apply it to the context of interaction with deceptive designs. As shown in Fig. 1, the macro-system includes the macro conditions where users and technology interact, like the economic and regulatory systems (see Section 6.1). Meso conditions pertain to the environment in which the interaction with technology occurs (e.g., communities and neighborhoods), namely the "everyday social, physical, and technical environment in which people live their lives" [72] (p. 24) (see Section 6.2). The micro-system refers to the individual and their specific personal conditions of users, such as temporary states like stress or fatigue, or more or less permanent states, such as mental health conditions (see Section 6.3). Exosystems are events or structures that indirectly affect the user, like the user's family workplace [76]. Chronosystems refer to how temporal changes in the environment influence the individual. Since exosystems and chronosystems are related to other conditions pertaining to other systems, such as socio-economic status or age, and they add an unnecessary layer of complexity to our model, we find it useful to confine our discussion to the micro, meso and macro systems which already provide a helpful framework of reference to understand vulnerability to deceptive designs. We situate deceptive designs in the informatics-system that is influenced by the macro-system. Moreover, the informatics-systems have a bidirectional influence over micro and meso-system, since deceptive designs affect the individual interaction of users, but at the same time factors of the meso-system shape this relationship and understanding of deceptive designs.

For us, vulnerability to deceptive designs occurs when one or more factors operate, giving rise to specific contextual and situated experiences that makes users vulnerable, following the ideas of drivers of

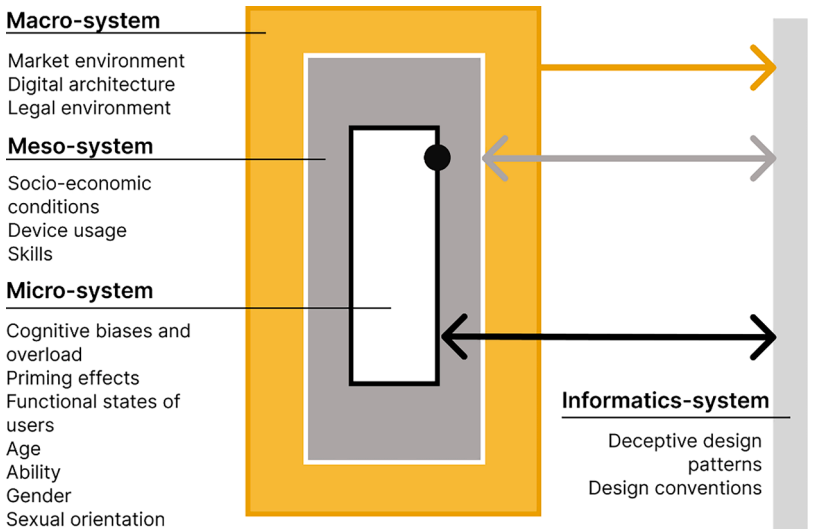

Fig. 1. The ecological systems theory model applied to the development of vulnerability to deceptive design. As the arrows show, the factors pertaining to the macro-system have an influence on the informatics-system and deceptive design deployment. The factors of the micro-system and the meso-system have an influence on the informatics-system and vice versa. At the same time, factors of the meso-system also influence the micro-system by mediating the users' interaction with deceptive patterns in the informatics system, represented with the dot.

vulnerability from Malgieri [80]. Hence, our intention is not to make an exhaustive list of conditions that strictly belong to one or another category, but rather to illustrate the complexity of vulnerability as a human condition within the (online and offline) systems where humans live and interact. In the following sections we explain how the interrelation of factors in every system contributes to experiences of vulnerability to deceptive design patterns.

6.1. Macro-factors: interface design is never neutral. The role of digital architecture in increasing vulnerability

The digitalization of all aspects of society such as banking, commerce, leisure, healthcare, and public administration is forcing individuals to continuously engage in digital transactions. The asymmetries with digital services and the very nature of the market [10] put individuals in situations of vulnerability [8], even though some people may be in a riskier position than others, given, for example, their level of credulity, or their difficulties navigating the Internet. For instance, in the context of online manipulation, older adults have shown the desire of disengaging from technology, reclaiming a right not to be bothered by manipulative practices. They express frustration because they feel like outsiders in the use of imposed technologies [81].

Market conditions and modern society's tendencies also result in a race to the bottom that causes dark patterns to be omnipresent in digital services. The concentration of power and information; the widespread reliance on personal data collection aimed at offering personalized advertising online; and the pervasive use of e-commerce services and social media are some of the market conditions that ease the proliferation of dark patterns. For instance, a 2022 European Commission's report [37] shows that almost all the most popular applications in the EU contain at least one deceptive design. In such settings, the complexity and interconnectedness of digital products embody higher risks [8], that can additionally increase the power asymmetries when users become dependent on such products [10]. For example, because their social networks heavily rely on a certain product [82] or because their employer mandates such use.

The set of actors, norms, architecture, and market that constitute the *digital architecture* is conceived to shape user interactions and afford certain actions, thereby impacting the behavior of individuals [83,84]. Thus, the emergence of potential deceptive elements that aggravate the vulnerability of users are determined by the various stakeholders that participate in the design of technologies at large, the position of the designer within the organisations that generate and commercialize certain designs, and the context in which the design activities take place [5,10,46,84–87].

In addition, the *legal norms* that regulate digital architectures also influence the protections embedded in the design of digital services, as the example of privacy preserving defaults illustrated in Section 5.2 shows. But they may also engender the paradoxical effect of affording certain deceptive practices. For instance, when transparency requirements produce off-putting lengthy explanations and long lists of options that fatigue readers (the so-called "transparency paradox" [31]) rather than enhancing their autonomous, informed decision-making capabilities.

6.2. Interaction between meso-factors and informatics-system: users perceive and resist deceptive designs differently. Intersectionality in vulnerability

People interact with interfaces through "affordances" and "signifiers" that enable, or disallow them to interact in specific ways [73,88]. "Affordances are opportunities for behaviours" [89] (p.189); therefore, certain interface design elements "afford" users different opportunities. However, these possibilities for action must be highlighted with "signifiers" to be perceived by users, for example, to discover more information, indicate or signify a choice, or accept or refuse the terms of a contract. In this section, we explore (i) how the design of the interface, embedded in the digital architectures, might increase the conditions of vulnerability and exploit them (ii) and how users' material conditions impact how they perceive and interact with online services.

6.2.1. Design conventions: visual perception based on salience and spatial organization of the elements on the screen

When it comes to graphical user interfaces, visual salience and cluttering are the main factors that channel attention to certain areas of the visual field [90]. The concept of visual salience describes the features of an object or region in a visual space that is distinct from its surroundings. Cluttering refers to the number of objects in that space and the complexity of the information organization, which makes visual tasks, such as searching for a stimulus (e.g., like a link or button on the screen), more difficult [90].

These two concepts can be engineered to drive users' attention

towards predetermined choices [91]. Studies show the influence of visual salience on users' shopping task performance [92], on choice [93,94] and recall [95] and on decision-making processes in general [96]. Cluttering can distort the processing of information and nudge the selection of one option over the other because it raises users' cognitive load such as amount of mental effort required to process and understand information. As a result, users may divert from activities not aligned with the system designer's goals. For example, the "disguised ad" patterns play on visual salience [4]. The user can be misguided by the false similarity between the elements of the interface and an external ad, and consequently, unintentionally click on it. Similarly, Bosch et al. [15] identify the wall of text of privacy policies as a dark pattern that overburdens readers and can make them abstain from engaging with the legal terms.

Visual salience of certain interface elements can be used to present options in a way that can nudge people [4] to select certain products or services over others or to take privacy invasive decisions instead of privacy friendly ones. Somehow complementary to this strategy is the low contrast between critical text information and layout, which makes certain options available but barely visible. For example, consent refusal on cookie banners or unsubscribe options from newsletters, therefore more costly for users in terms of time and cognitive effort. Jarovsky [97] also adds the contrast effect that plays on subtle color and contrast schemes to give or conceal visual saliency to some elements.

Even though the visual elements and how the choice architecture is presented in the user interface matter, there is no deterministic effect between one design element and the influence on behaviour, since personal factors also play a role. Although the literature that demonstrates the effects of manipulative designs is still growing [98], we can leverage some evidence concerning how manipulative design elements affect the users' choices online. Some studies have demonstrated that removing option buttons increases users' cookie consent [18,99] or privacy-invasive decisions when dark patterns are combined [100]. Luguri and Strahilevitz [100] found that playing with the information shown to the user (i.e., hidden information) and the availability of options (i.e., default choices), doubled the users' acceptance rates on subscriptions, while highlighting the "accept" option had no significant impact. Berens et al. [101] found that, in cookie consent banners, the reject option as a link instead of a button changes the responses significantly. They also found a weaker impact when the "accept" button is highlighted and the "reject" button is next to it.

Although there is still room to disentangle the effects, the experiences of users converge: users find it hard to identify manipulative design elements [27,100,102,103] and, even when they aware of their presence, feel powerless to counter them [27,102,104].

6.2.2. Socio-digital inequalities as internal meso-factor

Users' perception also depends on their immediate surroundings and other contextual factors. Belonging to a group that shapes their identity implies a set of values, norms, and attitudes that influence how users interact with technology [72]. When different conditions of class, education, race, or gender intersect [88], they may give rise to an intersectionality of vulnerabilities in the online realm. The outcomes of the use of technology (i.e., the extent to which users can take economic, cultural, or well-being benefits from technology) are determined by *socio-economic inequalities* [105–107] that are referred to as "digital inequalities" [105].

The experience and skills that are needed to navigate the digital environment are determined by formal and informal education, including family environment, support from friends, and personal attitudes [108]. These factors are all strictly related to socioeconomic conditions and impact how people process, understand, and relate to online and digital information. People living in an underprivileged state with less material, temporal, social, or cultural resources, find more complicated to develop digital skills to protect themselves online and to cope with harms afterwards [105,108–112]. Socio-digital inequalities are a meso-factor that in combination with the rest of the systems drives vulnerability.

As defined in the UCPD (see Section 3), "credulity" is one of the conditions of the "vulnerable consumer" and may be related to the level of education and digital literacy. Credulity may prevent users from understanding interfaces, the potential risks and harms from online interactions, and the business models that shape the design of an online interface [8,104]. On the contrary, living as a highly educated person in a resource-rich environment will enable individuals to be more cognizant of the benefits that the internet provides [105,108,109,113], as well as enable them to know how to be protected from online harm.

Prior studies show that the educational level of people plays a role in the resilience of people against dark patterns. For instance, Bongard-Blanchy et al. [27] found that people with an educational level lower than the Bachelor degree level were less likely to identify dark patterns online, while Luguri and Strahilevitz [100] showed that a lower level of education increased the likelihood of accepting privacy- invasive options. Conversely, Zac et al. [114] did not find significant results that relate education and income to dark patterns resistance.

These studies provide preliminary indications that certain socio-economic conditions can strengthen the influence of interface design, which in turn can aggravate the state of vulnerability, for example when it weakens people's privacy. Thus, DiPaola and Calo [115] point at the idea of socio-digital vulnerability: the environment mediates creating more opportunities for exploiting vulnerabilities and create new threats. Similarly, when Sanchez Chamorro et al. [81] explain the social component of resistance to manipulative design of teenagers at risk of social exclusion, they highlight the impact of socio-digital inequalities as a reinforcing factor of vulnerability online. Teenagers learn about the deceptive designs' risks from family and friends, who also help them to cope with the effects of these designs. However, if the environment is less educated or acquainted with technology, the effect might be the opposite. Similarly, the Stigler Committee [116] also notes that dark patterns imposing transaction costs, like cumbersome opt-out options, may particularly impact less tech-savvy users such as elderly or less educated people.

The socio-economic status also influences the *use of devices* since mobile phones are the most common devices among individuals pertaining to lower SES families [117]. Given that manipulative and addictive elements present in mobile applications are varied and pervasive [118], they are more likely to affect individuals from lower socio-economic classes [119]. Similarly, Radesky et al. [120] report how children of families in a lower socio-economic situation would be more likely to find mobile applications with manipulative designs. This may hint at a greater risk for those that mostly or exclusively recur to mobile devices to access digital services, which may predominantly coincide with a certain socio-economic status.

6.3. Micro-factors: internal human factors that influence vulnerability online

There are a number of human cognitive and perceptual elements that can influence the predisposition to being vulnerable to dark patterns. In this section, without the pretense of being exhaustive, we will analyze individuals' features that encompass cognitive biases and cognitive overload, priming effects, functional states and other personal conditions. Some of these factors may be permanent, some may be temporary, and some may evolve over time (e.g., age).

6.3.1. Cognitive biases and cognitive overload

As contemplated in Section 2, most of the literature in this domain focuses only on complex high-level psychological features, namely on cognitive biases [6,7,15,100] and psychological needs [15]. It has been noted that, even though cognitive biases are intrinsic to the human nature and are exacerbated by the online environment, not everyone suffers from all these biases, or we may do so with great variation [12].

Many cognitive biases and heuristics have been identified as having a primary role in the success of online manipulative designs that negatively impact behaviors and judgements. In terms of online choices related to one's personal data, Waldman [7] explains the main pervasive cognitive barrier as anchoring, hyperbolic discounting, loss-gain framing, and over choice. Anchoring makes people over-rely on available information when making their choices (e.g. what other people do), instead of basing their choices on the actual relevance of that information for the situation at hand. Hyperbolic discounting entails the overestimation of the immediate benefits of a certain action, while underestimating its future consequences – this is why people accept extensive digital tracking against free internet content. In loss-gain framing only the positive effects of a certain action are provided or highlighted, while the negative ones are glossed over or vice versa – for example in cookie banners this tactic is often employed to nudge people to consent to personalized advertisement [121,122]. Lastly, in over choice the excessive number of choices overwhelms and paralyzes users, for instance in terms of mobile application permissions and cookie installation, instead of enhancing their autonomous decision-making.

In their literature review of nudges used in privacy and security for more or less praiseworthy objectives, Acquisti et al. [17] identify additional hurdles. First, loss aversion, which causes people to value the personal information they have lost less in comparison to the information they still have (and thus resist losing it). Second, optimism bias leads people to take unjustified privacy risks based on wrong estimations of their chances of undergoing a negative event. Third, status quo bias which pushes people to stick with default options, like privacy-invasive pre-ticked boxes.

Concerning purchases and decisions on e-commerce services, many cognitive biases have been identified as having a role in transactional decisions, such as overconfidence, present bias and loss aversion [12]. Mathur et al. [6] add to the list more cognitive vulnerabilities. The bandwagon effect refers to when people value something more because others value it; the scarcity bias pushes people to value more what is in short supply; and according to the sunk cost fallacy, people continue with a course of action if they have invested resources into it, even when it is not reasonably worthwhile. Somehow related to this is the restraint bias, which designates people's tendency to overestimate their capacity for impulse control. This is a widely studied effect in addiction settings [97] and plays a role in purchasing and time spending decisions.

6.3.2. *Priming effects on purchase decisions*

Priming is the effect of stimulus exposure to the response of a later stimulus [123,124]. Research has demonstrated that the perception of a subsequent stimulus, known as the target, can be influenced by a preceding stimulus, referred to as the prime, even when the prime is visually masked to reduce its visibility or presented very briefly before the main stimulus [125]. However, even when the priming stimulus and the target are separated for a longer period of time and are thus distinguished, the prime is not necessarily processed in an active way [126]. That means, for example, that primes can go unnoticed, but be effective in terms of directing outcomes [127].

The concept of priming has been at the center of much research and practice in product advertising [128–130], where it can be considered as an instrument of persuasion [131] which is happening outside of the individual's conscious control. Petticrew et al. [132] and Costello et al. [133] demonstrate the use of the priming effect to implement 'dark nudges', i.e., possibly manipulative primes, which could direct users' actions in predetermined directions.

In the domains of e-commerce and online advertising, it has been proven that numeric and semantic primes can have a large impact on the customer's willingness to pay [134]. People mostly interact with e-commerce applications via screens and interfaces, the use of colour, form and other visual-oriented primes [125,135,136] in the choice architecture of interfaces can direct the customers' choice towards certain products. The typical tasks performed on e-commerce websites include primarily the visual evaluation of a product through photo galleries and carousels and easily accessible paying options available in a specific, visible compartment. Subliminal priming effectively affects the immediate decision of the user only when the purchase decision will directly follow the persuasive attempt [137], which can be easily engineered in the choice architecture of a digital application.

6.3.3. *Functional states of users*

The state of individuals affects task performance in complex computer-based systems [138]. Studies show that under conditions of stress, fatigue, and boredom, a deterioration in performance quality and an increased level of errors can be expected [139–142].

In the context of online deceptive designs, the system can be purposely designed as a responsive agent to these states. For example, several deceptive game patterns operate under the users' state of boredom [143] and even induce it (as in the "Pay to Skip" dark pattern) [144]. User's mental fatigue is also exploited by the so-called 'sneaking' types of patterns [4], which profit from lower levels of attention and cognitive functioning to impose additional purchases of goods or services on unwitting users. Several deceptive patterns that play on a sense of urgency such as fake countdown timers or scarcity such as fake limited availability of a product or offer can also be attributed to the exploitation of these functional states, because their primary goal is, at least partially, to induce stress in users [145] and nudge then to take fast, suboptimal decisions [145], like impulse buying [26].

6.3.4. *Other personal conditions*

A majority of studies in manipulative design that focus on users have looked at effects on behaviour caused by the existence of different UI elements [27,28,100,104,146]. However, very few of them have explored how user characteristics influence the experience when facing manipulative designs, like age [27,147,148], education [27,100] or socio-economic status [27,100].

Some preliminary studies hint towards the need of exploring age as a condition of vulnerability [27,147–149]. However, one cannot only look at age as a driver, but at other elements associated with that age. For example, childhood, adolescence, early youth, maternity and paternity, or retirement, embody situational aspects that can contribute to the experience of vulnerability beyond the age.

Children may be particularly prone to fall prey to manipulative designs due to factors such as their immature executive function, susceptibility to rewards, unfamiliarity with data privacy and lack of understanding of virtual currencies [2]. In this regard, prior studies show that children may be particularly exposed to manipulative design patterns due to the massive targeting in in-app advertisement that they are subject to [120,150] and to the deceptive strategies that are widespread in online games (i.e., loot boxes) [151].

Similarly, for teenagers, there is a social component in their interaction with manipulative designs: they are more exposed to them and tempted to interact with them because of their social relations with family and friends [149]. Sanchez Chamorro et al. [81] resorted to an utilitarian approach to explain why adults are susceptible to some manipulative designs like scarcity cues: adult users buy an item because there is only one good left, and they need it. However, teenagers may be aware of the fake scarcity, but still buy because they would otherwise miss the opportunity of belonging to a group, such as the group that has some cosmetics in a particular videogame. While there is still little research in this regard, among teenagers the utilitarian aspect seems to disappear to leave room for a social one. This social aspect may thus be a specific driver of vulnerability, that is not necessarily only present in teenagers but may be dominant in that age group and further aggravated by other conditions.

On the other hand, older adults may also be exposed to increased risk online, although the evidence in this regard is contrasting. Van Nimwegen and de Wit [148] used an experimental setting and found a negative relation between age and falling for some manipulative

designs: the younger the user, the more likely to fall into some deceptive designs like sneaking products into users' basket and the use of emotions. Conversely, via a survey study, Bongard-Blanchy et al. [27] showed that older participants had more difficulties identifying manipulative designs. Similarly, Avolicino et al. [147] found participants above 35 years old more vulnerable because of their lack of acquaintance with the online travel agencies, which was the context studied.

The difficulty of reaching general conclusions is probably because older age is associated to other conditions, such as disability, mental impairment, and digital exclusion, that can exhibit great variation from one person to the other [8]. For instance, the OECD has shown concern about how users with visual impairments can be affected by manipulative designs [8] and recent evidence shows that there can be indeed a troubling effect of deceptive graphical elements combined with poor design practices that do not ensure accessibility, such as low contrast of refuse links in cookie banners [152].

Even the UK's Competition and Markets Authority (CMA) [12] warns that personal traits (such as age, wealth, and health) as well as temporary situational characteristics (such as unemployment or experiencing scarcity of time, money and social connection) can exacerbate the susceptibility to being influenced by online design choices and can make people obtain worse outcomes from their digital experiences.

6.4. Examples of layered vulnerability

In the following, we gather some examples that illustrate how certain design elements of digital services can pose a threat to anyone, but are particularly risky for certain ones among their users, often due to a combination of factors that represent the (sometimes situational) layers of vulnerability [14,54] that people have.

6.4.1. Manipulative designs can increase risks for victims of intimate partner violence

PenzeyMoog [43] describes the real case of a fitness app that made visible their location with any other user nearby by default with the intention of increasing sociability among its users, such as exchanging with other athletes and enhancing the motivation to perform. However, such a social feature carries a great risk. Location sharing can reveal where a person lives and the route they take regularly for exercising, including weather that is a dark or secluded place. This information can be misused by anyone, but it becomes particularly dangerous in the case of survivors of domestic violence that run away from their abusers. Privacy invasive default settings may effortlessly vanish all the extreme efforts that survivors have made to stay alive. Given that privacy settings are often difficult to find, reach or decipher, which default has been selected for the user is not immediately visible nor easily editable, especially when data (e.g., location) sharing happens in an invisible manner, so the user may become aware of the risk too late or never at all. Although privacy risks can be generalized, for survivors of domestic abuse the effect of a privacy invasive default carries an additional danger to their physical safety.

6.4.2. Manipulative designs can increase risks for LGBTIQ+ community

Another example of how the design of privacy settings may excessively affect certain communities is given in Sannon and Forte [153]. Privacy settings that are difficult to find and manage are costly especially for those users that need to frequently switch between identities, for example because they are forced to conceal or separate certain aspects of their lives (e.g., members of the LGBTQ+ community, political opponents in autocratic regimes, etc.). The ability to adjust privacy preferences often requires a certain level of digital skills, which is often associated with one's socio-economic status, as recalled earlier. Data management may be further complicated by language barriers and the financial cost of connectivity in certain geographical areas. Failure to protect one's own privacy due to design barriers may result in serious implications on the lives and welfare of such communities.

6.4.3. Manipulative designs can increase risks for users with mental disorders

The UK's Money and Mental Health Policy Institute [154] denounce that many online gambling websites contain easy manners to make deposits while they conceive cumbersome, frustrating processes to withdraw funds. Moreover, tools that help setting spending limits and imposing self-exclusion are often hidden or ineffective. Although all users are exposed to such tricks and the nefarious (financial, emotional, relational) consequences of gambling may fall on anyone, people suffering from mental health issues are particularly vulnerable, due to a mix of factors. Common symptoms of such problems are impulsivity and low problem-solving ability which are associated with low control and difficult decision-making, including risk assessment [154]. When coupled with designs that nudge people to continue playing and spend money, the mix is lethal. The problem is exacerbated by online advertising and marketing strategies that are impossible to avoid [154] and lure users into gambling websites by targeting specifically those vulnerabilities.

6.4.4. Manipulative designs can increase risks for individuals with a predisposition to impulse buying

Internal conditions, as predisposition to impulse buying, can be accentuated, and even targeted, by manipulative designs. Moser and colleagues [26] carried out a study on dark patterns used on the top US e-commerce websites, noticing that they often include features that aggressively encourage impulse buying, such as lowering the perceived risk of carrying out transactions online, leveraging social influence and enhancing the perceived local and temporal proximity to the product. The study shows that such practices counter the will and capacities of consumers who desire to curb their impulse buying. What is more is that those that find it difficult to make ends meet experience a more severe detriment of such commercial practices [12].

7. Discussion: towards a multidisciplinary assessment of vulnerability

7.1. Building resilience against vulnerability

Throughout this article we have argued that deceptive design patterns may exert their effects on all users of digital technologies but can also target specific vulnerabilities and thereby be more detrimental to certain individuals or communities. Policy, practice, and research need to assess, and ultimately address, such challenges. It is hence important to discuss the implications of our arguments for building the resilience of individuals and groups against deceptive designs.

Since human beings are embedded in a common social, economic, institutional, and legal fabric, they are by nature interdependent. Their dependence on external conditions is the concrete manifestation of human vulnerability, which is universally present. This is why a state of non-vulnerability is impossible to achieve and should hence not be the goal of policies. Rather, resilience [32] should be promoted, namely the set of all those physical, human, environmental, and social resources that make human agency possible and enable people to cope with the implications of their inescapable vulnerable dimension. Resilience to dark patterns should not only be promoted as an *ex-post* mitigation mechanism, but also as an *ex-ante* solution. Many different actors of the digital economy can supplement the necessary resources to prevent the adverse effects of online manipulation.

7.2. Practical hurdles to risk assessment feasibility

With its purpose of proactive mitigation of the many drivers of vulnerability, risk-based regulation seems fitting to approach the ever-evolving reality of deceptive design patterns in existing and emerging

technologies. Whereas the domain of data protection can count on an established body of risk assessment methods, the implementation of the systemic risk assessment envisioned in the DSA is, at date, still to be defined. Regarding AI systems, the risk appraisal frameworks that are being developed are several and sometimes divergent [155]. Such an uncertainty leaves organizations unsure of how to plan and operate their development and commercialization activities in a safe and compliant manner, especially when it comes to implementing safeguards against digital manipulation, which is a young area of research and practice.

Further, if the issue is not addressed on a systemic level, individual safeguards on individual technologies risk to be ineffective. For instance, market- related macro-conditions foster the global tendency of ad personalization based on massive collection of personal information, while the competition for users' attention gave rise to a race to the bottom with the subsequent proliferation of dark patterns on the overwhelming majority of online services. However, macro conditions are often hard to account for in application-specific risk assessments.

Moreover, when risks are highly contextual, it is inaccurate to only ascribe them to a static set of functionalities of a certain technology (as a rigid reading of the DSA would propose) or to the purpose and domain of use of the technology (as the high-risk categorization of the AI Act suggests). In AI systems, moreover, deception is sometimes a mere strategy for a pleasant, functional interaction design. But it is because of this reason that, if the risks are not counterbalanced or removed *ex-ante*, these technologies can also produce greater harm than those that have been identified as harmful *a priori*. Given that the elements of vulnerability are many and intertwined, continuous assessments may result in a laudable theoretical effort, but be excessively costly and clash with the priorities, needs and timelines of business practice.

Such challenges risk jeopardizing the feasibility of risk assessment. A standardized approach would undoubtedly simplify the procedure and, supposedly, be experienced by organizations as less of a burden. Going beyond mere compliance reasons, technology developers need to be persuaded of the motivations why an accurate, reliable risk assessment can constitute a helpful tool. In addition, there is the need to develop a comprehensive methodology that includes the evaluation of the risks entailed by all sorts of digital manipulation within other kinds of risks, since it would prove burdensome to carry out separate *ad hoc* assessments. Accounting for and understanding the internal constraints for companies and actors involved in the digital architecture when applying risk assessments is a first necessary step, and potential future work we envision.

It is paramount that policymakers as well as civil society take part in the establishment of the methodologies for such an appraisal and the appropriate mitigation measures. Articles 34 and 35 of the DSA, for example, already foresee such a participatory approach, wherein the European Commission as well as the European Board for Digital Services play an active role in the oversight and the recommendation of measures for the management of systemic risks of very large online platforms. A similar role will be taken on by competent authorities in the activities carried out within the regulatory sandboxes established by Article 57 of the AI Act, that will offer controlled environments that enable the testing and validation of innovative AI systems before their placing on the market. Regulatory sandboxes will also provide the possibility to identify risks upfront and devise timely and effective mitigations measures.

7.3. The role of empirical evidence for transdisciplinary action

Existing empirical studies can be helpful in determining a broad range of potential threats that need to be factored in the risk appraisal methods. However, such an approach falls short of being complete, as it excludes all those risks and factors that have not been examined yet. The OECD [8], for instance, observes that most research studies so far have only examined internal drivers of vulnerability (e.g., age, socioeconomic status) while neglecting other relevant elements. There may be additional factors of vulnerability, though, that are difficult to identify, quantify, correlate and report (either directly or indirectly), since they relate to the experience of users and not to an observable change in behaviour. For example, inter-partner abuse victims may pertain to any socioeconomic class and experience privacy harms differently, but it is hard to quantify them in an observable way. Similarly, older adults are not a homogeneous group, and there is not an specific age threshold after which one should be considered an older adult [156]. Therefore, there is not an age in which users automatically become more vulnerable [8], and context plays again an important role. The fact that people can be vulnerable in one situation but not in another makes it cumbersome to exactly determine the drivers of vulnerability.

Social sciences methods and computational methods that collect empirical evidence are the necessary candidates to bridge this gap, as suggested indeed in Recital 90 of the DSA. As pointed out by Gray et al. [98], there is a pressing need for transdisciplinary approaches and knowledge transfer to understand and fight the effects of manipulative designs. For example, design scholarship can help regulators understand the impact of technology design on vulnerabilities. To broaden the understanding of the multifaceted reality created by digital markets, it is crucial to run user studies that include participants other than highly educated populations in developed countries with digital access to survey platforms (e.g., elderly, kids, teenagers, or low-educated people), as well as organize research designs in contexts that can expose vulnerability. In the same way, looking at the experience of interacting with manipulative designs, and not only at the effects in behaviour that design features have, will help to disentangle the contextuality and situatedness of vulnerability to deceptive design patterns. Hence, it is not only about "what makes users vulnerable", but about the "how and why".

Contacting and collaborating with organisations such as NGOs that work with specific populations in real-world contexts can be the first necessary step to carry out this endeavor with all the necessary ethical considerations [157]. The engagement of those that have firsthand experience and knowledge of certain realities is gaining importance for risk assessment. For instance, to determine the dangers that technology can cause to victims of intimate partner violence, Slupska and Tanczer [158] propose to involve affected groups and communities in the traditional threat assessment methodology: they can help mapping out the actual threats and devising measures for those types of harms that are less tangible than financial losses. The engagement with various stakeholders and the evidence that empirical science can thus provide can support realistic risk assessment and nurture good practices.

7.4. Implementing fairness and fair design patterns

Across various domains, there are proposals for a "fairness-by-design" duty [159] that could even be incorporated in the UCPD revision as a general obligation for businesses so that "products, user interfaces and commercial communications [...] be designed in a fair manner" ([160], p. 13). Fairness is more daring and more encompassing than the principle of transparency that has the goal of disclosing how a system or process works. The multidisciplinary community of researchers, regulators, civil rights defenders and businesses who work to contrast digital deception has so far mainly proposed transparency-enhancing measures (see e.g., [11]), which are necessary but not sufficient to fight dark patterns. It is now time to define and apply fair design practices as well as incentives for their adoption and determine their fitness for protecting vulnerable people and increasing their resilience.

The community has achieved astonishing results in the identification and exposure of problematic design practices, for instance through studies aimed at detecting dark patterns automatically [161,162] or at demonstrating their influence on people's decision-making (see Section 6 and for an overview [98]). However, in the already highly regulated digital sphere, there is the urgent need to successfully promote fair design patterns that can be adopted easily and safely by businesses. It is paramount that the high-level (often intertwined) requirements

provided by existing and upcoming regulations are translated into simple, operational instructions, accompanied by examples of good practices, that designers of digital experiences can effortlessly understand and apply.

Proposing one-size-fits-all solutions is not the goal, because there is often a subtle distinction between design patterns that are legitimate and appropriate within a certain context and those that are not. Rather, there needs to be an inventory of good design practices that one can draw from, in combination with knowledge about the harms that prospect users could suffer and about the methods for assessing, mitigating and eliminating risks. For what concerns privacy policies, for instance, there exist libraries of design patterns that collect, organize and make readily available good practices [163], such as the French Data Protection Authority's library on transparency-enhancing design patterns [164]. Even though it is for now easier to find and copy-paste bad practices since they are so widespread and the incentives for their adoption are high, designers and developers should be empowered to reuse fair design patterns and adapt them to their specific contexts.

Designers can also play a crucial role in the implementation of fair design patterns as they can embed awareness to vulnerability factors within the design process, thanks to a wide range of methods through which they can evaluate their work's potential impact. The use of personas, understood as a sort of average user that the planned design would target, and anti-personas, as those users that might be excluded from the planned design, can help to assess the impact of the adoption of a certain design element on a diversified variety of users. Similarly, specific methodologies and toolkits enable the impact assessment for inclusive design and ethical design [165,166], thereby supporting the development of less harmful designs.

Moreover, businesses can empower designers in their decisions and entrust them with "ethics ownership" [5,87,167]. Fostering governance models in organizations where legal departments, decision-makers, and designers work together within a check-and-balance system can help to broaden the adoption of ethical and safe interface designs. Without changing the market conditions and without developing economic incentives for adopting vulnerability-aware fair design patterns, this paradigm shift will not happen. The Digital Markets Act[10] is a first, important step in this direction and, indeed, it contains provisions against dark patterns. Additionally, the Data Governance Act[11] and the fair data economy it aims to foster can promote digital services that are safe and respectful by design: data intermediaries, data cooperatives and data altruistic organizations have the precious opportunity to conceive and design experiences for data sharing and consent that are in stark contrast with the deceptive status quo of data-hungry digital services.

8. Conclusions

Counting on the multidisciplinary expertise of the authors and the transdisciplinary knowledge that was generated through their collaboration, in this article we have argued that the harmful deceptive design patterns in digital services can be more detrimental to certain individuals or communities due to macro, meso and micro conditions. In the age of service personalization and hypernudging, there is the risk that manipulative attempts will increasingly be able to exploit such vulnerabilities to strengthen their effectiveness and weaken people's resilience even more. Risk assessment is becoming inescapable to account for the ever-evolving nature of digital technologies and the vulnerabilities they engender, but there are open questions on how to carry it out in a reliable and practicable manner. All actors of digital markets need to be involved, and held accountable when appropriate, in the creation of fair-by-design experiences.

[10] Regulation (eu) 2022/1925 of the European Parliament and of the Council of 14 september 2022 on con- testable and fair markets in the digital sector and amending directives (eu) 2019/1937 and (eu) 2020/1828 (Digital Markets Act). Published: OJ L 265, 12.10.2022, p. 1–66

[11] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act). PE/85/2021/REV/1 OJ L 152, 3.6.2022, p. 1–44

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.

Acknowledgements

A. Rossi, A. Sergeeva and L. Sánchez Chamorro have received funding from the Luxembourg National Research Fund (FNR)'s grant agreement no. IS/14717072 (DECEPTICON: Deceptive patterns online). A. Fernandez has received funding from the Luxembourg National Research Fund (FNR) (PRIDE17/12251371). R. Carli is supported by the Joint Doctorate grant agreement No 814177 LAST-JD-Rights of Internet of Everything. A. Rossi was also partially funded by the PNRR/NextGenerationEU project "Biorobotics Research and Innovation Engineering Facilities "IR0000036" – CUP J13C22000400007". L. Sánchez Chamorro has also received funding from the Dutch Research Council (NWO) via RESOCIAL project (NWA.1540.21.001), from the thematic program "Vulnerability and resilience in an online society". We thank Sophie Doublet from the University of Luxembourg for her help in designing the graphical abstract. We are grateful to the participants of the Lorentz Center's workshop on "Fair design for online interfaces" (held between 29 January and 2 February 2024) for their feedback on the main findings of this research.

References

[1] Mathur A, Kshirsagar M, Mayer J. What makes a dark pattern... dark? Design attributes, normative considerations, and measurement methods. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Yokohama Japan: ACM; 2021. p. 1–18. https://doi.org/10.1145/3411764.3445610. URL: https://dl.acm.org/doi/10.1145/3411764.3445610.

[2] OECD. Dark commercial patterns. In: Number 336 in OECD Digital Economy Papers; 2022. https://doi.org/10.1787/44f5e846-en. URL:.

[3] Chivukula SS, Watkins C, McKay L, Gray CM. "Nothing comes before profit": asshole design in the wild. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. ACM; 2019. p. 1–6. https://doi.org/10.1145/3290607.3312863. URL: https://dl.acm.org/doi/10.1145/3290607.3312863.

[4] Gray CM, Kou Y, Battles B, Hoggatt J, Toombs AL. The dark (patterns) side of UX design. In: Proceedings of the 2018 CHI conference on human factors in computing systems; 2018. p. 1–14.

[5] Sánchez Chamorro L, Bongard-Blanchy K, Koenig V. Ethical tensions in UX design practice: exploring the fine line between persuasion and manipulation in online interfaces. Designing Interactive Systems (DIS). Pittsburgh PA USA: ACM; 2023. p. 15. https://doi.org/10.1145/3563657.3596013. ACM, NewYork,NY, USApages.

[6] Mathur A, Acar G, Friedman MJ, Lucherini E, Mayer J, Chetty M, Narayanan A. Dark patterns at scale: findings from a crawl of 11k shopping websites. Proc ACM Hum Comput Interact 2019;3:1–32.

[7] Waldman AE. Cognitive biases, dark patterns, and the 'privacy paradox. Curr Opin Psychol 2020;31:105–9.

[8] OECD, 2023. Consumer vulnerability in the digital age. Number 355 in OECD Digital Economy Papers, Paris. URL: doi:10.1787/4d013cc5-en.

[9] European Commission, London Economics, VVA Consulting, Ipsos Mori consortium. Consumer Vulnerability Across Key Markets in the European Union. 2016. https://doi.org/10.2818/056024. Final report. LuxembourgURL: https://commission.europa.eu/system/files/2018-04/consumers-approved-report_en.pdf.

[10] Helberger N, Sax M, Strycharz J, Micklitz HW. Choice architectures in the digital economy: towards a new understanding of digital vulnerability. J Consum Policy

2022;45:175–200. https://doi.org/10.1007/s10603-021-09500-5. URL: https://link.springer.com/10.1007/s10603-021-09500-5.
[11] European Data Protection Board, 2023. Guidelines 03/2022 on deceptive design patterns in social media platform interfaces: how to recognise and avoid them. Version 2.0. URL: https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-032022-deceptive-design-patterns-social-media_en.
[12] Competition and Markets Authority, 2022. Online choice architecture - how digital design can harm competition and consumers - discussion paper. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1066524/Online_choice_architecture_discussion_paper.pdf.
[13] Article 29 Data Protection Working Party, 2017. Guidelines on Data Pro- tection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. URL: https://ec.europa.eu/newsroom/just/document.cfm?doc_id=47711.
[14] Malgieri G, Niklas Jędrzej. Vulnerable data subject. Comput Law Security Rev 2020;37:105415. https://doi.org/10.1016/j.clsr.2020.105415. Doi:.
[15] Bosch C, Erb B, Kargl F, Kopp H, Pfattheicher S. Tales from the dark side: privacy dark strategies and privacy dark patterns. Proc Priv Enhancing Technol 2016;2016:237–54.
[16] Kahneman D. Thinking, fast and slow. MacMillian; 2011.
[17] Acquisti A, Sleeper M, Wang Y, Wilson S, Adjerid I, Balebako R, Brandimarte L, Cranor LF, Komanduri S, Leon PG, Sadeh N, Schaub F. Nudges for privacy and security: understanding and assisting users' choices online. ACM Comput Surv 2017;50:1–41. https://doi.org/10.1145/3054926.
[18] Graßl P, Schraffenberger H, Borgesius FZ, Buijzen M. Dark and bright patterns in cookie consent requests. J Dig Soc Res 2021;3:1–38.
[19] Utz C, Degeling M, Fahl S, Schaub F, Holz T. (un)informed consent: studying GDPR consent notices in the field. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security; 2019. p. 973–90. https://doi.org/10.1145/3319535.3354212.
[20] Simon HA. Rational choice and the structure of the environment. Psychol Rev 1956;63:129–38. https://doi.org/10.1037/h0042769. URL : http://doi.apa.org/getdoi.cfm?doi=10.1037/h0042769.
[21] Hodgkinson GP, Bown NJ, Maule AJ, Glaister KW, Pearman AD. Breaking the frame: an analysis of strategic cognition and decision making under uncertainty. Strateg Manag J 1999;20:977–85.
[22] Nikolić J. Biases in the decision-making process and possibilities of overcoming them. Ekonomski Horizonti, 20; 2018. p. 45–59.
[23] Hertwig R, Grüne-Yanoff T. Nudging and boosting: steering or empowering good decisions. Perspect Psychol Sci 2017;12:973–86. https://doi.org/10.1177/1745691617702496.
[24] Kozyreva, A., Lewandowsky, S., Hertwig, R., 2020. Citizens versus the internet: confronting digital challenges with cognitive tools, 54.
[25] Cox AL, Gould SJ, Cecchinato ME, Iacovides I, Renfree I. Design frictions for mindful interactions: the case for microboundaries. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery; 2016. p. 1389–97. https://doi.org/10.1145/2851581.2892410. URL:10.1145/2851581.2892410.
[26] Moser C, Schoenebeck SY, Resnick P. Impulse buying: design practices and consumer needs. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19. Glasgow, Scotland UK: ACM Press; 2019. p. 1–15. https://doi.org/10.1145/3290605.3300472. URL: http://dl.acm.org/citation.cfm?doid=3290605.3300472.
[27] Bongard-Blanchy K, Rossi A, Rivas S, Doublet S, Koenig V, Lenzini G. I Am Definitely Manipulated, Even When I am Aware of IIt. It's Ridiculous!" - dark patterns from the end-user perspective. In: Designing Interactive Systems Conference 2021, ACM, Virtual Event USA; 2021. p. 763–76. https://doi.org/10.1145/3461778.3462086. URL: https://dl.acm.org/doi/10.1145/3461778.3462086.
[28] Di Geronimo L, Braz L, Fregnan E, Palomba F, Bacchelli A. UI dark patterns and where to find them: a study on mobile applications and user perception. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; 2020. p. 1–14. https://doi.org/10.1145/3313831.3376600. ght/content/10.1108/S2050-206020150000010002/full/html, 10.1108/S2050-206020150000010002.
[29] Matte C, Bielova N, Santos C. Do cookie banners respect my choice?: measuring legal compliance of banners from IAB Europe's tranparency and consent framework. In: 2020 IEEE Symposium on Security and Privacy (SP); 2020. p. 791–809. https://doi.org/10.1109/SP40000.2020.00076.
[30] Article 29 Data Protection Working Party, 2018. Guidelines on transparency under regulation 2016/679, 17/en wp260 rev.01. adopted on 29 November 2017. as last revised and adopted on 11 April 2018. URL: https://ec.europa.eu/newsroom/article29/redirection/document/51025.
[31] Nissenbaum H. A contextual approach to privacy online. Daedalus, 140; 2011. p. 32–48. https://doi.org/10.1162/DAED_a_00113.
[32] Fineman MA. The vulnerable subject: Anchoring equality in the human condition. Yale JL & Feminism 2008;20:1.
[33] Baker SM, Gentry JW, Rittenburg TL. Building understanding of the domain of consumer vulnerability. J Macromarket 2005;25:128–39. https://doi.org/10.1177/0276146705280622.
[34] Yeung K. 'hypernudge': big data as a mode of regulation by design. Inf, Commun Soc 2017;20:118–36.
[35] European Commission, 2021. Commission Notice – Guidance on the interpretation and application of Directive 2005/29/EC of the European Parliament and of the Council concerning unfair business-to-consumer commercial practices in the internal market. URL: https://op.europa.eu/en/publication-detail/-/publication/c608fff7-687a-11ec-9136-01aa75ed71a1/language-en.
[36] ISO, 2023. ISO 31700-1:2023 consumer protection — privacy by design for consumer goods and services — part 1: high-level requirements.
[37] European Commission. Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation: final report. Publications Office of the European Union, LU; 2022. https://doi.org/10.2838/859030. URL: data.eur opa.eu/doi/.
[38] Gunawan J, Santos C, Kamara I. Redress for dark patterns privacy harms? a case study on consent interactions. In: Proceedings of the 2022 Symposium on Computer Science and Law; 2022. p. 181–94.
[39] European Commission, 2022. Digital fairness – fitness check on EU consumer law. URL: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13413-Digital-fairness-fitness-check-on-EU-consumer-law_en.
[40] ICO, 2022. Overview of data protection harms and the ICO taxonomy v1. URL: https://ico.org.uk/media/about-the-ico/douments/4020144/overview-of-data-protection-harms-and-the-ico-taxonomy-v1-202204.pdf.
[41] Citron DK, Solove DJ. Privacy harms. BUL Rev 2022;102:793.
[42] Courmont, A., 2022. Le plaignant type? Un homme, diplômé et cadre. URL: https://linc.cnil.fr/fr/le-plaignant-type-un-homme-diplome-et-cadre.
[43] PenzeyMoog, E., 2021. Design for safety. a book apart.
[44] Kaplan S, Garrick BJ. On the quantitative definition of risk. Risk Anal 1981;1:11–27.
[45] Efroni Z. The Digital Services Act: risk-based regulation of online platforms. Internet Policy Rev 2021. URL: https://policyreview.info/articles/news/digital-services-act-risk-based-regulation-online-platforms/1606.
[46] Chivukula SS, Gray CM, Brier JA. Analyzing value discovery in design decisions through ethicography. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; 2019. p. 1–12. https://doi.org/10.1145/3290605.3300307. URL: https://dl.acm.org/doi/10.1145/3290605.3300307.
[47] Chivukula SS, Brier J, Gray CM. Dark intentions or persuasion? UX designers' activation of stakeholder and user values. In: Proceedings of the 2018 ACM Conference Companion Publication on Designing Interative Systems. Hong Kong China: ACM; 2018. p. 87–91. https://doi.org/10.1145/3197391.3205417. URL: https://dl.acm.org/doi/10.1145/3197391.3205417.
[48] Norman DA. Emotional design: why we love (or hate) everyday things. Civitas Books; 2004.
[49] Burgemeestre B, Hulstijn J, Tan YH. Rule-based versus principle- based regulatory compliance. IOS Press; 2009. p. 37–46.
[50] Calo R. Privacy, vulnerability, and affordance. DePaul L Rev 2016;66:591.
[51] European Data Protection Board, 2019. Guidelines 4/2019 on article 25 data protection by design and by default. URL: https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_201904_dataprotection_by_design_and_by_default.pdf.
[52] Susser D, Roessler B, Nissenbaum HF. Online manipulation: Hid- den influences in a digital world. Geo L Tech Rev 2018;4:1.
[53] Article 29 Data Protection Working Party, 2018. Guidelines on Automated individual decision- making and Profiling for the purposes of Regulation 2016/679 17/EN WP251rev.01.
[54] Luna F. Elucidating the concept of vulnerability: layers not labels. IJFAB: Int J Femin Approaches Bioethics 2009;2:121–39.
[55] Leiser MR, Santos C. Dark patterns, enforcement, and the emerging digital design acquis: manipulation beneath the interface. enforcement, and the emerging digital design acquis: manipulation beneath the interface. Eur J Law Technol 2024;15(1). https://ejlt.org/index.php/ejlt/article/view/990.
[56] Mantelero, A., 2023. Fundamental rights impact assessment in the DSA. Verfassungsblog ed. Berlin. p. 107–119.
[57] Seymour W, Van Kleek M. Exploring interactions between trust, anthropomorphism, and relationship development in voice assistants. In: Proceedings of the ACM on Human-Computer Interaction 5. 371; 2021. p. 1–371. https://doi.org/10.1145/3479515. 16URL:10.1145/3479515.
[58] Lacey C, Caudwell C. Cuteness as a 'dark pattern' in home robots. In: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE; 2019. p. 374–81.
[59] De Conca S. The present looks nothing like the Jetsons - deceptive design in virtual assistants and the protection of the rights of users. Comput Law Security Rev 2023;51:105866. https://doi.org/10.1016/j.clsr.2023.105866. URL:.
[60] Carli R, Calvaresi D. Reinterpreting vulnerability to tackle deception in principles-based XAI for human-computer interaction. International workshop on explainable, transparent autonomous agents and multi- agent systems. Springer; 2023. p. 249–69.
[61] Natale S. Deceitful media: artificial intelligence and social life after the Turing test. USA: Oxford University Press; 2021.
[62] Bermudez JP, Nyrup R, Deterding S, Moradbakhti L, Mougenot C, You F, Calvo RA. What is a subliminal technique? An ethical perspective on ai-driven influence. In: 2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS). West Lafayette, IN, USA: IEEE; 2023. p. 1–10. https://doi.org/10.1109/ETHICS57328.2023.10155039. URL: https://ieeexplore.ieee.org/document/10155039/.
[63] Laux J, Wachter S, Mittelstadt B. Trustworthy artificial intelligence and the European union AI Act: on the conflation of trustworthiness and acceptability of risk. Regul Gov 2023.
[64] Carli R, Najjar A, Calvaresi D. Risk and exposure of XAI in persuasion and argumentation: the case of manipulation. International workshop on explainable,

transparent autonomous agents and multi- agent systems. Springer; 2022. p. 204–20.
[65] Wäfler T, Schmid U. Explainability is not enough: requirements for human-AI-partnership in complex socio-technical systems. In: Proceedings of the 2nd European Conference on the Impact of Artificial Intelligence and Robotics (ECIAIR 2020); 2021. p. 184–94. https://doi.org/10.20378/irb-49775.
[66] Haataja M, Bryson JJ. The European Parliament's AI regulation: should we call it progress? Series 2, 4. Amicus Curiae; 2022. p. 707.
[67] Ebers, M., Hoch, V.R., Rosenkranz, F., Ruschemeier, H., Steinrotter, B., 2021. The European Commission's proposal for an Artificial Intelligence Act—a critical assessment by members of the robotics and ai law society (rails). J 4, 589–603.
[68] Smuha, N.A., Ahmed-Rengers, E., Harkens, A., Li, W., MacLaren, J., Piselli, R., Yeung, K., 2021. How the EU can achieve legally trustworthy AI: a response to the European Commission's proposal for an Artificial Intelligence Act. URL: https://ssrn.com/abstract=3899991 or doi:10.2139/ssrn.3899991.
[69] Stahl BC, Rodrigues R, Santiago N, Macnish K. A European agency for artificial intelligence: protecting fundamental rights and ethical values. Comput Law Security Rev 2022;45:105661.
[70] Zanca F, Brusasco C, Pesapane F, Kwade Z, Beckers R, Avanzo M. Regulatory aspects of the use of artificial intelligence medical software. Seminars in radiation oncology. Elsevier; 2022. p. 432–41.
[71] Gibson JJ. The theory of affordances. Lawrence Erlbaum Associates; 1986. p. 127–37.
[72] Helsper EJ. Digital world: from divides to socio-digital inequalities. The digital disconnect. The social causes and consequences of digital inequalities. SAGE Publications; 2021. p. 27–45.
[73] Norman DA. The design of everyday things. Revised and expanded edition ed. New York, New York: Basic Books; 2013.
[74] Bronfenbrenner U. The ecology of human development: experiments by nature and design. Harvard University Press; 1979.
[75] Bronfenbrenner U, Morris PA. The bioecological model of human development. Handbook of child psychology. Wiley; 2007. p. 793–828.
[76] Johnson GM, Puplampu KP. Internet use during childhood and the ecological techno-subsystem. Canadian J Learn Technol /La revue canadienne de l'apprentissage et de la technologie 2008;34. https://doi.org/10.21432/T2CP4T. URL: http://www.cjlt.ca/index.php/cjlt/article/view/26428.
[77] O'Neill B. Ecological perspectives and children's use of the internet: exploring micro to macro level analysis. Eesti Haridusteaduste Ajakiri Estonian J Educ 2015;3:32–53. https://doi.org/10.12697/eha.2015.3.2.02b.
[78] Murnane EL, Walker TG, Tench B, Voida S, Snyder J. Personal informatics in interpersonal contexts: towards the design of technology that supports the social ecologies of long-term mental health management. Proc ACM Hum Comput Interact 2018;2(CSCW):1–27. https://doi.org/10.1145/3274396.
[79] Costa Figueiredo M, Chen Y. Health data in fertility care: an ecological perspective. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; 2021. p. 1–17. https://doi.org/10.1145/3411764.3445189.
[80] Malgieri G. Who is the vulnerable individual?. In: Vulnerability and Data Protection Law. 1st edition. Oxford: Oxford University Press; 2023.. https://academic.oup.com/book/46055/chapter/404533553.
[81] Sánchez Chamorro L, Lallemand C, Gray, Colin M. My mother told me these things are always fake" — understanding teenagers' experiences with manipulative designs. In: Designing Interactive Systems Conference (DIS '24), July 1–5, 2024. Denmark: IT University of Copenhagen; 2024. p. 13. https://doi.org/10.1145/3643834.3660704. ACM, New York, NY, USApages.
[82] Mantelero A. The future of consumer data protection in the E.U. Rethinking the "notice and consent" paradigm in the new era of predictive analytics. Comput Law Security Rev 2014;30:643–60. https://doi.org/10.1016/j.clsr.2014.09.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S026736491400154X.
[83] Lessig Lawrence. Code and other laws of cyberspace. New York, NY United States: Basic Books, Inc.Division of HarperCollins 10 E. 53rd St; 1999.
[84] Lockton D. Persuasive technology and digital design for behaviour change. SSRN Electron J 2012. https://doi.org/10.2139/ssrn.2125957. URL, http://www.ssrn.com/abstract=2125957.
[85] Botes M. Autonomy and the social dilemma of online manipulative behavior. AI Ethics 2022. https://doi.org/10.1007/s43681-022-00157-5. URL: link.springer.com/10.1007/s43681-022-00157-5.
[86] Gray CM, Chivukula SS. Ethical mediation in UX practice. In: Proceedings of the 2019 CHI conference on human factors in computing systems; 2019. p. 1–11. https://doi.org/10.1145/3290605.3300408. URL: https://dl.acm.org/doi/10.1145/3290605.3300408.
[87] Wong RY. Tactics of soft resistance in user experience professionals' values work. Proc ACM Hum Comput Interact 2021;5:1–28. https://doi.org/10.1145/3479499.
[88] Constanza-Chock S. Design values: hard-coding liberation? In *design justice*. Community-led practices to build the worlds we need. The MIT Press; 2020. p. 48–89.
[89] Käufer S, Chemero A. Phenomenology: an introduction. Wiley; 2021. URL: https://books.google.lu/books?id=qR_azQEACAAJ.
[90] Hicks JM, Still JD. Examining the effects of clutter and target salience in an e-commerce visual search task. In: Proceedings of the human factors and ergonomics society 2019 annual meeting; 2019. p. 1761–5.
[91] Armel KC, Beaumel A, Rangel A. Biasing simple choices by manipulating relative visual attention. Judgm Decis Mak 2008;3:396–403.
[92] Still J, Still M. Influence of visual salience on webpage product searches. ACM Trans Appl Percept (TAP) 2019;16:1–11.
[93] Milosavljevic M, Navalpakkam V, Koch C, Rangel A. Relative visual saliency differences induce sizable bias in consumer choice. J Consum Psychol 2012;22:67–74.
[94] Towal RB, Milosavljevic M, Koch C. The effect of visual salience on multiple-alternative, value-based decisions. J Vis 2012;12:167. –167.
[95] Veas EE, Mendez E, Feiner SK, Schmalstieg D. Directing attention and influencing memory with visual saliency modulation. In: Proceedings of the SIGCHI conference on human factors in computing systems; 2011. p. 1471–80.
[96] Jarvenpaa SL. Graphic displays in decision making—the visual salience effect. J Behav Decis Mak 1990;3:247–62.
[97] Jarovsky L. Dark patterns in personal data collection: definition, taxonomy and lawfulness. SSRN Electron J 2022:048582. https://doi.org/10.2139/ssrn.4048582. URL, https://papers.ssrn.com/abstract=4.
[98] Gray CM, Sanchez Chamorro L, Obi I, Duane JN. Mapping the landscape of dark patterns scholarship: a systematic literature review. In: Designing interactive systems conference (DIS companion '23); 2023. https://doi.org/10.1145/3563703.3596635.
[99] Nouwens M, Liccardi I, Veale M, Karger D, Kagal L. Dark patterns after the GDPR: scraping consent pop-ups and demonstrating their influence. In: CHI Conference on Human Factors in Computing Systems; 2020. p. 13. https://doi.org/10.1145/3313831.3376321. URL: http://arxiv.org/abs/2001.02479. arXiv: 2001.02479.
[100] Luguri J, Strahilevitz LJ. Shining a light on dark patterns. J Legal Anal 2021;13:43–109. https://doi.org/10.1093/jla/laaa006.
[101] Berens BM, Dietmann H, Krisam C, Kulyk O, Volkamer M. Cookie disclaimers: impact of design and users' attitude. In: Proceedings of the 17th International Conference on Availability, Reliability and Security. Vienna Austria: ACM; 2022. p. 1–20. https://doi.org/10.1145/3538969.3539008. URL: dl.acm.org/doi/10.1145/3538969.3539008.
[102] Gray CM, Chen J, Chivukula SS, Qu L. End user accounts of dark patterns as felt manipulation. In: Proceedings of the ACM in human-computer interaction. 5; 2021. p. 26. https://doi.org/10.1145/3479516. URL:.
[103] Maier M, Harr R. Dark design patterns: an end-user perspective. Hum Technol 2020;16:170–99. https://doi.org/10.17011/ht/urn.202008245641.
[104] Borberg I, Hougaard R, Rafnsson W, Kulyk O. So I Sold My Soul": effects of dark patterns in cookie notices on end-user behavior and perceptions. In: Workshop on Usable Security and Privacy (USEC); 2022. p. 11.
[105] Helsper EJ. Survey on the use of information and communication technologies in Brazilian households: ICT households 2015. São Paulo, Brasil: Núcleo de Informação e Coordenação do Ponto BR; 2016.
[106] Scheerder A, van Deursen A, van Dijk J. Taking advantage of the internet: a qualitative analysis to explain why educational background is decisive in gaining positive outcomes. Poetics 2020;80:101426. https://doi.org/10.1016/j.poetic.2019.101426.
[107] Van Deursen AJAM, Helsper EJ. The third-level digital divide: who benefits most from being online?, 10. Emerald Group Publishing Limited; 2015. p. 29–52. URL: https://www.emerald.com/insi.
[108] Van Dijk J, van Deursen AJAM. Digital skills. New York: Palgrave Macmillan US; 2014. https://doi.org/10.1057/9781137437037. URL: http://link.springer.com/10.1057/9781137437037.
[109] Eynon R, Geniets A. The digital skills paradox: how do digitally excluded youth develop skills to use the internet? Learn Media Technol 2016;41:463–79. https://doi.org/10.1080/17439884.2014.1002845.
[110] Freese J, Rivas S, Hargittai E. Cognitive ability and internet use among older adults. Poetics 2006;34:236–49. https://doi.org/10.1016/j.poetic.2006. 05.008.
[111] Hargittai E, Hinnant A. Digital inequality: Differences in young adults' use of the internet. Commun Res 2008;35:602–21. https://doi.org/10.1177/0093650208321782.
[112] Helsper EJ, van Deursen AJAM. Do the rich get digitally richer? quantity and quality of support for digital engagement. Inf, Commun Soc 2017;20:700–14. https://doi.org/10.1080/1369118X.2016.1203454.
[113] Van Dijk J, Hacker K. The digital divide as a complex and dynamic phenomenon. Inf Soc 2003;19:315–26. https://doi.org/10.1080/01972240309487.
[114] Zac, A., Huang, Y.-C., von Moltke, A., Decker, C., Ezrachi, A. Dark patterns and consumer vulnerability (August 22, 2023). Available at SSRN: https://ssrn.com/abstract=4547964.
[115] DiPaola D, Calo R. Socio-digital vulnerability (SSRN Scholarly Paper 4686874). doi:10.2139/ssrn.4686874.
[116] Zingales, L., Morton, F.S., Rolnik, G., 2019. Stigler committee on digital platforms: final report . URL: https://www.chicagobooth.edu/-/media/research/stigler/pdfs/digital-platforms–committee-report–stigler-center.pdf.
[117] Tsetsi E, Rains SA. Smartphone internet access and use: extending the digital divide and usage gap. Mob Media Commun 2017;5:239–55. https://doi.org/10.1177/2050157917708329.
[118] Gunawan J, Pradeep A, Choffnes D, Hartzog W, Wilson C. A comparative study of dark patterns across web and mobile modalities. Proc ACM Hum Comput Interact 2021;5:1–29. https://doi.org/10.1145/3479521.
[119] Lee CS, McKenzie K. Socioeconomic and geographic inequalities of internet addiction in Korean adolescents. Psychiatry Investig 2015;12:559–62. https://doi.org/10.4306/pi.2015.12.4.559.
[120] Radesky J, Hiniker A, McLaren C, Akgun E, Schaller A, Weeks HM, Campbell S, Gearhardt AN. Prevalence and characteristics of manipulative design in mobile applications used by children. JAMA Netw Open 2022;5:e2217641. –e2217641.
[121] Kampanos G, Shahandashti SF. Accept all: the landscape of cookie banners in Greece and the UK. In: Jøsang A, Futcher L, Hagen J, editors. ICT systems security and privacy protection. Cham: Springer International Publishing; 2021. p. 213–27. https://doi.org/10.1007/978-3-030-78120-0_14.

[122] Santos C, Rossi A, Sanchez Chamorro L, Bongard-Blanchy K, Abu- Salma R. Cookie banners, what's the purpose? Analyzing cookie banner text through a legal lens. In: Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society. ACM; 2021. p. 187–94. https://doi.org/10.1145/3463676.3485611. Virtual Event Republic of KoreaURL: https://dl.acm.org/doi/10.1145/3463676.3485611.
[123] Bermeitinger C. Priming. Psychology and mental health: concepts, methodologies, tools, and applications. IGI Global; 2016. p. 42–88.
[124] Hsu N, Schütt Z. Psychology of priming. Nova Science Publishers; 2012.
[125] Breitmeyer BG, Ro T, Singhal NS. Unconscious color priming occurs at stimulus-not percept-dependent levels of processing. Psychol Sci 2004;15:198–202.
[126] Chartrand TL, van Baaren RB, Bargh JA. Linking automatic evaluation to mood and information processing style: consequences for experienced affect, impression formation, and stereotyping. J Exp Psychol: General 2006;135:70.
[127] Francken JC, van Gaal S, de Lange FP. Immediate and long-term priming effects are independent of prime awareness. Conscious Cogn 2011;20:1793–800.
[128] Yeun Chun K, Hee Song J, Hollenbeck CR, Lee JH. Are contextual advertisements effective? the moderating role of complexity in banner advertising. Int J Advert 2014;33:351–71.
[129] Yi Y. The influence of contextual priming on advertising effects. Adv Consume Res 1991;18:1.
[130] Yoo CY. Unconscious processing of web advertising: effects on implicit memory, attitude toward the brand, and consideration set. J Interact Market 2008;22:2–18.
[131] Wänke M. Primes as hidden persuaders. Curr Opin Psychol 2016;12:63–6.
[132] Petticrew M, Maani N, Pettigrew L, Rutter H, Van Schalkwyk MC. Dark nudges and sludge in big alcohol: behavioral economics, cognitive biases, and alcohol industry corporate social responsibility. Milbank Q 2020;98:1290–328.
[133] Costello FJ, Yun J, Lee KC. Digital dark nudge: an exploration of when digital nudges unethically depart. In: Proceedings of the 55th Hawaii International Conference on System Sciences; 2022.
[134] Dennis AR, Yuan L, Feng X, Webb E, Hsieh CJ. Digital nudging: numeric and semantic priming in e-commerce. J Manage Inf Syst 2020;37:39–65.
[135] Jiang X, Jiang Y, Parasuraman R. The visual priming of motion- defined 3d objects. PLoS One 2015;10:e0144730.
[136] Kareklas I, Muehling DD, King S. The effect of color and self-view priming in persuasive communications. J Bus Res 2019;98:33–49.
[137] Elgendi M, Kumar P, Barbic S, Howard N, Abbott D, Cichocki A. Subliminal priming—state-of-the-art and future perspectives. Behav Sci 2018;8:54.
[138] Matthews G, Hancock PA. The handbook of operator fatigue. CRC Press; 2017.
[139] Crump C, Walenchok S, Johnson C, Pauszek J, Young D. Stress, operator error, and transportation accidents. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Los Angeles, CA: SAGE Publications Sage CA; 2021. p. 1453–8.
[140] Hockey GRJ. Operator functional state: the assessment and pre- diction of human performance degradation in complex tasks, 355. IOS Press; 2003.
[141] Kahol K, Leyba MJ, Deka M, Deka V, Mayes S, Smith M, Fer- rara JJ, Panchanathan S. Effect of fatigue on psychomotor and cognitive skills. Am J Surg 2008;195:195–204.
[142] Ricciardi O, Maggi P, Nocera FD. Boredom makes me nervous: fidgeting as a strategy for contrasting the lack of variety. Int J Hum Factors Ergon 2019;6:195–207.
[143] Chaudhary A, Saroha J, Monteiro K, Forbes AG, Parnami A. Are You Still Qatching?": exploring unintended user behaviors and dark patterns on video streaming platforms. In: Designing Interactive Systems Conference, ACM, Virtual Event Australia; 2022. p. 776–91. https://doi.org/10.1145/3532106.3533562. URL: https://dl.acm.org/doi/10.1145/3532106.3533562.
[144] Zagal JP, Björk S, Lewis C. Dark patterns in the design of games. Foundations of digital games, 2013; 2013.
[145] Amirpur M, Benlian A. Buying under pressure: purchase pressure cues and their effects on online buying decisions. In: Proceedings of the Thirty Sixth International Conference on Information Systems; 2015. p. 1–18.
[146] Bielova, N., Litvine, L., Nguyen, A., Chammat, M., Toubiana, V., & Hary, E. (2024). The effect of design patterns on (Present and Future) cookie consent decisions. Proceedings of the 33rd USENIX Security Symposium, 1–9.
[147] Avolicino, S., Di Gregorio, M., Palomba, F., Romano, M., Sebillo, M., & Vitiello, G. (2022). AI-based emotion recognition to study users' perception of dark patterns. In M. Kurosu, S. Yamamoto, H. Mori, M. M. Soares, E. Rosenzweig, A. Marcus, P.-L. P. Rau, D. Harris, & W.-C. Li (Eds.), *HCI International 2022—Late Breaking Papers. Design, User Experience and Interaction* (Vol. 13516, pp. 185–203). Springer International Publishing. doi:10.1007/978-3-031-17615-9_13.
[148] van Nimwegen, C., & de Wit, J. (2022). Shopping in the dark: effects of platform choice on dark pattern recognition. In M. Kurosu (Ed.), *Human-computer interaction. User experience and behavior* (Vol. 13304, pp. 462–475). Springer International Publishing. doi:10.1007/978-3-031-05412-9_32.
[149] Sánchez Chamorro L, Toebosch R, Lallemand C. Manipulative design and older adults: co-creating magic machines to understand experiences of online manipulation. In: Designing Interactive Systems Conference (DIS '24), July 1–5, 2024. Denmark: IT University of Copenhagen; 2024. p. 17. https://doi.org/10.1145/3643834.3661513. ACM, New York, NY, USApages.
[150] Meyer M, Adkins V, Yuan N, Weeks HM, Chang YJ, Radesky J. Advertising in young children's apps: a content analysis. J Dev Behav Pediatr 2019;40:32–9.
[151] Forbrukerradet, 2022. Insert coin. How the gaming industry exploits consumers using loot boxes. URL: https://storage.forbrukerradet.no/media/2022/05/2022-05-31-insert-coin-publish.pdf.
[152] Clarke, J.M., Mehrnezhad, M., Toreini, E., 2023. Invisible, unreadable, and inaudible cookie notices: an evaluation of cookie notices for users with visual impairments URL: http://arxiv.org/abs/2308.11643, 10.48550/arXiv.2308.11643. arXiv:2308.11643 [cs].
[153] Sannon S, Forte A. Privacy research with marginalized groups: what we know, what's needed, and what's next. In: Proceedings of the ACM on Human-Computer Interaction. 6; 2022. p. 1–33. https://doi.org/10.1145/3555556.
[154] Holkar M, Lees C. A safer bet? Online gambling and mental health. 2020. London, UK. URL, https://www.moneyandmentalhealth.org/wp-content/uploads/2020/07/A_Safer_Bet.pdf.pdf.
[155] Xia B, Lu Q, Perera H, Zhu L, Xing Z, Liu Y, Whittle J. Towards concrete and connected AI risk assessment (c2aira): a systematic mapping study. In: 2023 IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN); 2023. p. 104–16. https://doi.org/10.1109/CAIN58948.2023.00027.
[156] Beimborn M, Kadi S, Köberer N, Mühleck M, Spindler M. Focusing on the human: interdisciplinary. Aging and Technology. Transcript Verlag; 2016.
[157] O'Brien JE, Brewer KB, Jones LM, Corkhum J, Rizo CF. Rigor and respect: recruitment strategies for engaging vulnerable populations in research. J Interpers Violence 2022;37:NP17052–72. https://doi.org/10.1177/08862605211023497.
[158] Slupska J, Tanczer LM. Threat modeling intimate partner violence: tech abuse as a cybersecurity challenge in the Internet of Things. Emerald Publishing Limited; 2021. p. 663–88. https://doi.org/10.1108/978-1-83982-848-520211049. Emerald Studies In Digital Crime, Technol- ogy and Social HarmsURL: doi.org/10.1108/97 8-1-83982-848-520211049.
[159] Siciliani P, Riefa C, Gamper H. Consumer theories of harm: an economic approach to consumer law enforcement and policy making. Bloomsbury Publishing; 2019.
[160] BEUC, 2022. "Dark patterns" and the EU consumer law. Recommendations for better enforcement and reform. URL: https://www.beuc.eu/sites/default/files/publications/beuc-x-2022-013_dark_patters_paper.pdf.
[161] Hasan Mansur SM, Salma S, Awofisayo D, Moran K. Aidui: toward automated recognition of dark patterns in user interfaces. In: 2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE); 2023. p. 1958–70. https://doi.org/10.1109/ICSE48619.2023.00166.
[162] Kirkman D, Vaniea K, Woods DW. Darkdialogs: automated detection of 10 dark patterns on cookie dialogs. In: 8th IEEE European Symposium on Security and Privacy. IEEE; 2023.
[163] Rossi A, Ducato R, Haapio H, Passera S. When design met law: design patterns for information transparency. Droit de la Consommation= Consumenterecht: DCCR. 2019. p. 79–121.
[164] CNIL, n.d.. Design patterns données & design par LINC. URL: https://design.cnil.fr/en/design-patterns/.
[165] Holmes K. The cycle of exclusion. MIT Press; 2020. p. 27–40.
[166] Mendez C, Letaw L, Burnett M, Stumpf S, Sarma A, Hilderbrand C. From gendermag to inclusivemag: an inclusive design meta-method. In: 2019 IEEE Symposium on Visual Languages and Human- Centric Computing (VL/HCC). Memphis, TN, USA: IEEE; 2019. p. 97–106. https://doi.org/10.1109/VLHCC.2019.8818889. URL: https://ieeexplore.ieee.org/document/8818889/.
[167] Chivukula SS, Watkins CR, Manocha R, Chen J, Gray CM. Dimensions of ux practice that shape ethical awareness. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Honolulu HI, USA: ACM; 2020. p. 1–13. https://doi.org/10.1145/3313831.3376459. URL: https://dl.acm.org/doi/10.1145/3313831.3376459.