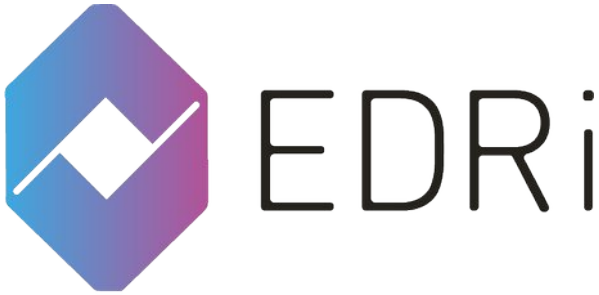

European Digital Rights

A Rights-Based Digital Fairness Act

Background Paper on Protecting People from Manipulative Design and Data Exploitation

October 2025

Lead authors: Dr. Itxaso Domínguez de Olazábal and Jan Penfrat
Review: Ella Jakubowska

European Digital Rights (EDRi) is sincerely grateful to our members and partners who have contributed to this exploratory paper, notably Bits of Freedom, Electronic Frontier Foundation (EFF), Privacy International, SUPERRR Lab, Vrijschrift.org, and Jürgen Bering and Simone Ruf, who supported with the research, conceptualisation, drafting, and review of this paper.

We are also grateful for the generous insights and critical feedback shared by Dr. Gianclaudio Malgieri, Dr. Lorena Sáncez-Chamorro, Dr. Constanta Rosca and the whole ReSocial project, Dr. Monika Namylowska, Dr. Natali Helberger, Dr. Cristiana Santos, Dr. Lex Zard, and Dr. Mark Leiser, whose thinking and contributions have improved the depth and rigour of this research paper.

Acronyms

BDIA	Behavioural Design Impact Assessment
CPC	Consumer Protection Cooperation
CRD	Consumer Rights Directive
CCPA	California Consumer Privacy Act
CMPs	Consent Management Platforms
DFA	Digital Fairness Act
DMA	Digital Markets Act
DSA	Digital Services Act
EC	European Commission
GDPR	General Data Protection Regulation
GPC	Global Privacy Control
LAM	Large Action Model
RAD	Representative Actions Directive
UCPD	Unfair Commercial Practices Directive
UCTD	Unfair Commercial Terms Directive
VLOP	Very Large Online Platform
VLOSE	Very Large Online Search Engine

Table of Contents

I. Executive Summary and Policy Recommendations

Digital services today are increasingly designed to manipulate rather than empower. **Exploitative patterns are built into the interface, the data infrastructure, and the business model, not always to deceive directly, but to capture attention, suppress disengagement, and steer behaviour at scale.** These dynamics do not only distort markets: they undermine people's dignity, autonomy, and equality, and corrode the conditions under which fundamental rights can be exercised meaningfully online. They also generate collective effects, weakening trust in digital environments and limiting democratic participation. Yet **current EU consumer protection law lacks the tools to address these systemic interferences with rights and freedoms**.

The DFA must take a systemic approach, addressing not only isolated practices but the embedded logic of manipulation in today's digital economy. It must recognise manipulative design as a structural rights issue, not a marginal consumer inconvenience. Specifically, it must recognise that:

- Digital environments structurally vulnerabilise all users, though not all to the same degree;
- Manipulation is often systemic, opaque, and continuous, not one-off or visible;
- Design is not neutral: it is a mechanism of power;
- And fairness is not a matter of efficiency, but of justice. It must be embedded by design and by default.

Manipulation by Design: A Three-Part Framework

Today's digital environments are structurally designed to manipulate. The DFA must confront these threats to fundamental rights as a systemic pattern, not a series of isolated tricks. We identify three interrelated categories of manipulative practices that distort autonomy and fairness online:

- **Addictive design:** feedback loops and interface features that **impair people's ability to disengage or exercise time-bound control** over their use of digital services.
- **Deceptive design:** design choices that **obstruct or distort consent and decision-making, turning rights into empty formalities**.
- **Unfair personalisation:** profiling-based systems that **distort decision-making or result in discriminatory or exclusionary outcomes, restricting equal access to information, services, and opportunitie**s, and subordinating people's rights to business optimisation goals.

These practices are not neutral. They are embedded in business models that prioritise data extraction, attention maximisation, and behavioural steering over people's agency. They interfere with the freedom to make uncoerced choices, the right to non-discrimination, and the ability to participate in democratic and social life without structural manipulation. The DFA must treat them as manipulative by design, and regulate them as such.

Core Legal and Policy Recommendations

1. Adopt a Hybrid Legislative Architecture

To ensure consistency and enforceability, the DFA should be adopted as a Regulation with a hybrid structure, combining self-standing, directly applicable provisions with targeted amendments to the UCPD, UCTD, and CRD. This design would allow the Regulation to introduce new, rights-based duties such as fairness-by-design and non-profiling-by-default, while aligning existing Directives with the realities of digital markets. It would also ensure a coherent enforcement architecture in which the European Commission plays a direct role, as it does under the DSA, avoiding fragmentation and uneven transposition across Member States.

2. Modernise Legal Concepts to Align Them with a Rights-Based Approach

Update definitions of consumer, vulnerable consumer, and trader to reflect the digital context:

- Consumers include all users subject to profiling or personalised interfaces, regardless of payment;
- Vulnerability should be understood as systemically induced. It does not reside in fixed individual traits but emerges from the interaction between user characteristics, situational conditions, and exploitative design environments. This includes both structural and situational factors, and requires regulators to focus on how digital systems actively produce and exacerbate vulnerability rather than treating it as pre-existing or exceptional. ;
- Traders include all entities involved in providing, deploying, or determining the operation of digital services, such as developers of recommender systems, providers of AI infrastructure, or intermediaries whose design or optimisation choices shape user interaction or decision-making, regardless of any direct contractual relationship with the end user.

3. Introduce a Structural Fairness Duty

- **Recognise fairness not only as a procedural standard but a structural obligation** rooted in fundamental rights, not as a procedural safeguard that traders can circumvent with formal compliance.
- Embed **fairness by design and by default**, especially for:
  - Interface architecture;
  - Consent and personalisation flows;
  - Profiling-based interaction models.
- **Shift the burden from individual users to traders**.
- **Reverse the burden of proof:** where behavioural profiling, personalisation, or opaque interface optimisation is involved, traders must demonstrate that systems respect user autonomy, do not exploit vulnerabilities, and comply with fairness obligations.

- **Codify interface fairness as an enforceable standard**: Traders should be subject to a general duty to design for fairness, with reference to published guidelines and black/grey list prohibitions.
- **Mandate proactive documentation**: Traders must maintain auditable records of system objectives, design decisions, and known or anticipated effects on people, and conduct Behavioural Design Impact Assessments (BDIAs) where a presumption of unfairness applies, to enable regulatory oversight and preventive enforcement.

4. Address Addictive Design

- Recognise addictive design as **structurally manipulative and presumptively unfair** when:
  - ➢ It removes disengagement cues;
  - ➢ It uses emotional profiling to trigger compulsive feedback loops;
  - ➢ It interferes with user-set boundaries.
- Introduce a **grey list for design strategies that may be unfair unless proven otherwise**.
- Enshrine a **right not to be disturbed by default**: push notifications, autoplay, streaks, and profiling-based suggestions should be opt-in only.

5. Tackle Deceptive Design Patterns

- **Prohibit interface strategies that render consent meaningless in practice**.
- **Mandate fair consent design**, support machine-readable consent browser signals and hold CMP providers accountable for manipulative consent templates deployed across services.
- **Expand the consumer law blacklist to include**:
  - ➢ Misleading or coercive consent interfaces;
  - ➢ Hidden opt-outs, countdown timers, misleading visuals;
  - ➢ Emotional framing or language manipulation used to steer user decisions.
- Codify **interface fairness as a general principle** with measurable standards.
- Require **impact assessments for systems that materially influence user decision-making**.

6. Regulate Unfair Personalisation

- Treat profiling-based personalisation as unfair when it:
  - ➢ Exploits cognitive or emotional vulnerabilities;
  - ➢ Affects visibility of options, pricing, or content without transparency;
  - ➢ Relies on inferred traits users cannot verify or contest.
- Ban personalisation based on distress, addiction, or marginalised status; as well as delegated decision-making based on inferred emotional states.
- Systems that adapt environments based on profiling must be presumed unfair where they compromise equality, autonomy, or transparency, regardless of whether formal consent has been obtained.

- Introduce a right to access a non-personalised version of any service where personalisation is not technically necessary.
- Require purpose limitation and transparency for the optimisation logic of recommender systems.

7. Strengthen Enforcement

- **Empower regulators with meaningful access**: Authorities must be able to access internal optimisation data, A/B testing results, and behavioural analytics to detect manipulative practices at scale.
- **Adopt pattern-based detection frameworks**: Regulators should be equipped to recognise manipulation even in the absence of deception, using indicators such as friction asymmetry, emotional steering, or structural nudging.
- **Facilitate inter-authority cooperation**: Strengthen links between consumer protection bodies, data protection authorities, competition authorities, media regulators, and potentially others, to tackle manipulative ecosystems across legal domains.
- **Strengthen deterrence and remedy**: Introduce penalties that reflect the scale of systemic manipulation, and remedies that include design reconfiguration, not just fines.

Too many people in the EU are faced with a digital playing field that is stacked against them. So, naturally, **any DFA ought to have as its policy goal to level the playing field for everyone, while recognising that some communities are more exposed and less protected due to how digital systems intersect with existing inequalities**. This also would stop a secondary harm, namely that of the economic interests of legitimate economic actors that refrain from using dark patterns and other deceptive and/or subliminal design choices in their digital services and products.

All in all, **the DFA must be more than a patch to current consumer law: it must be a foundational instrument for safeguarding fundamental rights in a digital economy built on surveillance and behavioural control**. By adopting structural obligations, reversing the burden of proof, and tackling manipulative design at its root, the DFA can help restore dignity, accessibility, and agency to people across the EU, while also protecting rights-respecting businesses from unfair competitive pressure.

II. Introduction

This background paper has been prepared by European Digital Rights (EDRi), a network of civil society organisations working to defend and advance human rights in the digital environment. Drawing on decades of collective experience in privacy, data protection, commercial and state surveillance, and platform accountability, EDRi has developed this analysis in response to the European Commission's (EC) plans for a Digital Fairness Act (DFA), in particular the Commission's eed for evidence and analysis as part of its July 2025 consultation into the future DFA., It precedes a future network position paper which will establish a common position of the EDRi network.

The EC has rightly recognised that **existing EU law is not sufficiently equipped to address the structurally manipulative practices that define today's digital economy**[1]. Fragmented enforcement, legal uncertainty, and outdated legal concepts have created serious gaps in the field of digital consumer protection, particularly in relation to practices such as manipulative interface design, attention-maximising features, and exploitative personalisation[2]. The Commission's 2024 Fitness Check confirmed what civil society has long warned **digital environments systematically distort people's ability to make meaningful choices about what they can see or do online through structurally manipulative practices, while existing law struggles to respond.** The Fitness Check Report confirms that while the Unfair Commercial Practices Directive (UCPD), the Unfair Contract Terms Directive (UCTD), and the Consumer Rights Directive (CRD) remain broadly coherent and relevant, they are not effective at ensuring the protection of people's fundamental rights across digital markets[3].

While consumer law remains relevant, it is grounded in a transactional model that does not reflect the continuous, pervasive ways in which people's digital environments are shaped. For example, recommender systems that endlessly

1 European Commission, Study to support the fitness check of EU consumer law on digital fairness and report on the application of the Modernisation Directive, 4 October 2024. https://commission.europa.eu/publications/study-support-fitness-check-eu-consumer-law-digital-fairness-and-report-application-modernisation_en (hereinafter, the 'Fitness Check Report').

2 While the more precise term might be digital choice architecture, this paper uses terms like 'platform', 'interface', and 'digital service' for clarity and accessibility. The intention is not to confine regulation to specific actors or front-end elements, but to make clear that consumer law must apply across the entire design and optimisation environment. In digital contexts, every aspect of a service, from interface layout to backend personalisation, can operate as part of a persuasive system, shaping user behaviour through defaults, friction, and dynamic adaptation, rather than through transparent information or genuine choice.

3 While this paper operates within the framework of consumer law, we have opted to use the term user rather than consumer throughout. This choice reflects the reality that people engage with digital services in multiple roles that go beyond transactional consumption. The term consumer traditionally refers to individuals acting for personal (non-professional) purposes in a purchase context, whereas user better captures the broader range of interactions, dependencies, and vulnerabilities that characterise people's engagement with digital environments, including when no economic transaction occurs. More broadly, we strive to recognise the impacts of digital systems on people as a whole, not only on individual users, as the fundamental rights at stake are often structural, collective, and social in nature.

autoplay content are not a single transaction, but an ongoing design choice that undermines people's autonomy and right to data protection; dark patterns in consent flows compromise the right to privacy, data protection and free choice; and algorithmic targeting based on sensitive traits can reinforce discrimination, limiting equal participation online. These problems are not marginal: they define how digital interfaces are currently built, optimised, and monetised across sectors, with profound consequences for fundamental rights, equality, and democratic participation.

The DFA represents a critical opportunity to address these long-standing and evolving challenges by putting people's digital rights at the heart of consumer law to ensure fair, equitable and just treatment for everyone. To do so effectively, it must be framed as a Regulation with a dual function. On the one hand, it must stand on its own as a modern legal instrument that lays down directly applicable rights and obligations in relation to unfair digital practices. On the other hand, it should also serve as the legislative vehicle to update key parts of the existing consumer *acquis*, namely the UCPD, the UCTD, and the CRD.

Without a hybrid approach, the DFA would either lack direct effect and rely too heavily on fragmented national transposition, or it would fail to modernise outdated concepts in the core consumer directives. By clearly establishing this structure from the outset, the DFA can both embed new rules on manipulative and unfair design and bring the existing directives into line with the realities of a data-driven digital economy.

The Commission's report also found widespread enforcement gaps, regulatory fragmentation, and a chilling effect on public and private enforcement due to legal uncertainty and high evidentiary burdens, especially in cross-border and technologically complex cases. Some have argued that no new regulation is needed, only better enforcement. But this framing obscures how today's commercial practices exploit legal grey zones and enforcement delays. The reality is that **the manipulation of user behaviour through interface design, profiling, and personalisation has evolved into systems that govern what people see, how they can act, and under what conditions, without being fully captured by existing law**. Clear legal red lines and systemic oversight are needed not because current rules are irrelevant, but because they were never designed to govern dynamic, data-driven persuasion architectures.

To be effective, the DFA must go beyond superficial fixes. It should establish enforceable rules that address the root causes of digital unfairness: asymmetries of knowledge and power, systems of opaque personalisation, and business models built on emotional and cognitive exploitation. **Crucially, this means embedding systemic accountability, design fairness, and structural oversight into the law.**

While the EU's General Data Protection Regulation (GDPR), Digital Services Act (DSA), and ePrivacy Directive offer partial safeguards, they leave important regulatory gaps. The GDPR's focus on individual consent and personal data does not fully capture collective and systemic dynamics of manipulation that undermine autonomy and equality. The DSA imposes due diligence obligations on large platforms, but does not address how design and profiling are used to distort choice and undermine agency. It is essential to avoid any carve-out for online platforms. Platforms are part of the broader

category of digital services, and limiting the DFA to them would leave significant gaps. For instance, video games, dating apps, and immersive environments can all deploy manipulative design and exploit vulnerabilities just as much as platforms covered by the DSA. The DFA must ensure that all digital services are bound by the same fairness principles, irrespective of sector. And voluntary codes of conduct have proven inadequate to secure environments where rights and autonomy are effectively protected, such as against manipulative design.

Moreover, sectors like video gaming have long operated under self-regulatory models that sidestep core human rights protections. Practices like in-game currencies and emotionally immersive monetisation loops thrive due to under-enforcement and regulatory ambiguity. **The DFA must ensure no digital business model is treated as exceptional when it comes to fairness and user agency.**

Traditional consumer law was built on the principle of providing people with freedom of choice. It was never intended to create new rights or intervene in systemic environments where choice is structurally constrained. The DFA must therefore move beyond information provision, recognising that manipulative design practices undermine autonomy in ways consumer law alone cannot remedy. Additionally, existing consumer law applies horizontally, but typically only wherever there is a contract. Manipulative environments, however, often operate before or around contractual relationships: in the design of recommender systems, the structuring of default settings, or the bundling of consent into sign-up flows. The DFA must explicitly address these pre-contractual and non-contractual environments, since these are precisely the moments where autonomy, rights, and fairness are most at risk.

The DFA must bridge these gaps not by duplicating existing protections, but by targeting their gaps. It must address how digital services use recommender systems, predictive nudging, and design-driven persuasion to steer behaviour and deepen inequality. And it must do so by setting legal red lines, reversing burdens of proof, and equipping regulators with the tools to detect and address manipulation at scale.

This background paper explores a series of concrete proposals for how to the Commission could achieve this in the future DFA. **These suggestions cover unfair personalisation, manipulative and addictive design, coercive consent flows, harmful profiling, and liability across the digital advertising ecosystem.** Our research shows that the DFA should advance fairness as a unifying legal principle that bridges consumer protection, data protection, and platform governance. **Without such an integrated approach, the EU will continue to address threats to fundamental rights in a piecemeal and ineffective manner.**

While this paper focuses on the DFA, the principles we propose are equally relevant to future reforms of other parts of the EU digital rulebook, especially any revised ePrivacy piece of legislation. Digital fairness is not a siloed issue: it cuts across legal domains and demands a coherent, rights-based response. The DFA is the place to begin.

This brings us to one of the most pressing policy debates today: the protection of children online. While essential, this debate is too often framed narrowly, leading

policymakers towards flawed solutions, rather than addressing the harmful design of digital environments as a whole.

While the protection of children and young people online is essential, new legislation must be grounded in a commitment to uphold the rights of all people. A narrow focus on minors alone risks encouraging flawed measures, like ineffective and intrusive age verification, that exclude rather than protect, while leaving the broader ecosystem harmful by design. Crucially, minors are not only harmed when they go online. They are harmed by the online environment, whether or not they are connected: when toxic content and manipulative platforms shape their peers, families, schools, and cultures; when advertising and profiling systems target them indirectly through others; when default designs normalise surveillance, addiction and commercial exploitation as acceptable standards. Ensuring that digital services are fair and safe by design for everyone would create an internet that protects minors and their rights meaningfully, without isolating them or pushing them into riskier and more opaque spaces.

At the same time, **key protective frameworks such as the GDPR are under growing attack**, framed by some stakeholders as barriers to innovation, AI competitiveness, or small and medium enterprise (SME) growth. These arguments often echo earlier deregulatory narratives: that Europe must 'catch up', that harmonisation requires simplification, or that consent and transparency obligations are too burdensome.

The DFA risks becoming a collateral victim of this deregulatory push. Industry narratives increasingly suggest that existing laws already protect people, that the real issue is overlap, and that digital literacy would be enough to fix the rest. But enforcement experience and the EC's own Fitness Check show otherwise: practices that threaten fundamental rights persist precisely because current frameworks were never designed to address them. There is no meaningful overlap when rights violations remain untouched, no literacy remedy for systems engineered to mislead, pressure, or addict by design, and no fairness when people are forced to navigate opaque systems without transparency, redress, or meaningful alternatives. **The DFA's role is not duplication, but closing the gaps that existing legislation - however valuable - cannot reach**.

Measures proposed for simplification - including reduced information obligations, weakened withdrawal rights, or vague calls for 'burden reduction' – increasingly serve as cover for a broader political agenda that questions the legitimacy of rights-based regulation The emphasis on simplification is driving a deregulatory agenda, weakening the EU's normative commitment to ensuring a high level of protection for individuals and eroding the broader societal function of consumer law. **Simplification, when framed only around compliance costs, obscures the value of regulatory friction: the very safeguards that prevent abuse, enforce accountability, and sustain trust – ultimately protecting our fundamental rights.**

In digital environments defined by opacity and structural asymmetries, weakening protective rules does not create fairness or innovation. It creates impunity. As such, **the DFA must explicitly reject the false opposition between regulatory protection and competitiveness and put people first, not dominant market players**.

III. Digital Fairness and Vulnerabilities

Executive Summary

Current EU consumer law is no longer sufficient to address the structural manipulation embedded in digital environments. Platforms shape behaviour not through isolated deception, but through systemic practices that exploit persistent power asymmetries. **The DFA must embed fairness as a structural obligation, not a feature-level fix.**

Digital environments today put all users in a vulnerable position by design through opaque optimisation, personalisation, and data-driven design. Vulnerability is not limited to fixed traits but emerges from lived experience, context, and systemic exclusion. The DFA must protect people not only as consumers but as individuals situated within unequal socio-technical systems.

Key Policy Recommendations

- **Redefine legal concepts**:
  - Consumer: widen the definition to explicitly include any user subject to profiling or personalisation, even where no monetary payment is made.
  - Vulnerable consumer: expand current demographic categories to instead define 'vulnerability' functionally, to reflect contextual and system-induced exposure, not only demographic categories.
  - Trader: clarify that the updated definition of trader includes all actors involved in providing or operating digital services, understood broadly to cover those who deploy algorithmic systems, providers of recommender or ranking mechanisms, and other intermediaries whose optimisation or design choices influence user interaction or decision-making, even where no direct contractual relationship exists with the end user.
- **Recognise societal impact**: Digital service design shapes access to services, participation, and rights. Fairness must function as a safeguard for democracy, not only as a condition of individual choice.
- **Address structural manipulation**:
  - Regulate not just deceptive features, but the optimisation logic, personalisation architecture, and feedback systems that steer behaviour.
  - Treat manipulative interaction models as violations of professional diligence, even without overt deception.
- **Codify fairness by design and by default**:
  - **Structural fairness obligations:** Move beyond banning isolated practices and embed fairness-by-design and resilience-by-default as general duties. Require services to support goal-oriented use, natural exit points, and meaningful consent, especially at onboarding.
  - **Reverse the burden of proof:**
    - ✔ Traders must prove fairness and non-exploitation with auditable evidence.
    - ✔ Regulators must have access to internal optimisation data and risk assessments.

- ✔ Introduce a dynamic grey list that presume certain practices unfair unless traders demonstrate otherwise. Keep the list open and adaptable to new forms of manipulation.
- ✔ **Mandatory Impact Assessments**: Require traders to conduct and document **Behavioural Design Impact Assessments** (BDIAs) for all adaptive or personalised systems where a presumption of unfairness applies. BDIAs should detail the system's intended behavioural effects, affected or vulnerable groups, testing results, and safeguards, and must be made available to regulators to support effective oversight and preventive enforcement.

- **Address the scale and structural nature of manipulation in digital markets:**
  - ➢ Propose the DFA as a Regulation with a hybrid structure, combining self-standing, directly applicable provisions with targeted amendments to core consumer law Directives.

The DFA as a Structural Shift in Modernising Consumer Law

Within EU consumer law, the concept of unfair commercial practices originated with the aim to safeguard people from deception, coercion, or exploitation of information asymmetries in their dealings with traders. While these protections remain important, they no longer capture the full extent of the challenges to people's rights in today's digital environments. **As the design of digital services increasingly determines whether people can exercise their rights to autonomy, equality, and participation in society, a broader and more ambitious understanding of fairness is needed, one that explicitly protects fundamental rights.** The DFA must recognise that digital systems are not neutral: they are deliberately built to maximise engagement, data extraction, or monetisation, often at the expense of people's agency and dignity. When design choices systematically prioritise those commercial optimisation strategies over fundamental rights, they generate structural imbalances of power with far-reaching consequences for individuals and society.

The DFA offers a timely opportunity to ensure that the definition of fairness places people's rights – including self-determination, the protection of personal data and privacy, and freedom of expression – at its core. **While building on existing consumer protection frameworks, the DFA should mark a broader shift: from protecting individuals solely in their role as consumers during discrete isolated economic transactions to protecting people more comprehensively in their everyday digital lives. This means safeguarding them across diverse roles, social positions, and situations – as workers, learners, patients, tenants, or people on the move – where manipulative design and commercial optimisation practices can** restrict visibility, limit autonomy, and undermine equal participation. Such systems not only distort individual choice but can also reproduce discrimination, reinforce existing inequalities and exploit vulnerabilities over time, regardless of whether someone is formally acting in a consumer capacity.

To speak of fairness in such settings therefore requires more than just protecting 'the average user'. The very notion of an average user is ill-suited to (digital) regulation, as

it erases the structural conditions and intersecting inequalities that shape how people engage with digital services. By presuming that there can be a standard or neutral person, anyone with a minoritised identity (e.g. a person with a disability) is constructed as being abnormal. This is not only unfair in market terms but directly undermines equality and non-discrimination principles. Digital asymmetries also create environments of induced vulnerability, in which **all users may be exposed to manipulation, though not all are affected equally, while also recognising that specific contexts (e.g. age), constraints, or lived experiences can further intensify this vulnerability**. A structural approach, therefore, requires acknowledging that most people will, at some point, face reduced agency or be subject to manipulation. This is not due to individual characteristics, but because digital services are optimised to exploit behavioural vulnerabilities, particularly under conditions of cognitive load, emotional distress, or persuasive design. These forms of design-driven exploitation interfere with people's right to self-determination and can undermine both mental health and data protection rights.

A growing body of research identifies that **vulnerability is not an individual deficit, but a condition actively produced by the design of digital choice architectures to exploit asymmetries of information, agency, and visibility**. In such environments, all users are made vulnerable by default, though threats to rights are often intensified for those already facing social or economic disadvantage[4].

Studies show that most people, regardless of education or experience, struggle with online choice architectures. For instance, an OECD study into consumers in the UK found that they spent over GBP 1.6 billion annually on unwanted subscriptions. These practices systematically undermine autonomy and informed choice, eroding trust and restricting the effective enjoyment of rights, with concrete harms such as financial loss, wasted time, frustration and diminished trust in digital services. Subscription traps were shown to affect everyone but disproportionally impact older users, lower-income groups, and those with health conditions, demonstrating how manipulative design can systematically undermine people's ability to make and act on informed choices[5].

A key starting point for this is the recognition that a majority of digital spaces are fundamentally characterised by structural asymmetries of information and power. **These asymmetries are not incidental; they are designed, but also hidden from view.** Platforms and digital services operate on the basis of extensive knowledge about users, derived from data collection, profiling, and behavioural prediction. Users, in contrast, have extremely limited understanding of how these systems work, how decisions are made, or how their behaviour is being steered.

[4] OECD, "Consumer vulnerability in the digital age", OECD Digital Economy Papers, No. 355 2023 https://www.oecd.org/en/publications/consumer-vulnerability-in-the-digital-age_4d013cc5-en.htm.

[5] OECD, "Consumer vulnerability in the digital age", OECD Digital Economy Papers, No. 355 2023 https://www.oecd.org/en/publications/consumer-vulnerability-in-the-digital-age_4d013cc5-en.htm.

Exploitation in this context takes many forms: some practices distort perception and nudge users toward particular outcomes, while others eliminate meaningful alternatives altogether[6]. These dynamics are embedded in digital service design, operating through emotional triggers, persuasive defaults, and personalised content distribution. **In such environments, users are placed in positions of exposure they cannot fully control, engaging on terms they cannot fully see or influence.**

The language of 'vulnerable users' risks misplacing responsibility and narrowing the scope of protection. Instead, it is important to recognise that these are digital environments that induce or exacerbate vulnerability. Rather than treating vulnerability as an intrinsic personal trait, the DFA should recognise that today's toxic digital systems actively produce exposure through profiling, opacity, coercive design, and emotional manipulation. In these environments, **exposure to manipulation is widespread but uneven, intensified where systems intersect with pre-existing inequality**. Some people are more closely profiled, more heavily targeted, or more easily coerced based on realities like socio-economic status, race, gender, disability, or age. **Fairness by design therefore requires moving from individualised notions of vulnerability to a systemic understanding of how platforms deliberately put people in vulnerable situations.**

Exposure to manipulation is thus not evenly distributed. It is intensified - structurally and contextually - for certain groups and communities[7]. In particular, children, older people, neurodivergent people, people with disabilities, racialised communities, individuals with access barriers to digital literacy and confidence in navigating digital systems, or people living in precarious socio-economic conditions often face additional, layered barriers to digital autonomy.

These are not simply 'special cases', but manifestations of how digital systems reproduce and amplify structural inequalities and contextual disadvantages[8]. Moreover, **this differential exposure does not operate in isolation: it intersects with one another and with broader systemic injustices**[9]. For example, a recent German study shows how digital literacy and digital confidence are unequally distributed along lines of gender and class. Many low-income or care-related professions, which are disproportionately occupied by women, offer limited access to digital tools, reducing opportunities to build the skills needed to resist manipulation online[10].

---

6 Rebrean, Maria-Lucia and Malgieri, Gianclaudio, Vulnerability in the EU AI Act: building an interpretation (November 28, 2024). FAccT '25: Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. https://ssrn.com/abstract=5058591

7 SUPERRR Lab, Thesenpapier: Für eine gerechte digitale Zukunft – Acht Thesen zu Digitaler Fairness, Berlin, April 2023. https://superrr.net/media/pages/projects/forum-digital-fairness/9c957db06e-1734347932/digital-fairness-thesenpapier.pdf

8 Miller, A. D. (2024, forthcoming). Invisible Allies: Algorithmic Consumer Profiling and the Rise of New Group Harms

9 Rossi, A., Carli, R., Botes, M. W., Fernández, A., Sergeeva, A., & Sánchez Chamorro, L. (2024). Who is vulnerable to deceptive design patterns? A transdisciplinary perspective on the multi-dimensional nature of digital vulnerability. Computer law & security review, 55, Article 106031.

10 Der Paritätische Gesamtverband, 2024. Armut führt zu digitaler Ausgrenzung. https://www.der-paritaetische.de/alle-meldungen/neue-studie-armut-fuehrt-zu-digitaler-ausgrenzung/

Digital vulnerability is not static or reducible to fixed socio-demographic traits. It emerges through contextual and experiential conditions, such as lacking social support, feeling unable to assert one's preferences, or facing digital environments that present impossible trade-offs. These conditions increase both people's exposure to manipulation and their difficulty in resisting or recovering from its effects[11].

The DFA should thus move away from the static category of 'vulnerable users' and instead recognise that digital systems routinely produce environments of exposure, where anyone can be manipulated, and some more systematically than others. This requires **amending the UCPD to reflect that digital environments create structural asymmetries that necessitate generalised protections for everyone, with additional safeguards where inequality intersects with digital exposure.**

In this regard, recent academic and civil society work on a dedicated digital fairness framework rightly suggest codifying a **trader's obligation not to exploit digital asymmetry or vulnerability**[12], something that legal scholars call "digital professional diligence"[13]. These contributions point to the need for a broader framework that goes beyond targeting individual features. Manipulative design is not limited to discrete features but reflects a deeper logic of exploitation, where asymmetric knowledge and behavioural insights are used to extract data, attention, or consent. The DFA must enshrine this principle in law, anchoring fairness in a rights-based standard of professional diligence, ensuring that people's digital autonomy is respected.

The DFA must thus treat manipulative design as a **structural problem, not a matter of isolated tricks. Exploiting people through design is a violation of professional diligence, even when there is no obvious lie or single deceptive feature.** Legal obligations must apply not only to surface-level design choices, but also to the deeper interaction models and commercial optimisation strategies that steer, pressure, or trap people into unwanted outcomes[14]. A fairness framework that is fit for the digital age must start from the reality that digital systems often create vulnerability by design, and that companies profit from this exposure[15]. **Structural fairness means shifting the focus from labelling certain people as inherently 'vulnerable' and instead regulating the systems that generate and exploit vulnerability in the first place, and thereby threaten fundamental rights.**

---

11 Sánchez Chamorro, L. (2024). Disentangling Vulnerability to Manipulative Designs: An Experiential Perspective to Rethink Resistance Strategies (Doctoral dissertation, University of Luxembourg.

12 Helberger, N., Kas, B., Micklitz, H.-W., Namysłowska, M., Naudts, L., Rott, P., Sax, M. & Veale, M., Digital Fairness for Consumers, BEUC, Brussels, March 2024, https://www.beuc.eu/sites/default/files/publications/BEUC-X-2024-032_Digital_fairness_for_consumers_Report.pdf

13 Namysłowska, M., The Silent Death of EU Consumer Law and Its Resilient Revival: Reinventing Consumer Protection Against Unfair Digital Commercial Practices, Journal of Consumer Policy, 2025.

14 Leiser, M., & Santos, C. (2024). Dark Patterns, Enforcement, and the Emerging Digital Design Acquis: Manipulation beneath the Interface. European Journal of Law and Technology, 15(1).

15 Helberger, N., Sax, M., Strycharz, J. & Micklitz, H.-W., Choice Architectures in the Digital Economy: Towards a New Understanding of Digital Vulnerability, Journal of Consumer Policy, 2021; G. Malgieri, Vulnerability and Data Protection Law, OUP 2023.

Fairness by Design and by Default[16]

The DFA must therefore adopt a relational concept of fairness that accounts for structural dependencies and systemic asymmetries. A transactional model, based solely on individual choices and disclosures, fails to capture how platform dominance, addictive design, or profiling restrict genuine user agency. **Digital fairness must be more than procedural compliance. It is not sufficient for a system to obtain consent or disclose information.** The DFA should evaluate whether digital environments are structured to produce substantively fair outcomes, measured by autonomy, non-discrimination, and meaningful control.

Fairness by design and by default is needed to counter how digital systems structurally encode, normalise, and obscure exploitative dynamics, undermining people's rights, autonomy, and dignity. Fairness must be embedded into the design, deployment, and functioning of digital systems as an *ex ante* responsibility for traders. This includes creating an obligation on such providers to anticipate and prevent conditions that produce asymmetries of influence, structurally distort user autonomy, reinforce inequality, and enable exploitative targeting. **Such forward-looking duties are essential to restore user trust and protect the foundations of human dignity and democratic participation in digital environments[17].**

Fairness by design therefore requires anticipating risks and injustices in the conception and development of digital systems before they are entrenched. **It involves shifting the burden of action away from individuals, who are often expected to navigate opaque systems, manage consent flows, or protect themselves through vigilance, and instead placing the responsibility on traders (as defined here to include platforms, intermediaries, and developers)** as the architects of interaction environments and gatekeepers of influence.

At the level of everyday interaction, this means eliminating manipulative, confusing, or coercive patterns, including so-called 'dark patterns' that steer users into consent or engagement based on immediate urges rather than reflective, informed intent. It means making systems intelligible, accessible, and usable for people with different cognitive, linguistic, or physical capacities. And **it also goes further: fairness by design means recognising that digital services are not neutral or merely functional: they are sites of governance, where decisions about visibility, accessibility, nudging, and**

---

16 We are grateful to BEUC for being the first civil society organisation to introduce this terminology, and for their longstanding, outstanding work on digital fairness in consumer law. See, for example, BEUC (2023) Towards European Digital Fairness. BEUC framing response paper for the REFIT consultation. https://www.beuc.eu/sites/default/files/publications/BEUC-X-2023-020_Consultation_paper_REFIT_consumer_law_digital_fairness.pdf

17 Karen Yeung, A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework, Council of Europe, DGI(2022)11, November 2022. https://rm.coe.int/a-study-of-the-implications-of-advanced-digital-technologies-including/168096bdab

friction[18] actively determine who gets to participate, how, and under what terms, and with what consequences.

The Wider Societal Impact of Digital Services

This systemic understanding of fairness is crucial for addressing the increasingly infrastructural role that digital services play in society. Digital services do not simply mediate individual consumer choices: they shape the public sphere, influence access to services, determine labour conditions, and structure the flow of information. **The fairness of such design choices therefore has indirect consequences for fundamental rights, including equality, non-discrimination, data protection, freedom of expression, and the right to participate in democratic life. For example, when engagement-driven recommender systems amplify disinformation, they directly harm the right and the ability to participate in democratic life and distort the public sphere.** By focusing only on consumers and transactional fairness, current legal frameworks contain gaps and do not adequately address wider societal effects. The DFA should therefore embed fairness as a systemic safeguard, one that protects people as individuals and as members of a democratic society.

Importantly, **fairness is not only about mitigating discrete harms: it is about enabling justice.** A rights-based approach to fairness in digital design entails embedding equity, accessibility, and empowerment as guiding principles. It involves **creating conditions under which all people - not just the most digitally literate or resourceful - can exercise meaningful agency.** This requires inclusive design practices, participatory governance, and transparency around how design decisions are made and who is responsible for them. It also requires robust accountability mechanisms and enforcement tools that can identify and rectify unfair practices at scale, including where these practices disproportionately affect racialised communities, people with disabilities, migrants, or other structurally minoritised groups.

Only by embracing this ambitious and systemic vision of fairness can we begin to build digital environments that serve people, respect their rights, and resist the structural production of vulnerability and exclusion In the context of the DFA, this means **ensuring that fairness is not only a requirement for specific practices, but a structural obligation that guides the development, deployment, and assessment of digital services at every stage of their development and use.** Only by embracing this ambitious and systemic vision of fairness can we begin to build digital environments that serve people, respect their rights, and foster just and equitable societies.

The Need For New Concepts and Definitions

Throughout this background paper, we use terms such as 'deceptive design', 'addictive design', and 'unfair personalisation' to describe specific forms of manipulative practices. However, we do not understand these concepts as isolated or purely visual

[18] Understood as deliberate obstacles, extra steps, or delays that interfere with people's autonomy and obstruct the exercise of their rights.

design choices. Rather, they are **expressions of deeper system-level manipulation, rooted in data extraction, profiling, behavioural targeting, and optimisation logics embedded in digital infrastructures**[19].

The European Commission and co-legislators should use the DFA as an opportunity to introduce or revise the existing legal definitions of 'consumer', 'vulnerable consumer', 'trader', and 'digital unfairness' across the relevant EU directives[20]. **The existing definitions, conceived in an offline transactional context, do not sufficiently reflect the deeply embedded, opaque, data-driven, and increasingly automated asymmetries and dynamics that characterise today's digital environment.** While asymmetries of power and information have always existed in consumer-trader relationships, what distinguishes the digital context is their structural entrenchment and increasing invisibility to users. The traditional framing assumes a rational consumer making discrete decisions, and a trader offering goods or services in a transparent marketplace. This model fails to capture the structural asymmetries and highly personalised nature of digital markets, where users are persistently profiled, nudged, manipulated, and segmented in ways they cannot perceive or resist, and where their exposure to content, choices, and friction is determined by opaque personalisation systems and algorithmic design.

Definition: Digital Unfairness

The consumer *acquis* does not currently define 'digital unfairness'. We therefore propose a new definition to be included in the DFA: "Digital unfairness refers to the **structural and persistent distortion of people's digital environments in ways that exploit structural asymmetries of power, information, and visibility, and that manipulate behaviour, restrict autonomy, or deepen inequality**."

Why? Digital unfairness occurs when systems, including interfaces, algorithms, personalisation mechanisms, and data-driven architectures, manipulate behaviour, restrict autonomy, or deepen inequality, especially without meaningful transparency, contestability, or alternatives. Unfairness is not limited to individual deception or harm, but includes cumulative, collective, and often invisible impacts on attention, choice, consent, and participation.

[19] Mark Leiser, Dark Patterns, Deceptive Design, and the Law (Hart Publishing, 2025).

[20] This background paper builds on and aligns considerably with the European Parliament's 2023 resolution on addictive design and consumer protection (P9_TA(2023)0459) https://www.europarl.europa.eu/doceo/document/TA-9-2023-0459_EN.html, which recognises the growing evidence of systemic manipulation in digital interfaces and the need for a comprehensive regulatory response (hereinafter, 'INI Report').

Definition: Consumers in the Digital Environment[21]

The DFA must expand the definition of consumer to account for the realities of digital platforms and services. As such, we propose the following definition: ***"'Consumer' means any natural person who acts, within digital environments, for purposes which are outside their trade, business, craft or profession, regardless of whether the service is accessed for remuneration or free of charge. In the context of automated, personalised, or data-driven services, a consumer shall be understood to include any user who is subject to behavioural targeting, profiling, or personalisation initiated by a trader or an affiliated system."***

Why? In digital contexts, consumers often interact with services through asymmetrical systems of influence, where transactional boundaries are blurred and personal data functions as a form of currency or leverage. Exposure to profiling and manipulation is not evenly distributed, but shaped by the intersection of digital architectures with broader inequalities.

Definition: Vulnerable Consumers in the Digital Environment

The DFA needs a new definition which captures the structural vulnerability to which users are subjected: ***"'Vulnerable consumer' means any consumer who, due to individual characteristics, situational context, or intersecting inequalities is particularly susceptible to manipulation, coercion, or exploitation."***

Why? In digital contexts, all consumers face structural risks of manipulation, especially in environments characterised by profiling, automated decision-making, or personalised design. Vulnerability does not arise solely from individual characteristics, but from the interaction between people and systems deliberately designed to profile, optimise, and extract. It may be intensified by factors such as age, disability, digital literacy, socio-economic status, or emotional state, but it also emerges through systemic asymmetries in information, visibility, and control that characterise digital services. This proposed approach must complement, rather than replace, existing consumer law protections for groups already recognised as needing additional safeguards. The aim is to broaden protection to reflect how digital systems can create situations of vulnerability for all users, while retaining targeted additional safeguards for those who require them.

To avoid tensions with the full harmonisation nature of EU consumer law, the DFA should not replace the existing UCPD concepts of the 'average consumer' and the 'particularly vulnerable consumer'. Instead, it should amend these definitions to reflect the realities of digital environments. As mentioned, vulnerability online is structural, but also contextual and situational: interface design, profiling, or behavioural targeting can place any person in a position of disadvantage, regardless of their baseline

[21] While this paper adopts the terminology of 'consumer' to align with the legal architecture of the DFA and EU consumer law, it does so critically. The goal is not to reduce people to market actors, but to extend protections to everyone affected by manipulative digital systems, recognising individuals in their full range of roles, identities, and vulnerabilities, including those that emerge situationally through design.

characteristics. **Preserving the existing *acquis* while clarifying that vulnerability can be dynamic and design-induced would provide legal continuity and enforcement certainty**, while ensuring protection extends to the structural conditions that impair autonomy and fairness for all users.

Definition: Traders in the Digital Environment

The DFA also needs to update the definition of trader by amending the existing *acquis* (UCPD/UCR) to ensure clarity in digital markets: **"'Trader' means any natural or legal person who acts for purposes relating to their trade, business, craft, profession, or revenue-generating activity in the digital environment, including developers of digital services, adtech intermediaries, and service providers. This includes those who develop or deploy systems that influence consumer decision-making through profiling, personalisation, or design-based steering." This approach preserves continuity with current law while explicitly extending its scope to cover the full range of actors shaping digital choice architectures.**

Why? A trader may also include actors who, while not directly contracting with the consumer, shape the consumer experience through technological infrastructure, data processing, algorithmic recommendations, or embedded service systems, and who generate revenue from such involvement. This definition extends responsibility beyond direct contractual parties to include those who exercise influence over the consumer journey in a structured; for example, providers whose systems determine defaults, steer attention, or set the conditions under which choices are made.

Reversing the Burden of Proof

The DFA should introduce a **general principle of fairness by design, applicable across digital services**. This principle would apply to the same expanded category of traders in the digital environment outlined above, and would:

- Require traders to design and operate digital environments that respect users' autonomy, agency, and ability to make free and informed decisions;
- Prohibit traders from implementing practices that systematically distort or obstruct user decision-making through design, profiling, or manipulation;
- Apply *ex ante* to the architecture of digital services, rather than relying only on *post hoc* assessment of individual transactions.

This principle builds on, but goes beyond, Articles 5–9 of the UCPD by addressing structural forms of manipulation that we argue are currently under-regulated. While Articles 5–9 prohibit unfair practices in specific interactions, **the DFA should go further by embedding fairness as a design obligation. This means adopting the above three obligations as binding principles, and that fairness must be a design obligation, not merely as a case-by-case assessment of individual interactions.**

The UCPD was once praised for its flexibility allowing regulators and courts to adapt the notion of 'unfairness' to new market behaviours. But this flexibility is no longer enough: manipulation is now built into the algorithms and systems that shape digital

environments, and the UCPD cannot adequately address the resulting structural threats to autonomy and fairness. Despite being the cornerstone of the law, many scholars and practitioners consider Article 5 UCPD to be too vague to offer effective protection against manipulative and other exploitative design practices[22]. The general clause defines unfairness through abstract standards such as the failure to meet 'professional diligence' and the capacity to 'materially distort' consumer behaviour. However, **the ambiguity of these concepts has enabled traders to justify manipulative designs in digital services as innovation or legitimate commercial strategy, rather than unfair practice subject to legal constraint.**

Instead of empowering enforcers, this interpretative leeway has often resulted in delayed action, weak deterrence, and inconsistent enforcement across Member States. In short, **the legal flexibility offered by Article 5 UCPD has not benefited consumers as much as it was supposed to.** To ensure legal certainty and meaningful protection, **the general clause should be amended but also complemented by a broadening of the list of blocked practices via explicit amendments to Annex I of the law to include those that are known to be unfair**. The DFA should thus provide a modernised general prohibition tailored to digital unfairness, and explicitly ban practices known to undermine autonomy and rights, including profiling-based nudging, and real-time interface adaptation[23].

Voluntary tools like time-use dashboards or screen-limit settings are often presented as sufficient safeguards. But they leave the burden on individuals to resist designs that are deliberately calibrated to undermine self-regulation. What is needed is not hard-coded time caps, but clear legal obligations on traders not to deploy features that are structurally manipulative by design. The responsibility must rest with those who create and monetise the manipulative architecture, not with individuals to fight against it.

In addition, to account for evolving manipulation strategies and ensure future-proof enforcement, **the DFA should** establish **a structured annex model, akin to the UCPD, not only to list categorically unfair practices, but also to expand over time as new design patterns and exploitative techniques emerge**. This modular structure would enable the Regulation to reflect the systemic nature of manipulation in digital environments, distinguishing between outright prohibited practices (block list[24]), presumptively unfair strategies (grey list), and potential future annexes focused on structural obligations for traders.

A new grey list (which could take the form of an Annex III to the UCPD, introduced through the DFA) should be created. Unlike the Annex I list of the UCPD, which prohibits certain practices outright, a grey list would establish a category of practices that are presumed unfair unless the trader can demonstrate otherwise. Embedding this annex structure within the UCPD ensures continuity with existing consumer law,

22 M. Namysłowska, 'The Silent Death of EU Consumer Law and Its Resilient Revival: Reinventing Consumer Protection Against Unfair Digital Commercial Practices', Journal of Consumer Policy, 2025.

23 *Ibid.*

24 We note that the term 'blacklist' is often used in policy discussions. However, we prefer to avoid the term due to its racial connotation, and therefore use 'block list' instead.

while the DFA provides the modernisation needed to address systemic and evolving forms of digital manipulation. This does not mean that regulators must detect manipulation from the outside. What matters is the design logic inside digital systems themselves. Traders routinely shape users' decisions through what is often called Digital Choice Architecture (DCA)[25] - the way systems structure and present options to steer decisions: altering defaults, ordering options, manipulating salience, or sequencing prompts to nudge outcomes. The DFA should treat DCA as a commercial practice in its own right and assess its fairness based on its cumulative impact on autonomy and self-determination.

Where design practices are plausibly exploitative, the burden of proof must fall on those who deploy and exploit them. Users cannot meaningfully challenge what they cannot see, and regulators cannot assess hidden optimisation strategies[26]. **Traders are the only actors with full visibility and control, and should be legally required to show that their systems comply with fairness obligations and do not exploit digital vulnerability.**

To meet this burden, **traders should have to produce verifiable, auditable documentation demonstrating that their systems:**

- Are explainable and non-discriminatory;
- Enable meaningful user agency;
- Do not exploit vulnerabilities; and
- Do not rely on deception, coercion, emotional exploitation, or disproportionate friction.

The practices that should be placed on the grey list share structural features that indicate a high risk of manipulation, such as reliance on personalisation, emotional steering, or interface friction. This constitutes *prima facie* evidence of potential unfairness, sufficient to shift the procedural burden. Where doubts remain, the design should be assessed against published guidelines on design fairness, and presumed unfair unless convincingly justified under EU consumer law.

The DFA should therefore establish a grey list of practices that are presumed unfair unless traders can demonstrate otherwise. This list must cover emerging and evolving design patterns that exploit digital asymmetries, as explained in the relevant sections below. Crucially, **the grey list should be designed as an open, non-exhaustive list, capable of accommodating novel practices as they develop**. Without such a flexible tool, the law risks always lagging behind new forms of exploitation, leaving regulators without the means to address future violations of people's rights. This flexibility is essential to ensure regulators can respond to novel strategies without requiring new legislation each time. Crucially, however, this approach should not be confused with the DSA's systemic risk mitigation model[27]. Rather, it echoes a core tenet of EU

[25] The concept of DCA comes from behavioural economics and regulatory studies, and has been taken up in consumer law debates to describe how digital services structure options to steer behaviour.

[26] A. D. Miller, Invisible Allies: Algorithmic Consumer Profiling and the Rise of New Group Harms, 2024.

[27] While some legislation like the AI Act and DSA rely on broad risk categories and discretionary enforcement, the DFA's approach to manipulation should be rooted in clear, observable features of interface design and their effects on user agency. Rather than assess abstract risk levels, the grey list

consumer law: that fairness must be evaluated in light of evolving commercial practices. The grey list would give regulators the procedural tools to flag emerging manipulative strategies while still ensuring legal certainty through published criteria and rebuttable presumptions, rather than discretionary obligations. To address concerns about enforcement abuse, any inclusion in the grey list should require a demonstration of structural similarity to listed practices and must be guided by publicly-available criteria that focus on measurable distortions of user autonomy, not abstract risk.

Far from creating new burdens, a system of blacklists, greylists, and burden-shifting provides clarity and predictability. Compliant businesses know exactly which practices are prohibited, which are presumed unfair unless justified, and what kind of evidence they need to demonstrate fairness. This reduces legal uncertainty and ensures enforcement focuses on those who profit from manipulative design, not on businesses who already act fairly.

Reversing the burden of proof for these types of practices is essential to effective enforcement and reflects the power imbalance that defines digital markets. Specific recommendations are contained in the Chapter VIII below. Without this reversal, opacity and complexity will continue to shield exploitative practices from accountability. Digital technologies must be governed in ways that sustain the conditions for meaningful autonomy, non-discrimination, and collective self-determination. **The DFA must be part of a broader regulatory shift that addresses structural power asymmetries and curtails exploitative design at its source**[28].

Operationalising Fairness by Design and by Default[29]

Even if a few actors adopt fairer practices (e.g. transparent pricing), the presence of others who obscure key terms can cancel out those benefits. This underlines the **need for baseline regulatory duties that apply across the market, preventing strategic under-disclosure and pre-empting design asymmetries that would otherwise distort competition or punish traders who act responsibly**[30]. Regulators should be equipped to both detect and quantify systemic harm. While some forms of harm are difficult to

would identify structurally exploitative patterns that distort decision-making, with the burden placed on the trader to justify their design. This balances legal certainty with regulatory foresight, without outsourcing enforcement to future political discretion or technical standard-setters.

28 Karen Yeung, A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework, Council of Europe, DGI(2022)11, November 2022. https://rm.coe.int/a-study-of-the-implications-of-advanced-digital-technologies-including/168096bdab

29 Tim de Jonge, Hanna Schraffenberger, Jorrit Geels, Jaap-Henk Hoepman, Marie-Sophie Simon, and Frederik Zuiderveen Borgesius. 2025. If Deceptive Patterns are the problem, are Fair Patterns the solution? In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25). Association for Computing Machinery, New York, NY, USA, 3131–3137.

30 Behavioural Insights Team, The behavioural science of online harm and manipulation – and what to do about it, March 2022. https://www.bi.team/publications/the-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it/

observe directly, it is possible to estimate both the market at risk and the realised harm through a combination of user research, behavioural experiments, and top-down or bottom-up modelling methodologies. Without such tools, structural manipulation may remain legally invisible despite its wide impact[31].

To move beyond reactive enforcement and embrace positive duties, not just prohibitions, the DFA should therefore not only prohibit specific practices such as deceptive design, addictive design patterns, and unfair personalisation, but also establish fairness as a structural obligation in service design. **This should shift the burden away from overburdened users and toward those who shape the digital environment.**

Fairness must be embedded *ex ante* into system design. Traders should be required to demonstrate, through structured assessments, that their products do not rely on manipulation, emotional exploitation, or undue friction to drive engagement or monetisation. The following tools would help translate the concept of 'fair patterns' into legal practice:

- **Codify design fairness as a general duty**: while expanding the UCPD "block list" is essential to prohibit known structurally exploitative practices, the DFA must also go further by embedding a general duty of fairness in digital service design. This would allow regulators to address both listed and emerging forms of manipulation, and to intervene where design logic systematically undermines user autonomy, even in the absence of a clearly prohibited practice. To make this enforceable, the general duty should be reinforced by presumptions of unfairness for high-risk design techniques, and complemented by bright-line prohibitions in the UCPD "blacklist". Together, these tools offer a scalable framework: a structural baseline (the duty), targeted presumptions (to reverse the burden of proof), and categorical bans (for legal certainty). This layered approach ensures regulators can tackle deceptive design not only where it is obvious, but also where it is systemic, subtle, or deliberately evasive.

- **Mandate resilience by default in digital service design**: to reduce structural exposure to manipulation, the DFA should establish a legal obligation for resilience-by-default in digital service design. These obligations recognise that **vulnerability and exposure to manipulation arise not from intrinsic traits but from situational, relational, and systemic conditions shaped by digital architectures**. Resilient design is not an optional feature but a structural safeguard against exploitative environments. This would mean requiring that digital services:
  - **Support goal-oriented interaction**: Digital services should allow users to articulate or select a goal (e.g. finding specific information, making a one-time purchase) and be shielded from design features that divert attention or prolong engagement beyond that purpose.

[31] London Economics, Digital Consumer Harms: A taxonomy, root cause analysis and methodologies for measurement, Report prepared for the UK Department for Digital, Culture, Media and Sport (DCMS), March 2023. https://assets.publishing.service.gov.uk/media/63c6813ce90e074ee5bb7d4f/DCMS_consumer_harms_research_01-Jan-22.pdf

- **Enable meaningful pause and exit mechanisms**: Users should be able to pause or interrupt interactions without losing progress, facing penalties, or being subjected to pressure to continue. Absence of stopping cues or frictionless exits should be presumed unfair where they systematically override user self-regulation.
- **Include disengagement safeguards**: Design flows must offer natural points of closure (e.g. end screens, summary stages, or reminders to log off), particularly in environments prone to compulsive loops. Practices that obscure exit points or default to endless engagement should be banned or subject to strict scrutiny.
- **Prohibit bundling of consent with onboarding flows**: coercive design practices are most acute during the first moments of use, when individuals are least able to evaluate the implications of their choices. Design fairness requires that meaningful consent be decoupled from device or service activation[32].

- **Mandate systemic auditability of digital service design and personalisation systems via BDIAs**: to make fairness-by-design enforceable, the DFA must establish auditability as a legal obligation. This includes granting regulators access to internal A/B testing results, optimisation metrics, and documentation of design rationales. Such access is essential not only for verifying compliance, but for understanding the behavioural assumptions and incentives embedded in digital service architecture[33].
  - Given that the DFA should recognise cumulative manipulation and sequential friction as core elements of unfair design, it should specifically require **audits of multi-step interaction flows, not just individual screens**[34].
  - In particular, BDIAs would be core audit instrument: for all complex digital systems that shape or personalise user environments. A full BDIA must be conducted whenever a presumption of unfairness applies - for instance, where profiling, personalisation, or optimisation techniques are likely to distort user autonomy or appear on the grey or black lists.
  - **Scope and content:** each BDIA must describe the system's purpose, behavioural assumptions, design logic, and intended effects; include the outcomes of internal testing (e.g. A/B experiments and optimisation metrics); and document mitigation measures. A plain-language summary should be made available to regulators and, where appropriate, to the public.
  - **Empower regulators with access and oversight:** Enforcement authorities must have the power to obtain internal documentation, testing data, and experimentation results to assess the behavioural assumptions and incentives embedded in service architecture. They must also be able to conduct **behavioural audits and pattern-based investigations** capable of

[32] AlExis Hancock (2019) Designing Welcome Mats to Invite User Privacy, EFF February 14, 2019. https://www.eff.org/deeplinks/2019/02/designing-welcome-mats-invite-user-privacy-0

[33] Mark Leiser provides a broader analysis in Dark Patterns, Deceptive Design, and the Law (Hart Publishing, 2025) by framing auditability not only as a reactive investigative tool, but as a proactive obligation embedded in the design process itself.

[34] H. Brignull, Deceptive Patterns: Exposing the Tricks Tech Companies Use to Control You (Testimonium Ltd, 2023).

identifying collective harms, such as designs that disproportionately target inferred traits like low self-esteem, financial distress, or anxiety.

- **Formalise the process**: BDIAs should **shift the burden of explanation upstream**, requiring traders to demonstrate that systems respect autonomy and fairness before deployment. Without enforceable audit powers and access to experimentation data, regulators will remain structurally excluded from understanding the dynamics of manipulation that define today's digital services.
- **Apply enhanced BDIA requirements to systems with advanced behavioural capabilities** - such as agentic AI, Large Action Models, or anticipatory personalisation - that adapt to or pre-empt user states and behaviours. These assessments must evaluate how such systems affect autonomy (free choice), emotional state (e.g. distress or fatigue), and goal-setting (diversion from user-intended objectives toward trader-benefiting outcomes).
- **Integrate meaningful stakeholder engagement[35]:** BDIA processes should involve affected stakeholders, including civil-society and vulnerable-group representatives, at early design stages. This ensures that fairness obligations are informed by lived experience, not only by internal testing or abstract ethical review.
- **Provide coordinated guidance**: The European Commission, together with enforcement authorities and civil society, should publish binding examples of fair and unfair design to guide implementation. Guidance should reflect established fairness principles such as symmetry, neutrality, non-bundling, and accessibility.
- **Ensure interoperability across regulatory regimes**: To prevent duplication and loopholes, BDIA templates should be interoperable with Data Protection Impact Assessments (DPIAs) under Article 35 GDPR and Fundamental Rights Impact Assessments (FRIAs) under the AI Act. Joint guidance and model templates should ensure that no single assessment can be used to bypass another's obligations.
- **Support cross-authority cooperation**: Effective oversight requires coordination between consumer protection bodies, data protection authorities, competition regulators, and media regulators to address systemic unfairness consistently across the EU digital rulebook.

- To make rights meaningful in practice, the DFA should also **guarantee users the right to human interlocution in digital environments**[36]. Automated systems often manage refusals, complaints, or consent withdrawal through standardised flows that are difficult to contest. Users must have the ability to seek clarification, challenge decisions, and escalate concerns to a human representative when needed. This is particularly important for people in vulnerable situations, or where automated interfaces deploy coercive defaults, block refusal pathways, or

[35] ECNL and Access Now, FrAmework For Meaningful Engagement: Human Rights Impact Assessments Of AI, 8 March 2023. https://ecnl.org/publications/framework-meaningful-engagement-human-rights-impact-assessments-ai

[36] SUPERRR Lab, Thesenpapier: Für eine gerechte digitale Zukunft – Acht Thesen zu Digitaler Fairness, Berlin, April 2023. https://superrr.net/media/pages/projects/forum-digital-fairness/9c957db06e-1734347932/digital-fairness-thesenpapier.pdf

obscure accountability. Without this right, fairness risks becoming a purely procedural abstraction with no meaningful redress.

Enforcement frameworks must be designed for responsiveness, allowing regulators to confront emerging structurally exploitative features and adapt to evolving exploitation patterns without waiting for legislative reform. This, as mentioned throughout the document, requires **moving beyond rigid or narrowly defined categories of unfairness and embracing a dynamic model in which design techniques are assessed in light of their effects on fundamental rights and on the cumulative conditions of autonomy and equality, not just their visual form or declared intent**.

Regulators should therefore be empowered to **identify and prohibit new structurally exploitative or addictive practices as they arise, including through updated guidance, delegated acts, and rebuttable presumptions of unfairness.** In a rapidly evolving digital environment, exploitative design strategies can be quickly rebranded, gamified, or layered into new digital services to avoid formal definitions. A responsive model would allow enforcement to focus on the underlying logics of attention extraction and behavioural manipulation, rather than chasing individual features after harm has occurred.

Proportionality of Procedural Obligations

The DFA must ensure that the substantive fairness duties to avoid manipulative design, coercive consent flows, and unfair personalisation apply universally and without exception. **These duties are grounded in the protection of fundamental rights and cannot be weakened on the basis of company size, market share, or organisational form.**

Proportionality must be understood in terms of regulatory support, not regulatory dilution. **Traders should not face 'lighter' obligations, but regulators should provide the means to comply: clear templates, model assessments, practical guidance, and sector-specific examples.** This reduces administrative friction while preserving the universality of the substantive rules.

This approach avoids a risk-based framework: the obligation to respect rights is the same for all, while the tools for demonstrating compliance can be standardised and scaled to make accountability feasible in practice. Proportionality here means ensuring accessibility of compliance, not lowering the bar for fundamental rights.

A Hybrid Structure: Integrating Consumer Law for the Digital Age

One of the main findings of the Fitness Check was that **consumer law is largely underused in meeting its intended objectives**[37]. To meet the scale, automation and

[37] Fitness Check Report.

systemic nature of risks and harms embedded in today's digital markets, it is important for the DFA to be a Regulation, following the example of the DSA. It must also impose binding obligations on traders and equip enforcement authorities with the tools needed to address systemic unfairness, rather than isolated infringements. **Making the DFA a Regulation brings with it the need for a coherent enforcement architecture. Without clarity on who enforces it and how, its transformative potential will remain limited.**

Incremental amendments to outdated instruments have failed to address the structural and behavioural asymmetries of digital markets. As shown by recent scholarship[38], piecemeal reforms risk reinforcing fragmentation and entrenching inequality. The DFA must offer more than a patchwork of obligations. It must operate as a horizontal safety net across all digital business-to-consumer (B2C) environments.

To achieve this, the DFA should be designed as a hybrid instrument: (1) partly amending existing Directives (the UCPD, UCTD, and CRD) to modernise their concepts of unfairness, vulnerability, and trader responsibility, while (2) also introducing self-standing, directly applicable provisions - such as fairness-by-design duties and rights not to be profiled by default - that fall squarely within the European Commission's enforcement role under the EU digital rulebook. This combined structure is necessary to ensure both consistency and effectiveness: a Regulation alone would lack integration with existing law, while isolated amendments would perpetuate fragmentation.

In addition, the burden often falls on individuals to recognise and report violations, despite the fact that consumer law is meant to address power asymmetries and cognitive biases that people cannot reasonably overcome on their own. Without meaningful enforcement reforms, including stronger coordination, design-level remedies that address manipulation at its structural source and proactive investigations, the promise of consumer protection in the digital age remains largely unfulfilled.

Last but not least, and to ensure legitimacy and rights-based policymaking, **the legislative process for the DFA should include fundamental rights scrutiny and structured civil society participation.** The Commission must conduct a formal fundamental rights impact assessment when preparing the proposal. Independent bodies such as FRA and the EDPS (but also potentially others) must provide opinions during negotiations. Civil society organisations must be directly involved throughout; not in one-off consultations, but as part of regular hearings, advisory fora, and co-creation of evidence. These mechanisms are essential to prevent fairness from being diluted under deregulatory pressure and to anchor the DFA in fundamental rights from the outset.

[38] Namysłowska, M., The Silent Death of EU Consumer Law and Its Resilient Revival: Reinventing Consumer Protection Against Unfair Digital Commercial Practices, Journal of Consumer Policy, 2025.

IV. Addictive Design and the Logic of Retention

Executive Summary

Addictive design refers to interface features, recommender systems, and other aspects of digital services that are deliberately optimised to encourage compulsive use or excessive engagement. These systems impair users' ability to control their time and attention by exploiting behavioural and emotional vulnerabilities through opaque optimisation, dynamic adaptation, and affective cues. Crucially, addictive design is not the result of a single deceptive interface choice, but the cumulative effect of multiple design features working in concert to maximise engagement and inhibit disengagement. These can include endless scroll, autoplay, variable rewards, persistent nudging, emotionally-framed prompts, and more – all reinforcing one another to reduce friction for continued use and increase resistance to opting out.

Addictive design has serious consequences for the enjoyment of fundamental rights, including human dignity, mental integrity, and democratic participation, with particularly severe impacts on children, marginalised users, and people in precarious situations. Users are often unaware of these dynamics and lack real alternatives. This undermines their autonomy and can lead to financial harm, especially in vulnerabilised contexts where psychological triggers are used to drive repeated spending, through mechanisms like loot boxes, gamified offers, or progression-based monetisation. These are not isolated flaws, but structural business strategies aimed at maximising engagement and data extraction, often reinforced by behavioural profiling and personalisation which violates people's rights to privacy and data protection.

Current Legal Gaps:

- The UCPD does not clearly prohibit persistent attention-manipulating systems.
- While the DSA increases transparency and mandates offering product options that are not based on profiling, it does not regulate how recommender systems exploit compulsive feedback loops.
- The GDPR regulates data processing, but not the core design architectures that impair autonomy and control, even in the absence of data misuse.

Key Policy Recommendations:

- Clarify through the UCPD general clause that digital practices which impair attentional autonomy, especially when designed to induce compulsive or excessive use, are to be treated as structurally unfair and incompatible with professional diligence under Articles 5–9 of the UCPD.
- Add features to the UCPD block list that are structurally manipulative by design, such as infinite scroll, autoplay without user controls, compulsive feedback loops, and loot boxes that combine variable-ratio rewards with monetisation or behavioural targeting. These features exploit user vulnerabilities and lack legitimate justification.

- Introduce a grey list of high-risk practices that are presumed unfair unless traders can demonstrate, through verifiable and auditable evidence, that they do not impair autonomy or induce compulsive use. These include streaks, emotionally framed re-engagement prompts, gamified pressure mechanisms, and non-monetised loot boxes or randomised rewards that mimic addictive dynamics.
- Enshrine a right not to be disturbed by default, requiring that push notifications, autoplay, and recommender systems are opt-in only and easy to turn off.
- Ban features that override user-set preferences, such as reactivating disabled settings or targeting boredom and distress to trigger re-engagement.
- Amend the Consumer Rights Directive (CRD) to mandate clear disclosures about engagement-optimised systems and allow users to switch off or modify them at any time.
- Shift regulatory focus to system-level behaviour, requiring pre- and post-market audits of compulsive design and emotional targeting.
- Treat addictive design as structurally exploitative, regulating not only visual tricks but the underlying data-driven logic of attention extraction.

By tackling addictive design as a systemic practice and therefore a risk to fundamental rights, rather than as a collection of user interface flaws, the DFA can restore attentional autonomy and protect people from manipulative digital environments.

'Addictive design' refers to elements of digital services that are intended to, or have the effect of, encouraging compulsive use or excessive engagement[39], in a manner that materially impairs people's ability to exercise time-aware, intentional, and autonomous use of digital systems. It encompasses not only to manipulative interface features, but also entire interaction infrastructures engineered to extract attention[40]. These practices are not trivial inconveniences: they undermine people's ability to exercise core rights such as autonomy, dignity, privacy, freedom of thought, and participation in society. By systematically overriding intentional disengagement, addictive design constrains how people use their time, how they encounter information, and how they relate to one another[41].

Such design often operates by **circumventing deliberate reflection**[42], and includes, though is not limited to, practices that:

[39] INI Report; see also Chauncey Neyman. 2017. A Survey of Addictive Software Design. 1, 1, Article 1 (June 2017). Some scholars refer to these practices as 'addicting design', emphasising the intentional strategies deployed by traders to trigger compulsive engagement and dependence through interface and system architecture. This term highlights the active role of design in producing addictive behaviours. In this paper, we opt for 'addictive design' as the more widely used formulation, which captures both the techniques and their outcomes while ensuring consistency with existing regulatory and policy discussions.

[40] Montag C, Lachmann B, Herrlich M, Zweig K. Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of Psychological and Economic Theories. Int J Environ Res Public Health. 2019 Jul 23;16(14):2612.

[41] Riccardo Chianella, 'Addictive digital experiences: the influence of artificial intelligence and more-thanhuman design, Conference Safe Harbors for Design Research 2011.

[42] Ilan Kilovaty, 'Legally Cognizable Manipulation' (2019) 34 Berkeley Technology Law Journal 449.

- Exploit psychological, neurological, or emotional vulnerabilities to prolong engagement[43];
- Use algorithmic systems, including recommender systems, to optimise for retention metrics such as screen time or interaction frequency rather than user benefit;
- Remove stopping cues or natural points of disengagement (e.g. infinite scroll, autoplay);
- Personalise content delivery or interaction loops based on profiling aimed at maximising usage regardless of the user's well-being or intention; and/or
- Undermine or circumvent user attempts to limit, pause, or disengage from the service.

Addictive design should be recognised as a systemic architecture, not an isolated flaw: it emerges from the interplay of multiple design features deliberately engineered to prolong use, suppress disengagement and deprive people of the ability to make rational and informed choices. A presumption of addictive design should apply where digital services, including interface elements or recommender systems, are tested or optimised to suppress disengagement. In such cases, the burden of proof must rest with the trader to show that this optimisation is strictly necessary for core functionality or user safety. Otherwise, the practice should be deemed unfair by default.

What Addictive Design Looks Like in Practice

To understand the real world implications of addictive design, we must move beyond abstract critiques and examine how specific design patterns are embedded into widely used digital services today, leveraging well-documented predictable cognitive and behavioural responses.[44].

Addictive design does not merely exploit attention, but reshapes what counts as a normal attentional state. As digital services push users toward constant reactivity, distraction becomes habitual and even expected, eroding the conditions for autonomy, reflection, and sustained focus[45]. **The goal is not merely to facilitate interaction or access, but to maximise user engagement, often without meaningful transparency about how personal data is used or how systems shape decisions and experiences on the user's behalf**[46].

---

43 Charman-Anderson, Suw. "Seeking Addiction: the Role of Dopamine in Social Media." Computer Weekly, 2009.

44 Anastasia Hronis, What Makes Us Keep Swiping?, University of Technology Sydney, 22 February 2024. https://www.uts.edu.au/news/2024/02/what-makes-us-keep-swiping

45 Aksoy, M.E. (2018). A Qualitative Study on the Reasons for Social Media Addiction. European Journal Of Educational Research, 7(4), 861-865.

46 Aimen Taimur, Manipulative Matchmaking - A Legal and Ethical Assessment of Addictive AI Design in Dating Apps Rivista di Scienze Giuridiche, Scienze Cognitive ed Intelligenza Artificiale. Vol. 18 n. 1 (2025).

Below are illustrative examples across several sectors that demonstrate how addictive design works in practice:

- **Infinite scroll, autoplay, and emotional loops:** services like YouTube, TikTok, and Instagram Reels are built around frictionless content consumption. Autoplay ensures that videos roll one into the next with no user input. Infinite scroll removes stopping cues, allowing users to watch for extended periods without conscious decision-making. In terms of fundamental rights, these design strategies limit freedom of thought and access to information by conditioning exposure. **These systems are not simply content delivery tools: they are oftentimes driven by recommender systems trained on granular behavioural profiles about every single user in order to serve whatever is most likely to prolong user attention.** For example, users searching for physical exercise tips, anxiety relief, or parenting advice are often funnelled toward more extreme, divisive, emotionally charged, or obsessive content loops that maximise engagement while reinforcing affective dependencies.
- **'Dopamine' loops and validation feedback on social media platforms:** services like Facebook, TikTok, and Snapchat optimise for social reward feedback. Notification systems are fine-tuned to deliver 'likes', comments, and other interactions in patterns that reinforce return visits[47]. Pervasive tracking and profiling allows services to personalise when and how these notifications are delivered, for instance by batching them or delaying them to coincide with known vulnerability windows (e.g. late at night; after a period of inactivity). These techniques create **dopamine-driven loops that deepen compulsive use, especially effective among teenagers and people seeking validation**[48]. Platform-led 'screen time' tools often place the burden on users - including children - to resist systems engineered to bypass their self-regulation. By targeting notifications to vulnerability windows, platforms structurally erode autonomy and dignity, and in the case of children and teenagers, amount to a direct violation of the right to special protection. These prompts rarely interrupt compulsion effectively and lack meaningful health warnings.
- **Recommender systems further amplify this by surfacing content designed to provoke a reaction, increasing both use and anxiety, shaping discourse, polarisation, and mental health, and undermining equal participation in the public sphere**. While reinforcement mechanisms are not inherently harmful (many apps, including language learning or meditation apps, use them to support self-directed goals), in the context of surveillance-driven business models[49] these same techniques are routinely deployed not to support user choice and wellbeing, but to override it. Recommender systems amplify this

[47] Alberto Monge Roffarello and Luigi De Russis. 2022. Towards Understanding the Dark Patterns That Steal Our Attention. In Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22). Association for Computing Machinery, New York, NY, USA, Article 274, 1–7. https://doi.org/10.1145/3491101.3519829

[48] Ana da Silva Pinho et al., Youths' sensitivity to social media feedback: A computational account. Sci. Adv. 10, eadp8775 (2024).

[49] EDRi (2021) Booklet: Surveillance-based advertising: An industry broken by design and by default, March 9, 2021. https://edri.org/our-work/surveillance-based-advertising-an-industry-broken-by-design-and-by-default

further by surfacing emotionally provocative content, increasing both time spent and emotional volatility. These mechanisms also pose a direct threat to democratic discourse by funnelling people into echo chambers, delivering content that reinforces increasingly extreme beliefs, and determining the parameters of the information that people do or do not see.
- Personalised praise or emotional cues from Large Language Models (AI agents).
- Community mechanisms that reinforce daily or frequent use, such as streaks, leaderboards, group challenges, or notifications that exploit social pressure to keep people engaged.

Why this Issue Matters

Addictive design is no longer a fringe concern, it is **central to how dominant digital services are built, monetised, and experienced**. **Digital services are designed not simply to attract users but to maximise their screen time, to generate longer engagement, and to increase their exposure to advertising and data extraction**[50]. It's critical to stress that addictive design is not the result of a single dark pattern or interface choice, but the cumulative effect of multiple design features working in concert to create compulsive engagement and inhibit disengagement.

Decades of research have shown how certain features of digital services exploit human cognitive biases. They are techniques that form part of a design toolkit intentionally deployed to sustain engagement and normalise compulsive use, thereby undermining people's ability to exercise freedom of thought, informational self-determination, and meaningful consent. These strategies, documented extensively in behavioural research and research into human-computer interaction (HCI), are now embedded in many digital choice environments[51].

Many addictive design patterns exploit emotionally charged contexts, such as intimacy, loneliness, or the desire for connection, to intensify user engagement. **The degree of exposure varies depending on how digital infrastructures intersect with emotional, social, and economic precarity, producing greater harm in contexts already marked by exclusion or dependence**. These vulnerabilities are particularly pronounced in services like dating apps, but the same techniques are increasingly replicated in social media, video platforms, and e-commerce[52]. This is not a technical by-product of innovation, nor merely a matter of user convenience. It is a commercial strategy deliberately engineered to exploit users' cognitive and emotional vulnerabilities in

[50] The Guardian, "Ex-Facebook president Sean Parker: site made to exploit human 'vulnerability'" 9 November 2017 https://www.theguardian.com/technology/2017/nov/09/facebook-sean-parker-vulnerability-brain-psychology; Ryan Calo, Digital Market Manipulation, 82 Geo. Wash. L. Rev. 995 (2014).

[51] Chauncey Neyman. 2017. A Survey of Addictive Software Design. 1, 1, Article 1 (June 2017).

[52] Aimen Taimur, Manipulative Matchmaking - A Legal and Ethical Assessment of Addictive AI Design in Dating Apps Rivista di Scienze Giuridiche, Scienze Cognitive ed Intelligenza Artificiale. Vol. 18 n. 1 (2025).

order to maximise engagement, extract more personal data, and drive revenue through prolonged exposure to advertising, in-app purchases, or other monetisation channels.

It is important to clarify that **addictive design is not limited to Big Tech platforms: many digital services use these techniques to maximise engagement**. However, in dominant ecosystems, addictive design becomes a competitive strategy: it deepens interface lock-in, extends user surveillance, and consolidates data power, reinforcing structural asymmetries in the digital economy[53]. Even when a more ethical, fee-based competitor enters the market, it struggles to compete due to network effects and user inertia. This dynamic entrenches monopolistic structures and narrows the conditions under which people can exercise their rights online.

At the heart of addictive design lies a **predictive surveillance logic, in which behavioural data are continuously collected, analysed, and fed back into algorithmic systems in order to refine user retention strategies without their awareness.** Every scroll, pause, reaction, or hesitation is mined for meaning, producing increasingly granular models of user behaviour. Such practices directly implicate the right to privacy, data protection, freedom of thought, and access to information, rights that are impossible to exercise meaningfully under conditions of constant behavioural governance and manipulation.

Courts in the U.S. are increasingly treating addictive design as a product safety issue but this framing misses the deeper point: what is at stake is not safety alone but the protection of fundamental rights and the harms that stem from structural design logic. **The close coupling of tracking, profiling, and unfair personalisation remains a missing link in many policy debates.** Design addiction is not solely a question of visual cues or persuasive interfaces. It is inseparable from the surveillance infrastructure that makes it possible to personalise user retention strategies at scale, in real time, and with behavioural precision. Yet consumer protection law rarely addresses this link and instead treats addictive design as a surface-level interaction, rather than a data-driven process of behavioural governance.

The gaming industry illustrates how addictive design and monetisation converge: microtransactions, loot boxes[54], and time-limited offers are often structured to exploit compulsive engagement loops. A review analysis of top-grossing games shows that players frequently describe these features as manipulative or psychologically coercive, even when no formal deception is involved[55].

---

53 Nie, M. (2025). Algorithmic Addiction by Design: Big Tech's Leverage of Dark Patterns to Maintain Market Dominance and its Challenge for Content Moderation. arXiv preprint arXiv:2505.00054.

54 Norwegian Consumer Council. INSERT COIN. How the gaming industry exploits consumers using loot boxes. May 2022 https://storage02.forbrukerradet.no/media/2022/05/2022-05-31-insert-coin-publish.pdf

55 Elena Petrovskaya, Sebastian Deterding, and David I Zendle. 2022. Prevalence and Salience of Problematic Microtransactions in Top-Grossing Mobile and PC Games: A Content Analysis of User Reviews. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 560, 1–12.

Critically, **consent alone does not provide adequate protection against addictive design**. Users may nominally agree to using a service, but they rarely understand or have meaningful control over the design logic shaping their experience[56]. Design elements that retain users, such as frictionless transitions, looping interfaces, and strategic colour or sound triggers, are not disclosed as mechanisms of influence, nor are the underlying algorithms. **While users may appear to engage voluntarily, addictive design patterns often structure choice through emotional pressure, constrained options, and asymmetries in information.** The result is not autonomous action, but habituated, nudged behaviour that reinforces traders' interests over user agency. Even users aware of these techniques may lack meaningful alternatives, especially in markets dominated by a handful of large digital services traders. When optimisation prioritises user retention over human agency, recommender systems become tools of behavioural manipulation that undermine fundamental rights, regardless of their apparent convenience.

Scholars rightly argue that addictive design should be considered a form of deceptive design, with terms such as 'attention-capture dark pattern'[57] or 'hyper-engaging dark pattern'[58]. Like deception, addictive design is intentional, manipulative, and exploits cognitive vulnerabilities to produce user harm. However, in this background paper we distinguish between the two for both analytical and legal reasons.

- Deceptive design is typically framed around discrete transactional manipulation: design choices that coerce, deceive or subvert immediate user decisions, such as consenting, purchasing, or subscribing.
- Addictive design, by contrast, refers to temporally-extended manipulation: systems engineered to foster compulsive use and reduce self-regulation over time.

While addictive design meets the functional criteria of deceptive design, its operation across time, its entanglement with personalisation systems, and its public health implications, make it useful to treat as a separate, though overlapping, category. This distinction reflects the way EU policy instruments currently address them, and allows for a more targeted discussion of structural risks to rights that extend beyond isolated interface choices.

Many addictive systems do not rely on profiling at all. Instead, they use emotional and aesthetic design to produce compulsive engagement. For instance, popular educational apps and games often deploy stylised characters, gamified praise, or anthropomorphic feedback (e.g. sad faces when a user exits) to foster emotional attachment and discourage disengagement. These aesthetic-affective strategies are designed to bypass reflective choice, exploiting developmental or emotional attachment to generate habituation, especially in environments where design replaces human interaction with programmable affect. These design patterns are particularly

[56] Nita Farahany, The Battle for Your Brain: The Right to Think Freely in the Age of Neurotechnology (2023).

[57] Xin Ye (2025) Dark Patterns and Addictive Designs, Weizenbaum Journal Of The Digital Society 03; Volume 5 \ Issue 3.

[58] Esposito F, Maciel Cathoud Ferreira T. Addictive Design as an Unfair Commercial Practice: The Case of Hyper-Engaging Dark Patterns. European Journal of Risk Regulation. 2024;15(4):999-1016.

prevalent in environments directed at children, but their underlying logic affects all users. **When systems are designed to work precisely because they trigger developmental vulnerabilities or exploit affective attachment, they function as behavioural traps.** These dynamics demonstrate that manipulation can be embedded in the service itself – not just in data practices – and must be regulated as such[59].

It's therefore critical to stress that **addictive design does not depend on the existence of recommender systems.** While profiling-based curation often intensifies compulsive use, many addictive features, such as infinite scroll, autoplay, push notifications, or streak counters, function without personalisation. These mechanisms exploit attention and limit self-regulation through digital service design alone. A recent cross-platform study mapped 13 distinct design strategies and 37 interface features that can contribute to addictive use, from reward systems and visual cues to attention-capturing notifications and emotionally reinforcing interactions[60]. The analysis confirms that addictive design is not limited to recommender systems or social media feeds: nearly all user-facing features, depending on how they are structured, can be optimised to extract more time, attention, and emotional dependency. This reinforces the need for consumer protection rules that apply across the full digital service, not just to profiling or personalisation. Therefore, to be effective, regulation, and especially the DFA, must target addictive design patterns broadly, not just systems using AI or behavioural profiling using personal data.

Addictive Design and Recommender Systems

Addictive design is not limited to interface features: it also concerns the nature of the content being promoted. Recommender systems often amplify emotionally-charged, low-quality, or misleading user-generated content precisely because it is more likely to provoke compulsive engagement, regardless of informational value or wellbeing. This dynamic reinforces the addictive architecture of the system, turning content selection itself into a tool of manipulation. A 2023 Amnesty International report showed how TikTok's recommender system leverages personal data to fuel addictive and potentially harmful content loops, including disproportionate exposure to distressing content by vulnerable users[61].

Profiling-based recommender systems have become one of the main operational backbones of addictive design, even though they are not technically necessary: they play a central role in reinforcing addictive engagement loops[62]. **Under the guise of offering more relevant or desirable content, and while presented as tools for**

[59] Catherine Pescott, 'Children, Young People and Online Harm: An Overview', in Faith Gordon and Daniel Thomas (eds), Children, Young People and Online Harms (Bristol University Press 2024), ch 3.

[60] Granda, M.F., Sarmiento, MB., Nuñez, AG., Maldonado, R., Parra, O. (2025). Developing a Design Features Taxonomy of Human-Computer Interaction in Social Media that Affect User Engagement and Addictive Behaviors. In: Grabis, J., Vos, T.E.J., Escalona, M.J., Pastor, O. (eds) Research Challenges in Information Science. RCIS 2025. Lecture Notes in Business Information Processing, vol 547. Springer, Cham.

[61] Amnesty International (2023). "I feel exposed": Caught in TikTok's surveillance web. November 2023. https://www.amnesty.org/en/documents/pol40/7446/2023/en

relevance or personalisation, most are optimised to maximise engagement[63], typically defined in terms of observable metrics such as clicks, watch time, or scrolling velocity. **This optimisation relies not only on user behaviour within the service, but also on extensive profiling built from data collected across websites, apps, devices, and even purchased from third parties**. As users interact, these systems continuously experiment, iteratively refining what is shown, when, and how, based on shifting signals of what will provoke another tap, another scroll, another captured moment of attention.

In doing so, recommender systems are not merely personalising (see Chapter VII on Unfair Personalisation below), they are perpetually **curating environments of engineered compulsion**. For example, Amnesty International's report showed how a newly created TikTok account, set up to mimic a 13-year-old engaging with mental health content, was quickly flooded with videos that appeared to normalise or romanticise suicide[64]. This case exemplifies how these systems rapidly lock users into emotionally-charged content loops, combining prolonged exposure with extreme personalisation to heighten vulnerability rather than alleviate it.

Whether in social media feeds, on video sharing platforms, in online retail, or on streaming services, recommender systems often create behavioural loops that users struggle to exit not because of informed choice, but because the systems are fine-tuned to keep attention, not to support reflection or autonomy. While it's true that some recommender systems enhance usability and discovery, especially in environments like streaming platforms where content is not user-generated, the core issue lies in the goal of optimisation. **When optimisation focuses on time spent, frictionless flow, or predicted vulnerability rather than user agency, even seemingly helpful features become tools of disproportionate manipulation.**

Structural Rights Interferences of Addictive Design

Addictive design is not a marginal phenomenon but a defining feature of dominant digital services. By deliberately engineering environments that capture and retain attention, addictive design systematically interferes with people's ability to exercise their fundamental rights including autonomy, dignity, freedom of thought, and democratic participation. These practices transform digital environments into infrastructures of governance where behaviour is shaped, anticipated, and monetised through design logics optimised for compulsion rather than agency.

Addictive design cannot be understood as persuasion alone. When design features are powered by behavioural data, profiling, and predictive optimisation, they operate as

62 Context, "Tech platforms must drop addictive features that harm young people", EDRi 17 April 2024. https://edri.org/our-work/exploitative-designs-how-tech-platforms-harm-young-people/

63 Neil Richards & Woodrow Hartzog, Against Engagement, 104 B.U. L. Rev. 1151 (2024).

64 Amnesty International, "DrIven Into The Darkness. How TikTok encourages self-harm and suicidal ideation", 7 November 2023. https://www.amnesty.org/en/latest/news/2023/11/tiktok-risks-pushing-children-towards-harmful-content

coercive environments. **People do not encounter neutral interfaces, but carefully constructed systems that curtail their ability to act freely and knowingly.**

Autonomy and Self-Determination: Design against intention[65]

According to the Commission's Fitness Check, gamification techniques and other design features can create compulsive usage patterns and undermine consumers' ability to make autonomous choices[66]. Interviews with users reveal a common pattern of compulsive re-engagement: they describe reopening platforms like Instagram reflexively, sometimes seconds after closing them[67], reinforcing the loss of self-directed use. **These mechanisms replace decision-making with habituated responses, making choices for people rather than with them.**

Addictive design bypasses conscious decision-making through tactics like variable rewards (unpredictable pay-offs given after an uncertain number of actions, a mechanism well known from gambling)**, infinite scroll, and push notifications. These exploit attentional and affective vulnerabilities to provoke compulsive use, not deliberate choice**. Recommender systems often rely on opaque affective profiling, using subtle behavioural signals to infer moods or desires, yet rarely disclose the basis of these inferences or the logic driving system decisions. Even with disclosure, exploiting affective states to steer behaviour raises profound concerns: it risks manipulation rather than empowerment. At most, design should support people's self-defined goals (e.g. a learning app reminding users to practise), not use hidden emotional cues to trigger compulsive use. This leads to **decisions made for users, not with them, undermining both autonomy and transparency**[68].

Such manipulation distorts user agency in ways that would be impermissible offline[69]. Time and attention are finite resources for every person. The more they are extracted by design, the less space remains for critical reflection, off-screen life, or freely chosen alternative online activity. **Addictive design contributes to what scholars call 'time poverty'[70]: the loss of self-directed time caused by environments structured to override intention**[71]. It is a form of harm that is individually felt but structurally

65 Koopmans, F., & Sremac, S. (2011). Addiction and autonomy: Are addicts autonomous? Nova prisutnost: časopis za intelektualna i duhovna pitanja, 9, 171 – 185.

66 Fitness Check Report.

67 Cao, X., Gong, M., Yu, L., & Dai, B. (2020). Exploring the mechanism of social media addiction: an empirical study from WeChat users. Internet Research, 1305-1328.

68 Aimen Taimur, Manipulative Matchmaking - A Legal and Ethical Assessment of Addictive AI Design in Dating Apps Rivista di Scienze Giuridiche, Scienze Cognitive ed Intelligenza Artificiale. Vol. 18 n. 1 (2025).

69 Bhargava, V. R., & Velasquez, M. (2021). Ethics of the attention economy: The problem of social media addiction. Business Ethics Quarterly, 31(3), 321 – 342.

70 Schoch, Manfred and Weinert, Christoph, "IT-Related Time Poverty: Identifying Antecedents and Consequences of a Lack of Time Related to IT Use" (2023). Wirtschaftsinformatik 2023 Proceedings. 69; Ding K, Shen Y, Liu Q, Li H. The Effects of Digital Addiction on Brain Function and Structure of Children and Adolescents: A Scoping Review. Healthcare (Basel). 2023 Dec 20;12(1):15.

71 Statista, "Time Spent Online Worldwide by Region 2024", 8 March 2024. https://www.statista.com/statistics/1258232/daily-time-spent-online-worldwide/#statisticContainer; Watzl, S. (2023). What

produced[72]. **The result is not just distraction, but a breakdown in the conditions for autonomy: reflection, awareness, and the possibility of choosing otherwise.**

Neurological and Mental Health: Structural Impacts on Wellbeing[73]

Mounting research links addictive design to deteriorating mental health[74]. **The harm goes beyond screen time or distraction: it includes measurable psychological and physiological consequences.** Addictive systems intensify compulsive use, amplify anxiety and social comparison, and disrupt circadian rhythms and sleep quality. Emerging studies suggest that rapid, context-shifting environments, such as short-form video feeds, significantly degrade prospective memory, reducing the user's ability to remember tasks or intentions even after brief exposure[75]. Studies also show that **even brief interruptions to self-regulation caused by addictive features can produce guilt, irritability, and fatigue[76].**

Neuroscientific research has **linked early exposure to compulsive design environments with reductions in grey matter volume, impaired prefrontal cortex development, and ADHD-like symptoms**[77], especially among adolescents. While such effects are intensified for groups already facing social or economic disadvantage, they reflect a structural risk for all users exposed to high-friction, exploitative design systems without meaningful safeguards or off-switches[78]. Addictive design reduces people's ability to exercise control over their own mental integrity. Such interference

attention is: The priority structure account. Wiley Interdisciplinary Reviews: Cognitive Science, 14(5), Article e1632. https://doi.org/10.1002/wcs.1632

72 Gaia Bernstein, Unwired: Gaining Control Over Addictive Technologies (2023).

73 World Health Organization. (2018, September 13). Public health implications of excessive use of the internet and other communication and gaming platforms. World Health Organization. https://www.who.int/news/item/13-09-2018-public-health-implications-of-excessive-use-of-the-internet-and-other-communication-and-gaming-platforms; Word Health Organization. (n.d.). Addictive Behaviours: Gaming Disorder. World Health Organization. https://www.who.int/news-room/questions-and-answers/item/addictive-behaviours-gaming-disorder

74 Matt Lawrence, Public Health Law's Digital Frontier: Addictive Design, Section 230, and the Freedom of Speech, 3 Free Speech L. 299 (2023); Ujala Zubair, Muhammad K. Khan, Muna Albashari, Link between excessive social media use and psychiatric disorders, Annals of Medicine & Surgery (2023) 85:875–878.

75 Will Moore, "Mindless Scrolling: The Science Behind Why It's So Addictive" 29 November 2024. https://mooremomentum.com/blog/what-is-mindless-scrolling-and-the-science-behind-why-its-so-addictive

76 MUJICA, Alejandro L. et al. ADDICTION BY DESIGN: Some Dimensions and Challenges of Excessive Social Media Use. Medical Research Archives, [S.l.], v. 10, n. 2, February 2022. ISSN 2375-1924; European Parliamentary Research Service, Harmful internet use. Part I: Internet addiction and problematic use, January 2019. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/ 624249/EPRS_STU(2019)624249_EN.pdf; Kumar M, Mondal A. A study on Internet addiction and its relation to psychopathology and self-esteem among college students. Ind Psychiatry J. 2018 Jan-Jun;27(1):61-66. doi: 10.4103/ipj.ipj_61_17.

77 BBC, "Web addicts have brain changes, research suggests", 12 January 2012. https://www.bbc.com/news/health-16505521

78 Zubair U, Khan MK, Albashari M. Link between excessive social media use and psychiatric disorders. Ann Med Surg (Lond). 2023 Mar 27;85(4):875-878.

directly implicates human dignity as recognised in EU and international human rights law.

These risks are further amplified when addictive design elements mimic gambling mechanics. Research links loot boxes and other gambling-like features, such as prize wheels or social casino games, to increased impulsivity, anxiety, and problematic gambling behaviours, particularly in adolescents. Simulated gambling in games may act as a gateway to monetary gambling, reinforcing cycles of psychological harm in vulnerable users.[79]

Democratic Participation: Extractive Design Erodes Civic Capacity

When attention is persistently diverted and behaviour subtly shaped by opaque systems, people have less time and energy to engage with political content critically, or to engage at all. **Algorithmic design choices that favour outrage, repetition, and virality also crowd out deliberation and reinforce polarisation**[80]. In this way, addictive design undermines the informational and attentional conditions required for meaningful democratic participation. Structural time poverty becomes not just a personal cost, but a civic one.

Unequal Exposure, Structural Effects

While all people are exposed to addictive features, the degree and consequences of exposure vary. Children and adolescents are particularly at risk, given their developing self-regulation capacities and susceptibility to social validation loops[81]. But adults, too, are affected, especially those in precarious life circumstances, for whom digital services may offer temporary escape while simultaneously deepening dependency. This demonstrates that addictive design is not merely about individual traits, but about how digital systems reproduce and intensify structural inequalities.

---

79 Villalba-García et al., The relationship between loot box buying, gambling, internet gaming, and mental health: Investigating the moderating effect of impulsivity, depression, anxiety, and stress, May 2025, https://doi.org/10.1016/j.chb.2025.108579; Grosemans et al., More than loot boxes: the role of video game streams and gambling-like elements in the gaming-gambling connection among adolescents, January 2024, https://www.researchgate.net/publication/379514259_More_than_loot_boxes_the_role_of_video_game_streams_and_gambling-like_elements_in_the_gaming-gambling_connection_among_adolescents.

80 Colomina, C., Sanchez Margalef, H., & Young, R. (2021). The impact of disinformation on democratic processes and human rights in the world. European Parliament. https://www.europarl.europa.eu/thinktank/en/document.html

81 Nancy Costello et. al., Algorithms, Addiction, and Adolescent Mental Health: An Interdisciplinary Study to Inform State-Level Policy Action to Protect Youth from the Dangers of Social Media, 49 Am. J Law Med. 135 (2023); People vs. Big Tech (2024) 07.11.24 Briefing: protecting children and young people from addictive design. https://peoplevsbig.tech/briefing-protecting-children-and-young-people-from-addictive-design

Why Existing Rules are Not Working

Some scholars argue that addictive design already falls within the scope of the UCPD[82]: these arguments point to how prolonged manipulation of time and attention undermines freedom of choice and erodes individual autonomy. However, in practice, **the UCPD has proven ill-equipped to address such systemic threats.** Its current framing was not designed to capture behavioural profiling, dynamic interface experimentation, or compulsive loops. Addressing addictive design as a form of structural unfairness requires targeted reform: new presumptions, default-off obligations, and a reversal of the burden of proof to better reflect the profound asymmetries of the digital environment and to safeguard fundamental rights.

Likewise and as mentioned above, current interpretations of consumer vulnerability rely on individual characteristics, rather than acknowledging the structural asymmetries that define digital environments. **Addiction-by-design is not about exceptional cases, but about digital services features that systematically undermine user autonomy at scale.** Techniques such as infinite scroll, autoplay, and algorithmic content loops are not marginal quirks they are core business strategies optimised for user retention, and thus fall outside the scope of existing laws unless reframed as inherently unfair.

While the DSA and DMA contain provisions against certain manipulative interface practices, their scope is limited. Article 25 DSA, for instance, only applies to online platforms and marketplaces (more on this in Chapter V on deceptive design below), and does not address the specific dynamics of addictive design, in particular its use as a business model to extract attention and data. These gaps point to the need for legislation that tackles not only deceptive tactics, but the structural incentives behind compulsive design.

While the DSA introduces obligations around transparency and some limited choice in recommender systems, notably Article 26 for all platforms and Article 38 for Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs), these provisions focus on information disclosure (e.g. explaining main parameters) and a very narrow user choice obligation for recommender systems that are not based on profiling. They do not regulate how profiling-based recommender systems are optimised, nor do they prohibit design patterns that exploit compulsive feedback loops, emotional volatility, or user fatigue.

What is more, most recommender systems remain so-called 'black boxes'[83], tested at scale to maximise engagement, but with no assessment as to how they influence user autonomy, mental health, or self-regulation. The DSA's emphasis on platform size and

[82] Esposito F, Maciel Cathoud Ferreira T. Addictive Design as an Unfair Commercial Practice: The Case of Hyper-Engaging Dark Patterns. European Journal of Risk Regulation. 2024;15(4):999-1016.

[83] Panoptykon Foundation (2020) Black-Boxed Politics: Opacity is a Choice in AI Systems, 17 January 2020.

information asymmetries leaves another regulatory gap: design features that induce addiction through structurally unfair feedback architectures, even without violating data protection or transparency rules. These systems are not merely personalised: they are behaviourally curated to reward compulsive engagement, and must be regulated as such.

Similarly, **while the GDPR plays a vital role in regulating personal data, it offers limited protection against addictive design, unless that design results in unlawful processing**. As mentioned above, even when profiling is used to optimise engagement, enforcement tends to focus narrowly on the validity of consent or the transparency of data use. It does not address how digital services architecture itself can induce compulsive use through patterns that remain largely invisible to individuals. Many of the most harmful techniques, such as infinite scroll, personalised notifications, or behavioural loops, operate below the threshold of conscious awareness, bypassing informed decision-making and undermining autonomy without *prima facie* violating data protection rules.

As a result, **there is currently no legal instrument in the EU toolkit that systematically addresses the cumulative harms and structural threats posed by design architectures that are optimised for dependency and compulsion**. This gap makes it difficult to sanction traders who prioritise engagement metrics over user well-being, or to impose structural limitations on attention-maximising systems. A new legal standard is needed to define, identify, and prohibit such exploitative strategies, and to rebalance power towards the user in ways that safeguard fundamental rights and ensure fairness by design.

Proposed Policy Changes to Address Addictive Design

This phenomenon cannot be reduced to interface tricks or behavioural psychology alone: it emerges from a convergence of business models, sociotechnical infrastructures, and ethical or cultural framings. A systemic lens supports the argument that regulatory interventions, particularly under consumer law, must address the full architecture and incentive logic behind addictive experiences, not merely their surface features.

As explained above, while current law addresses deception, coercion, and unfair influence in abstract terms, it has not yet evolved to recognise addictive design as a systemic practice of consumer exploitation, nor to regulate the role of personalisation and recommender systems in shaping user experience. **The DFA provides an opportunity to close this gap, define when persuasive design becomes exploitative, impose boundaries on attention extraction, and introduce safeguards that protect users not only from what they see, but from how and why they are made to see it.**

Such regulatory progress is long overdue. It requires **treating attention as a dimension of consumer vulnerability, not as an infinite commodity.** It also means recognising that recommender systems and the personalisation they enable are not neutral tools but behavioural engines whose design choices reflect commercial imperatives rather

than public interest. Fairness in this context must be more than a procedural notion, it must be substantive, structural, and enforceable.

As mentioned, addictive design undermines not only user autonomy in a given moment, but also users' long-term ability to exercise self-regulation. **When services systematically override user-set preferences such as reactivating disabled notifications or bypassing time limits, they disrupt individuals' attempts to disengage or moderate use.** These techniques weaken behavioural boundaries over time, contributing to compulsive engagement.

The following measures are tailored to address addictive design practices. While some safeguards (such as auditability or regulator powers) should apply across all deceptive design, these measures specifically target the structural drivers of digital addiction. General safeguards that are equally relevant to deceptive design have been moved to the following chapter to avoid duplication and ensure coherence.

Because developer intent is often invisible or denied, **addictive design must be regulated based on observable techniques and real-world effects.** Enforcement should target compulsive feedback structures, usage patterns, and known manipulative design choices. Regulation must intervene at both the development and deployment stages: **pre-market controls should prohibit known addictive techniques, while post-market oversight must ensure accountability where design practices undermine agency, fairness, or the effective exercise of fundamental rights.** As mentioned above, by shifting the burden of responsibility onto traders, the DFA can help dismantle dependency loops and restore baseline conditions for attentional autonomy.

Turning Off by Default: Systemic Safeguards Against Compulsive Engagement

To address the systemic nature of attention extraction, and in line with the European Parliament's position[84], **the DFA should enshrine a right not to be disturbed by default.** This principle would ensure that engagement-optimised features such as autoplay, push notifications, alerts, or profiling-based content suggestions are disabled by default and activated only through informed, unbundled, and reversible user choice. Implementing this right means promoting rights-respecting design by default, including up-front opt-ins for potentially addictive features, greyscale modes, and time-awareness nudges. These measures help reduce dependency-driven engagement loops and rebalance user agency. This right should apply regardless of business model or design of the digital service, and must be enforceable across systems that induce engagement through profiling, reward anticipation, or coercive interface loops.

Amend the UCPD's General Clause

To address addictive design, the DFA should also introduce targeted amendments across core EU consumer protection instruments. >Amending Article 5 UCPD to expressly mention these practices, and complementing it with both an expanded block

[84] INI Report.

list and a new grey list would provide much-needed clarity for enforcers and a stronger legal basis for timely intervention. **This dual-list approach would enable regulators to prohibit inherently harmful design practices while also flagging emerging manipulative techniques that warrant closer scrutiny.**

Block List: Prohibiting Structurally Manipulative Features

Under the UCPD, then, **the DFA should propose that features such as infinite scroll, autoplay without user controls, and compulsive feedback loops be added to the block list of prohibited commercial practices**. These design elements are structurally manipulative: they are not designed to inform or assist users but to discourage disengagement and exploit behavioural vulnerabilities, with no legitimate justification.

Special attention should also be given to A/B testing and real-time behavioural experimentation[85]**.** While these techniques can serve legitimate purposes such as improving usability or accessibility, they can also be misused to manipulate user behaviour by exploiting emotional responses, cognitive load, or situational vulnerabilities. Where such testing or dynamic adjustment is used to influence engagement, attention, or spending rather than to enhance user welfare, it should be subject to strict transparency, accountability, and audit requirements. Traders should be required to disclose when and how behavioural experiments are conducted, the parameters being tested, and their potential impact on people's autonomy and well-being.

To prevent circumvention, **the law must also prohibit indirect reactivation of disengagement choices via nudges, deceptive prompts, or default resets**. This includes features that systematically override user intention or disrupt self-regulation, such as:

- Persistent interruption of user-set time limits,
- Default reactivation of previously switched-off features (like notifications or autoplay),
- Behavioural predictions (e.g. boredom, sadness) used to trigger re-engagement, or
- The absence of effective 'stopping cues'.

Beyond classic deception, the UCPD must be updated to address manipulation aimed at monetisation, such as artificially-induced scarcity, friction in free use, and behavioural prompts tied to spending pressure. As seen in gaming, these designs exploit compulsive loops to drive revenue rather than informed choice. This includes monetisation through variable-ratio reward systems (a technique borrowed from gambling, where rewards are given unpredictably after an uncertain number of actions, making people repeat the behaviour in the hope of a payoff) that create

[85] A/B testing refers to the process of comparing multiple versions of a design element (e.g. a button or message) to determine which variation performs best. Users are unknowingly assigned different versions, and their behaviour (e.g. clicks, time spent, or conversion rates) is tracked to assess impact. Real-time behavioural experimentation goes further: it involves dynamic, algorithmic adjustments to interface elements or content during a user session - based on live interaction data, emotional states, or inferred behavioural patterns - with the aim of maximising engagement, conversion, or other business outcomes.

addictive spending incentives, particularly when linked to in-game currencies or psychologically timed offers.

Loot boxes should be block-listed when they use variable-ratio reward mechanisms linked to monetisation, especially when: they obscure the real-money cost through indirect currencies, they are triggered by behavioural profiling (e.g. loss, boredom), or they target structurally vulnerable users, including children and those with impulsivity profiles. This form of loot box design is structurally manipulative and engineered to induce compulsive spending, not user enjoyment.

Grey List: Presumed Unfair Practices

In addition, **the DFA should introduce a grey list of presumptively unfair design strategies, including:**

- **Gamified pressure mechanisms** such as streaks, progress-based rewards,
- **Emotionally framed re-engagement prompts,** especially those triggered by behavioural or emotional profiling,
- **Engagement-optimised recommender systems,** particularly those designed to extend usage duration by exploiting known psychological biases,
- **Progression-based monetisation nudges,** such as 'buy now to skip wait time' offers,
- **Simulated urgency cues,** including countdown timers or pseudo-scarcity banners ('only 2 left!') that are not based on factual availability.

Grey-listed practices may include gamified pressure mechanisms such as loot boxes that do not involve monetisation or behavioural targeting, but still introduce compulsive use patterns. Where loot boxes rely on transparency, clear probability disclosure, and are not tied to real-money currencies, they may warrant contextual assessment. Otherwise, monetised and variable-ratio loot box mechanics should be categorically prohibited.

These practices should be presumed unfair unless traders can demonstrate through verifiable, auditable evidence that they do not exploit user vulnerabilities, induce compulsive use, or undermine autonomy.

A recent academic proposal offers a structured enforcement tool which classifies addictive patterns into four categories: forced action, social engineering, interface interference, and persistence[86]. **This taxonomy could enable regulators to detect structural manipulation even where no explicit deception occurs, and could serve as a baseline for establishing new presumptions of unfairness under the DFA**.

Additional Measures: Beyond the UCPD

The CRD should be amended to require **clear pre-contractual information about the use of engagement-optimised recommender systems** and to **give users the ability to modify them or turn them off**.

[86] M. Beltrán, "Defining, Classifying and Identifying Addictive Patterns in Digital Products," in IEEE Transactions on Technology and Society, vol. 6, no. 3, pp. 314-323, Sept. 2025.

V. Deceptive Design and the Limits of Current Protections

Executive Summary

Deceptive design refers to interface and interaction strategies that distort or impair users' ability to make free and informed decisions. These tactics exploit cognitive and emotional vulnerabilities, use misleading defaults, create friction around refusal, and personalise manipulative flows through profiling. **They are not isolated tricks, but systemic, data-driven strategies embedded in commercial business models**.

Current Legal Gaps:

Although the DSA's Article 25 recognises and prohibits some forms of deceptive design, it only covers online platforms, a narrow subset of providers engaged in the deployment of deceptive design patterns, and fails to address dynamic, adaptive forms of manipulation powered by recommender systems, behavioural testing, and profiling.

While the GDPR requires consent to be freely given, informed, and specific, it does not regulate how the design of digital services can undermine these conditions. It remains silent on visual nudges, emotional manipulation, or personalised consent flows that steer users toward agreeing. As a result, structurally manipulative designs can persist even where formal consent requirements appear met.

Key Policy Recommendations:

- **Mandate design fairness:** The DFA should require that consent mechanisms are symmetrical, neutral, accessible, and free of emotional or adaptive manipulation. Consent should never be coerced through digital service structure.
  - The DFA must support interoperable, machine-readable consent signals and prohibit their circumvention.
  - The DFA should hold Consent Management Platform (CMP) providers accountable for deceptive designs that they standardise across services.
- **Define and ban deceptive design under the UCPD**, including:
  - Profiling shall not be used for manipulative purposes, even where such processing might otherwise comply with data protection rules;
  - Emotional coercion, obstructed opt-outs, hidden or visually biased consent options;
  - Adaptive consent flows or recommender systems that personalise digital services to steer outcomes; and
  - Fatigue-based design and strategies that delay, confuse, or impair user decision-making.
- **Create a narrow grey list for borderline practices (e.g. urgency prompts) subject to burden-shifting and strict conditions, while ensuring the block list covers the most widespread manipulative tactic**s.

- Establish a **right to non-manipulative interaction**: Amend the CRD to enshrine a right to design fairness and require disclosure of optimisation and profiling when used to steer user behaviour.
- **Mandate design auditability:** Require A/B test documentation, optimisation metrics, and behaviour-based targeting logs to be made available to regulators.
- Amend the UCTD to **presume unfairness in contract terms accepted through manipulative design flows**.

By treating deceptive design as a structural violation of fundamental rights rather than a usability flaw, the DFA can restore user agency, rebalance power, and create a digital environment where meaningful consent and fair choice are actually possible.

'Deceptive design', often referred to as 'dark patterns', means any design pattern, interface element, system architecture, or interaction flow that is intended, or has the effect, of materially distorting or impairing the ability of a person to make free, informed, and autonomous decisions, in a manner contrary to the requirements of professional diligence[87]. These are not minor tricks but systemic forms of exploitation that interfere with fundamental rights.

Deceptive design includes, but is not limited to, practices that:
- Exploit cognitive, emotional, or behavioural biases;
- Present choices in an asymmetrical, confusing, or coercive manner;
- Obscure, delay, or hinder access to rights or key functionalities;
- Adapt or personalise interface elements on the basis of profiling or behavioural data to increase compliance;
- Create friction, fatigue, or urgency to discourage refusal or disengagement.

Deceptive design should be presumed to exist where a digital service is structured in a way that systematically steers users toward a particular outcome that primarily benefits the trader, especially where such outcomes are optimised through testing, data-driven personalisation, or behavioural prediction.

What Deceptive Design Looks Like in Practice

Deceptive design patterns are not one-off tricks: they are **recurring strategies that exploit users' attention, expectations, and decision-making**. Research has grouped them into categories that reveal how these tactics function in practice[88]:

[87] See the reference website on deceptive patterns at: https://www.deceptive.design/about-us; While the term 'deceptive design' is now recognised in EU consumer and platform law - and this is why this background paper opts to use it -, many of these practices are better understood as forms of manipulative design: design strategies that exploit cognitive or emotional vulnerabilities, often powered by personal data and recommender systems. These go beyond simple deception and must be addressed as systemic, structural forms of exploitation.

[88] Colin M. Gray, Cristiana Teixeira Santos, Nataliia Bielova, and Thomas Mildner (2024) An Ontology of Dark Patterns Knowledge, CHI '24: Proceedings of the CHI Conference on Human Factors in Computing Systems May 2024.

- **Obstruction**: making it easy to sign up, but hard to cancel (e.g. account deletion hidden behind multiple steps).
- **Sneaking**: slipping items into a shopping cart or adding hidden charges at checkout.
- **Coercive framing:** using visual tricks, nudging language ('No, I like paying more'), or defaults that favour data extraction.
- **Emotional pressure**: exploiting urgency ('Only 2 left!'), social proof ('Your friend bought this'), or guilt ('Don't you care about your privacy?').
- **Choice manipulation**: hiding privacy settings, making fair choices harder to find, or overwhelming users with too many options.

Practical examples include[89]:
- Pre-selected friend suggestions (Snapchat)
- Fake urgency and countdown timers (Shein)
- Repeated notification pop-ups (Snapchat)
- Profiling-based content ordering without easy alternatives (Facebook)

Importantly, these techniques are **rarely deployed in isolation.** In practice, multiple deceptive patterns are often bundled together, reinforcing one another **to maximise engagement, drive monetisation, or extract data[90]**. This bundling contributes to making deceptive design systemic by constructing entire environments in which the exercise of rights becomes difficult, costly, or illusory.

These techniques share a common goal: to steer users toward outcomes they might not freely choose, often by bypassing reflection, making refusal harder than consent, or using interface control to shape behaviour invisibly. They have become embedded in user experience design. Regulatory scrutiny must recognise these not as marginal anomalies, but as systemic strategies that exploit interface control to drive engagement, monetisation, or data extraction, undermining fairness, consent, and autonomy.

Why this Issue Matters

As confirmed by the EC's Fitness Check, **interface design features which exploit behavioural bias, attention fragility, emotional triggers, and the structural imbalance of power between digital businesses and the individuals they claim to serve, are now entrenched across digital markets**[91]. Deceptive design has become one of the most visible and pernicious expressions of manipulation in digital markets[92]. Empirical

[89] Chitra Mohanlal (2025) Exploratory Study of Manipulative Design, Bits of Freedom 28 May 2025. https://www.bitsoffreedom.nl/wp-content/uploads/2025/06/20250616-report-exploratory_study_manipulative_design.pdf

[90] Morel, V., Karegar, F., & Santos, C. (2025). "I will never pay for this" Perception of fairness and factors affecting behaviour on'pay-or-ok'models. arXiv preprint arXiv:2505.12892.

[91] Fitness Check Report.

[92] European Commission. (2023, March 8). Consumer Protection: Manipulative Online Practices Found on 148 out of 399 Online Shops Screened. European Commission. https://ec.europa.eu/commission/

studies show that manipulative design[93] is widespread across digital services. For instance, one review found that over 90% of popular apps deploy at least one form of deceptive or coercive design element to steer user behaviour[94]. **These are not merely poor user experiences or ambiguous buttons; they are deliberate and calculated mechanisms, implemented to extract consent, retain attention, maximise sales, discourage user choices, or steer users towards outcomes that benefit the trader at their expense.**

Other manipulative strategies simulate a sense of obligation, for example implying that users 'owe' attention or data because a service is 'free', 'tailored', or 'curated for you'. These appeals distort user agency by presenting emotional reciprocity as a precondition for refusal, and should also be considered structurally coercive[95]. **Some systems manipulate users by exploiting their social networks**, nudging them to invite contacts, auto-enabling tagging, or defaulting to public visibility. These forms of coerced virality instrumentalise users' relationships for commercial reach. **A growing class of manipulative designs simulate user control without granting it.** Dashboards present false granularity, toggles appear customisable but have no back-end effect, and complex wording masks default tracking, manufacturing the illusion of consent, undermining both transparency and fairness.

Deceptive design has thus evolved far beyond static tricks. In today's digital ecosystems, **it is dynamic and adaptive**: elements of digital services change depending on what a system knows or infers about its users, like their habits, location, hesitation, income level, or mood. This means that not all users see the same pathways. Some are nudged gently, others are trapped aggressively. Deceptive design is no longer just manipulative but is also discriminatory.

This creates a structural asymmetry: while firms use data to optimise for user compliance, individuals must shoulder the full burden of resistance. Refusing, opting out, cancelling, or protecting one's rights requires disproportionate time and effort. **The market rewards those who can capture the most attention, extract the most data, or make refusal practically impossible**.

Recommender systems are a key enabler of deceptive design. They mediate what users see, when, and how, yet their logic is often obscured, and their outputs are

presscorner/detail/en/ip_23_418.

93 Throughout this paper, the term 'manipulative design' refers to structural features of digital services that systematically distort user autonomy or decision-making. This includes practices that have manipulative effects, regardless of whether intent can be proven. While the term reflects widespread usage in academic and policy literature, where scholars have adopted 'manipulative design' to describe a wide range of autonomy-impairing practices, the paper acknowledges that, in legal contexts, the requirement to demonstrate intent may limit enforceability. Therefore, in its legal recommendations, the paper uses the term 'structurally exploitative design' to refer to these same practices in a way that aligns with effect-based enforcement under EU consumer law.

94 Behavioural Insights Team, The behavioural science of online harm and manipulation – and what to do about it, March 2022. https://www.bi.team/publications/the-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it/

95 H. Brignull, Deceptive Patterns (Testimonium Ltd, 2023).

designed to maximise engagement rather than support informed choice. Deceptive design emerges when these systems are configured to steer attention while systematically obstructing disengagement, such as by hiding or complicating access to non-profiling alternatives[96]. By reducing transparency, pre-selecting defaults, and minimising control, recommender systems operationalise interface-level manipulation under the guise of personalisation.

Design strategies that undermine user choice are frequently the product of extensive testing and commercial optimisation. This makes them both deliberate and scalable. Policy must therefore focus not only on the presence of individual deceptive patterns, but on the incentive structures that drive their continuous refinement and deployment at scale[97]. **These strategies frequently go undetected not because they are subtle, but because they are only visible at scale or through testing across user profiles.** What looks like a clean interface to one person may be a coercive maze to another.

In addition to that, **deceptive design often unfolds as a sequence, not a one-off trick**[98]. Seemingly minor nudges, like urgency messages, visual misdirection, or default selections, are regularly followed by additional friction, escalating commitments, or blocked exit paths. **These patterns work cumulatively to steer users into outcomes they might otherwise avoid.** Such dynamics risk being overlooked by enforcement mechanisms focused on single screenshots or isolated user interface (UI) elements[99]. The design is not deceptive because of one click, but because of the overall path it constructs: a funnel that minimises resistance at each step while amplifying behavioural inertia.

People also face being locked into certain digital services due to barriers to switching between different services or to using several competing services in parallel (multi-homing). This vendor lock-in may be reinforced by:

- Network effects (a service becomes more valuable the more people use it);
- The lack of data portability (the ability to transfer an account to a competing service);
- User interface strategies that discourage disengagement or portray the costs of switching as high.

This reinforces a service's market dominance and diminishes meaningful user control[100].

---

96 Chitra Mohanlal (2025) Exploratory Study of Manipulative Design, Bits of Freedom 28 May 2025. https://www.bitsoffreedom.nl/wp-content/uploads/2025/06/20250616-report-exploratory_study_manipulative_design.pdf

97 UK Competition and Markets Authority (CMA), Online Choice Architecture: How digital design can harm competition and consumers, Discussion Paper CMA 155, April 2022. https://www.gov.uk/ government/publications/online-choice-architecture-how-digital-design-can-harm-competition-and-consumers

98 H. Brignull, Deceptive Patterns (Testimonium Ltd, 2023).

99 UI elements refer to the discrete components of a digital interface (for example, a button, checkbox, pop-up, or progress bar) which on their own may not appear manipulative but can become so when sequenced together.

Deceptive design practices are thus the antithesis of fairness, and they are incompatible with a rights-respecting digital environment. To address such practices, the DFA must be anchored in a broader and more accurate definition of deceptive design. A future-oriented approach must capture not only static interface tricks but also:

1. **Asymmetry of choice:** for instance, when opting out requires more time or effort than opting in;
2. **Emotionally coercive prompts:** such as shame-based messaging or 'one-time only' urgency signals;
3. **Multi-layered exit friction:** for instance, subscription cancellation flows that require navigation through multiple menus, misleading language, or phone-based confirmation;
4. **Personalised manipulation:** situations in which profiling, behavioural tracking (including for advertising purposes), or recommender systems are used to dynamically adapt service design in ways that increase user compliance, limit resistance, or re-trigger previous decisions.

Deceptive and addictive design patterns are related but distinct. Deceptive design interferes with decision-making, steering users toward choices they might not otherwise make, such as consenting, buying, or staying subscribed. Addictive design, by contrast, interferes with disengagement, encouraging prolonged or compulsive use through features like infinite scroll, streaks, or emotionally reactive content.

Where deceptive design manipulates choice architecture, addictive design manipulates time and attention. Both exploit behavioural asymmetries, but they operate on different levers of user control, and both require targeted legal responses. Importantly, these forms of manipulation do not impact all users equally: individuals and communities facing structural marginalisation may be disproportionately exposed or vulnerable, depending on how profiling systems categorise and target them. Legal safeguards must account for these intersectional effects to avoid reproducing or deepening existing inequalities.

The Cumulative Impact of Deceptive Design on People's Rights

Deceptive design is not a usability flaw or an inconvenience: it is a structural practice that systematically undermines people's fundamental rights. **At the individual level, manipulative interfaces undermine people's ability to make conscious, self-directed choices.** By engineering consent through fatigue, distraction, or emotional pressure, manipulative interfaces interfere with the right to autonomy, dignity, and data protection. Users are nudged into unwanted subscriptions, tricked into data sharing, delayed from cancelling, or emotionally manipulated into spending. **These tactics leave behind not only financial or contractual consequences, but also feelings of**

[100] London Economics, Digital Consumer Harms: A taxonomy, root cause analysis and methodologies for measurement, Report prepared for the UK Department for Digital, Culture, Media and Sport (DCMS), March 2023. https://assets.publishing.service.gov.uk/media/63c6813ce90e074ee5bb7d4f/DCMS_consumer_harms_research_01-Jan-22.pdf

regret, anxiety, and loss of control. Over time, repeated exposure creates decision fatigue, diminishes trust in one's own judgement, and leads to behavioural resignation, the sense that resisting is futile[101].

Manipulative design interferes with rights and freedoms in multiple ways. It can cloud judgement through confusion, information overload, or attentional hijacking. Others are emotional, exploiting insecurities, urgency, or guilt to push people toward compliance. Many risks have to do with time: wasting users' time by forcing them through long and obstructive flows. And some are existential: shaping what people see, choose, and feel, often without their knowledge. These practices strike at the core of autonomy, dignity, and equality in the digital environment. While their impact is intensified for people already subject to stress, time poverty, low digital confidence, or structural discrimination, the underlying threat affects everyone, because manipulation is embedded into the very design of everyday digital services[102].

The use of in-game currencies, especially when they conceal real-money equivalence, constitutes a systemic form of both deception and emotional manipulation. These currencies intentionally obscure cost visibility and create immersive environments that distort users' sense of value and consequence. This is particularly concerning in child-oriented or gamified commercial environments, where design is engineered to extract users from reality. **In highly immersive digital environments, such as games or gamified shopping apps, design deliberately disorients users and detaches them from the real-world consequences of their decisions**. Emotional and cognitive manipulation becomes embedded in the logic of the digital service. In-game currencies, dynamic rewards, and hidden pricing flows exemplify this.

At the societal level, deceptive design distorts markets. It rewards coercive tactics that undermine users' ability to compare options or make informed choices, as illustrated by the UK Competition and Market Authority's investigation into e-commerce company Emma Sleep for the use of discounts and urgency claims, including countdown timers and high demand prompts that mislead consumers[103]. Separately, a 2025 investigation by digital rights NGO Bits of Freedom revealed the widespread use of manipulative design by large online platforms, highlighting how default settings, profiling, and emotional triggers are systematically used to distort user choice and reinforce disproportionate platform power[104].

---

101 OECD (2022) Dark Commercial Patterns. Oecd Digital Economy Papers. October 2022 No. 336. https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/10/dark-commercial-patterns_9f6169cd/44f5e846-en.pdf

102 Santos, Cristiana and Morozovaite, Viktorija and De Conca, Silvia, No harm no foul: how harms caused by dark patterns are conceptualised and tackled under EU data protection, consumer and competition laws (June 26, 2024). Accepted for publication in Information & Communications Technology Law journal.

103 UK Competition and Markets Authority, Emma Group: consumer protection case. https://www.gov.uk/cma-cases/emma-group-consumer-protection-case

104 Chitra Mohanlal (2025) Exploratory Study of Manipulative Design, Bits of Freedom 28 May 2025. https://www.bitsoffreedom.nl/wp-content/uploads/2025/06/20250616-report-exploratory_study_manipulative_design.pdf

The effects are also unequal and discriminatory. They disproportionately affect people who lack the resources to resist: those without time to read small print, without the digital literacy to identify manipulation, or without the psychological defences to push back against coercive design. As mentioned above, vulnerability here is not the exception but the default condition in asymmetrical digital environments.

Deceptive design also corrodes democratic resilience. It creates environments where dishonesty is rewarded, and where the most extractive designs become standard. Platforms and other digital services that respect user autonomy are commercially punished, outcompeted by those that extract more data, time, or money through hidden tactics. This generates a race to the bottom in the ethics of digital service design, where manipulation is no longer an anomaly but the norm. **This undermines not only individual rights but also collective conditions for democratic participation, equality, and trust in innovation.**

For these reasons, deceptive design must be recognised in the DFA as a structural violation of rights, not as a secondary consumer inconvenience. **It thus undermines the legitimacy of the entire digital economy**. Trust is corroded when users repeatedly feel tricked, manipulated, or disempowered. And in the long term, this damages the promise of digital innovation: when services are optimised for abuse, not transparency, participation becomes riskier, especially for those already marginalised.

Why Existing Rules are Not Working

A Limited Step Forward: Article 25 DSA and its Scope

Despite growing recognition of service-level manipulation, **current EU law addresses deceptive design only partially and unevenly**. The adoption of the DSA marked a critical shift in this regard. For the first time, a horizontal EU law recognised and banned certain forms of deceptive design. Article 25 prohibits specific manipulative practices on online platforms, establishing a valuable legal precedent: that digital service design can indeed manipulate people and materially distort or impair their ability to make free and informed decisions.

However, **the scope of Article 25 is narrow,** both in substance and application[105]. It only applies to online platforms as defined by the law. This **excludes a vast number of commercial actors whose revenue models rely on similar or more invasive practices**: e-commerce sites, consumer-facing apps, fintech and wellness platforms, and many others. These actors are outside the scope of Article 25 even though they may deploy identical strategies: nudging users towards high-priced options, making opt-outs harder than opt-ins, or hiding cancellation paths behind multiple confirmation pages.

[105] Santos, Cristiana and Bielova, Nataliia and Ahuja, Sanju and Utz, Christine and Gray, Colin and Mertens, Gilles, Understanding the scope of Article 25 of the DSA in regulating dark patterns (July 22, 2024).

Gaps in the DSA: Personalisation, Profiling, and Adaptive Design

Even where Article 25 applies, its definition of deceptive design is too narrow. It focuses on clearly coercive or misleading techniques but fails to capture how manipulation is personalised, continuously optimised, and embedded in surveillance infrastructures[106]. In today's digital economy, deceptive design is not static. It is dynamically adapted based on tracked behaviour, emotional state, location, or inferred vulnerability. The same design element may appear differently to different users: one might be shown a bright 'accept all' cookie prompt late at night; another receives a subtly adjusted version based on previous refusals or hesitation. Design is tailored through recommender systems, A/B testing, and behavioural analytics not only to influence, but to exhaust resistance.

This convergence of digital service design manipulation and commercial surveillance therefore remains largely unregulated. Recommender systems, consent banners, and digital service design are increasingly shaped by profiling and behavioural targeting. Recommender systems learn which types of nudges work best on specific individuals, based on their past behaviour and inferred traits. Profiling tools enable digital services to adjust design elements such as button placement, wording, or friction in real time, depending on what is most likely to elicit a desired response from each user. Consent banners, for example, are often configured to exploit fatigue or confusion, nudging users to click 'accept' through repeated exposure or misleading layouts. What links these techniques is the integration of behavioural design with the underlying data economy, where the digital service is not static but dynamically optimised to steer individuals based on their predicted vulnerabilities. Yet no current legal instrument explicitly addresses this convergence.

The Value of the DFA: Addressing Design as Structure, Not Violation

The DFA is better positioned to address manipulative design because, unlike the DSA, it can apply to the full spectrum of digital services and commercial actors. As outlined above, the DSA's scope and framing remain too narrow to address the systemic, data-driven nature of behavioural manipulation. **The DFA can fill this gap by recognising structurally exploitative design not as a set of isolated practices, but as a structural feature of business models built on personalisation, profiling, and attention extraction.** It can move beyond reactive enforcement and help establish fairness as a design obligation across the digital economy.

Additionally, **while Articles 34 and 35 DSA introduce obligations around systemic risk assessment and design audits, they apply only to Very Large Online Platforms (VLOPs) and Search Engines (VLOSEs), and are anchored in a public interest logic** (e.g. disinformation, harm to minors). These provisions do not establish a general duty of fairness in design, nor do they apply to the vast majority of commercial services that deploy the same structurally exploitative techniques. The DFA can also fill this gap by framing manipulation as a horizontal consumer protection issue, one that cuts across services, business models, and sectors, regardless of size.

[106] Becker Castellaro, Sebastian; Penfrat, Jan: The DSA fails to reign in the most harmful digital platform businesses – but it is still useful, VerfBlog, 2022/11/08. https://verfassungsblog.de/dsa-fails

Digital Markets Act (DMA) and Data Act: Promising but Falling Short

While the DMA does not include direct prohibitions on deceptive or manipulative design, certain practices deployed by gatekeepers could be assessed under the anti-circumvention clause in Article 13, where they undermine compliance with core obligations such as user choice, data portability, or fair access. However, this mechanism is limited to specific contexts involving a few designated gatekeeper companies and cannot substitute for unified protections against manipulative design across the digital economy. This further reinforces the **need for a consumer law-based instrument like the DFA to address deceptive design as a structural and cross-cutting issue, beyond the scope of platform-specific obligations**.

In contrast, **the Data Act takes a more promising approach by grounding its anti-dark pattern provisions in general principles of fairness**, particularly in the context of data access and sharing. Rather than listing specific practices, it prohibits strategies that exploit fatigue, inattention, or lack of knowledge, allowing regulators to consider cumulative and systemic effects. However, its scope is limited to data sharing scenarios and does not cover the broader ecosystem of manipulative practices embedded in consent flows, personalisation interfaces, or monetisation architectures. Crucially, it does not address design strategies that exploit behavioural vulnerabilities to drive retention or influence, especially when no data access is involved.

Gaps in the GDPR: Consent Without Autonomy

The GDPR, while protecting fundamental rights, also does not provide the necessary protections. **Deceptive design directly undermines the conditions for valid consent. But enforcement under the GDPR often focuses on whether information was provided or a consent box clicked, rather than how the design context shapes the user's choice.** Many real-world consent flows secure agreement through nudging, visual dominance, or cognitive overload; yet regulators rarely interrogate how digital services are structured to extract compliance. Personalisation tools and UX testing are used to optimise consent rates, not respect user agency.

DPAs currently lack the mandate or tools to address coercive design as a structural strategy. Existing frameworks address manipulative design only indirectly, through transparency and consent requirements, rather than by recognising the manipulative design logic itself as a legal violation. The DFA should fill this gap by directly prohibiting structurally manipulative design strategies, without forcing enforcement authorities to rely solely on the indirect route of invalidating consent.

This points to a **core legal contradiction: consent may formally comply with GDPR requirements while still being obtained through manipulative design.** The result is that coercive or engineered agreement becomes legally valid, undermining the principle of freely given and revocable choice. Neither the GDPR nor the UCPD gives regulators clear powers to act against interfaces or entire services used as an instrument of manipulation.

UCPD and the Illusion of Free Choice

The UCPD defines unfair practices in terms of professional diligence and material distortion. But it offers little guidance on how to apply these standards to adaptive, personalised, and dynamic digital environments. It does not recognise design as a technical system engineered to induce behaviour, especially where CMPs are used to create the illusion of free choice. As a result, most deceptive patterns are treated as usability flaws or borderline legality, rather than as forms of unfair commercial practice requiring intervention.

In February 2025, the Higher Regional Court of Bamberg, Germany, ruled that consumer protection organisations could not rely on Article 25 DSA to challenge manipulative design on a ticketing platform, on the grounds that the practice fell under the scope of the UCPD[107]. The case illustrates a structural flaw in current law: legal frameworks may displace one another rather than reinforce protection. This increases the urgency for a dedicated legal instrument like the DFA, capable of addressing structurally exploitative design as a systemic practice across all digital services.

Legal fragmentation is part of a deeper regulatory failure. While consumer law, data protection law, and competition law each conceptualise harm differently[108], the real obstacle is systemic. Deceptive design is not incidental; it is a core feature of dominant digital business models, designed to be scalable, dynamic, and difficult to trace. Its effects are often cumulative, emotional, and structurally embedded, yet regulation continues to focus on isolated violations and individual harm. Many regulators are under-resourced, lack access to testing environments, and operate within mandates that were not built to address adaptive manipulation. Meanwhile, users bear the burden of detecting and resisting manipulative tactics they cannot even see. **The DFA must confront this reality. It must shift the burden to traders, recognise manipulation itself as a form of harm, and embed enforceable standards of fairness into the design of digital services.**

Proposed Policy Changes to Address Deceptive Design

The Fitness Check identifies personalisation strategies that exploit psychological or emotional vulnerabilities as a source of consumer harm that current rules struggle to capture[109]. Indeed, and as shown above, current legal frameworks continue to treat deceptive design as a series of isolated compliance failures rather than recognising it as a systemic feature of digital business models. As we have seen, **manipulative design is not accidental or peripheral: it is structurally embedded in the logic of data extraction, behavioural targeting, and engagement maximisation.** The fragmentation of enforcement across the applicable legal frameworks has proven inadequate to

[107] Oberlandesgericht Bamberg, Judgment of 5 February 2025 – 3 U 324/23.

[108] Santos, Cristiana and Morozovaite, Viktorija and De Conca, Silvia, No harm no foul: how harms caused by dark patterns are conceptualised and tackled under EU data protection, consumer and competition laws (June 26, 2024). Accepted for publication in Information & Communications Technology Law Journal.

[109] Fitness Check Report.

address how design practices are continuously optimised to influence user behaviour. Consumer law must adapt to this shift by recognising digital service design itself as a vector of commercial power[110].

This section consolidates all structural and horizontal measures that should apply across manipulative design practices, including those found in addictive systems. **The following recommendations apply to deceptive design broadly, but also serve as a baseline for addressing addictive design where it involves attention manipulation, coercive nudging, or the systematic undermining of user autonomy**.

Reimagining Consent: From Manipulation to Structural Fairness

Deceptive design is one of the most pervasive ways in which people's agency is undermined online. It shapes how information is presented, how choices are framed, and how consent is extracted. This is particularly clear in the case of **consent interfaces, which have become a primary site of manipulation**. While the GDPR defines valid consent as freely given, informed, specific, and revocable, it does neither address how consent is requested, nor how interfaces are used to bypass refusal. This is where the DFA must intervene.

Consent is a cornerstone of data protection law and an essential condition for enabling consumers to protect their fundamental right to privacy. Where it is lawful and meaningful, it serves as a critical barrier against illegitimate data processing. But **today's digital environments frequently strip consent of its substance**. Especially in commercial personalisation, it often becomes a procedural formality, undermined by information overload, or lack of meaningful alternatives. Users mostly 'agree' to tracking as a condition of access, with little understanding of the implications, or awareness that profiling frequently continues even after refusal.

The problem thus lies in the service-level dynamics that routinely distort it. **Most consent interfaces are not designed to inform or empower but to extract compliance.** Visual asymmetry, obstructive opt-outs, emotional cues, and profiling-based adaptation are routinely used to steer users toward acceptance. These practices collapse the distinction between consent and coercion, and degrade the very notion of choice.

This places an **impossible burden on individuals**. Expecting every user to parse dense privacy notices, resist emotional triggers, and repeatedly exercise refusal across dozens of services is neither fair nor sustainable. It externalises accountability from those who profit from surveillance and onto already overburdened individuals. And it compounds structural inequalities: those with less time, education, or cognitive capacity are disproportionately targeted and manipulated. Restoring fairness in digital

[110] Sánchez Chamorro, L. (2024). Disentangling Vulnerability to Manipulative Designs: An Experiential Perspective to Rethink Resistance Strategies (Doctoral dissertation, University of Luxembourg, Faculty of Humanities, Education and Social Sciences, Psychology Department). Defence held on 19 November 2024 in Esch-sur-Alzette.

environments therefore requires more than fixing consent: it requires regulating the architecture that distorts it.

The DFA cannot and should not redefine what counts as valid consent under the GDPR, nor does it replace the need for its strong enforcement[111]. But it should regulate the strategies that make meaningful consent impossible in practice. This includes interface-level coercion, fatigue, deception, and behavioural targeting. Where design is used to simulate choice while foreclosing refusal, it should be prohibited as a form of unfair commercial practice. The DFA's role is not to duplicate the GDPR, but to address the manipulative environments that render its protections ineffective.

While the 2023 Cookie Pledge[112] failed to deliver structural change, it did propose useful design principles: symmetry, meaningful withdrawal, and clarity of architecture. These principles should not remain aspirational. The DFA must make them enforceable through binding, auditable standards that ensure fairness in every digital interaction. A fair consent interface should be:

- **Automatable**: For frequent, similar requests (such as for tracking or marketing) consent should be expressible via interoperable, machine-readable signals (e.g. Global Privacy Control or Advanced Data Protection Control). These signals must be based on open protocols and accessible to browsers and OS providers.
- **Clear, symmetrical, and visually neutral**: 'Accept' and 'Reject' must be equally visible, accessible, and free of visual or emotional bias.
- **Non-conditional and non-bundled**: Services must not be made conditional on consenting to unnecessary processing, nor should refusal come with degraded service quality.
- **Comprehensible and accessible**: Explanations must use plain language and consistent icons. Layered information should be genuinely accessible, not buried in interfaces.
- **Revocable and changeable**: Consent dashboards should be easy to find and use, with no hidden friction.
- **Postponable without penalty**: Consent must not be forced through repetition, denial of access, or psychological pressure.
- **Emotionally neutral**: Design should not invoke guilt, urgency, or fear in connection with refusal.
- **Unaffected by profiling or adaptation**: The service design must be the same regardless of user traits, past choices, or predicted behaviours.

---

111 EDPB Guidelines 03/2022 on deceptive design patterns in social media platform interfaces. https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-032022-deceptive-design-patterns-social-media_en

112 The 2023 Cookie Pledge was a voluntary initiative launched by the European Commission to encourage companies to improve the design of cookie banners in line with the ePrivacy Directive and GDPR. While non-binding, it highlighted persistent concerns over coercive consent interfaces and aimed to promote good practices around user choice and transparency. https://commission.europa.eu/live-work-travel-eu/consumer-rights-and-complaints/enforcement-consumer-protection/cookie-pledge_en

These principles are not about rigid templates, but about ensuring that consent is a space for accessible and dignified choice, not a battleground of manipulation. To treat consent as meaningful, the design asking for it must be fair.

CMPs are a critical node in this ecosystem[113]. As intermediaries between users, publishers, and advertisers, they shape how consent is requested and conveyed. Yet many CMPs embed structurally exploitative patterns at scale: pre-ticked boxes, misleading toggles, deceptive grouping. These are not one-off failures, but design features standardised across millions of sites. **The DFA should close this accountability gap by holding CMP providers directly responsible for unlawful or structurally exploitative designs.**

Lastly, **the DFA must address structural inefficiencies that fuel consent fatigue**[114]. Making users repeat the same decision across services is not neutral: it is a form of friction that punishes refusal and normalises acquiescence. The solution is **interoperable, machine-readable signals that allow people to express their preferences once and have them respected across contexts.** Binding browser-level signals, as proposed in Article 9(2) of the (now withdrawn) ePrivacy Regulation, offer a viable and scalable approach. These must be recognised under consumer law to prevent service-level circumvention and eliminate repeated prompting.

Any future DFA fairness framework must make clear that so-called 'pay-for-privacy' or 'Consent or Pay' models are incompatible with fundamental rights: conditioning access to rights on payment is a coercive consent structure that undermines autonomy and equal protection, and must be prohibited.

Such a shift would reduce reliance on structurally exploitative banners, close a key enforcement gap, and shift the burden of proof back onto those who process data. Tools like California's Global Privacy Control (GPC)[115] have shown this is technically feasible. **The DFA should mandate legal recognition of such signals and prohibit companies from circumventing them through banners or interface friction.** To ensure effectiveness, browser and operating system providers should also be required to support recognised global signals by default, so that individuals can exercise their rights without depending on corporate discretion.

---

113 Nataliia Bielova, Cristiana Santos, Colin M Gray. Two worlds apart! Closing the gap between regulating EU consent and user studies. Harvard Journal of Law and Technology, 2024, 37 (3), pp.1295-1333.

114 AlExis Hancock (2019) Designing Welcome Mats to Invite User Privacy, EFF February 14, 2019. https://www.eff.org/deeplinks/2019/02/designing-welcome-mats-invite-user-privacy-0

115 California's Global Privacy Control (GPC) is a technical signal that users can enable in their web browsers or extensions to automatically communicate their privacy preferences, specifically, their opt-out of the sale or sharing of personal information under the California Consumer Privacy Act (CCPA). Websites covered by the CCPA are legally required to honour the GPC signal as a valid opt-out request, without requiring users to click through individual cookie banners or settings. It is designed to give users a simple, universal way to assert their privacy rights across multiple sites. https://oag.ca.gov/privacy/ccpa/gpc

Amend Existing Consumer Law

To tackle deceptive design, the DFA should propose a coordinated set of reforms across EU consumer protection law to reflect how modern manipulation in digital services distorts user autonomy. **Consent-related design must be understood not as a neutral interface feature but as structurally shaped by service-level incentives and optimisation logics**. By requiring that digital services embody fairness, not merely functionality, the DFA can help re-establish consent as a meaningful expression of user agency.

Auditable Design and Default Settings

As mentioned, many deceptive designs go undetected not because they are hard to see, but because they are only visible at scale or through testing across user types. National authorities cannot be expected to reverse-engineer every digital service, nor can users be expected to file complaints about subtle nudges that only become coercive over time. The DFA must therefore **require auditable design justifications, alongside proportionate algorithmic impact assessments for systems that affect user autonomy**. More on this can be located in the Enforcement Chapter VIII below.

Deceptive design frequently operates through the manipulation of default settings that steer users toward choices they might not otherwise make. Defaults that enable profiling-based recommender systems, persistent notifications, or data-intensive features can function as silent nudges, exploiting inertia and cognitive bias. The DFA should require that such features remain deactivated unless users provide clear, informed, and unbundled consent. **Default configurations must reflect autonomy-preserving choices, rather than commercial optimisation**. Regulating defaults is therefore essential to prevent consent harvesting, reduce misleading digital service practices, and ensure that user choice is respected not only in form but in practice.

Regulating Behaviourally Optimised Digital Services

Particular concern arises where digital services use recommender systems or optimisation to exploit predicted user states such as fatigue, boredom, or emotional vulnerability. When such systems personalise prompts or content delivery based on inferred susceptibility, the line between engagement and manipulation collapses. The DFA should make clear that these practices constitute structural deception and merit specific regulatory intervention.

Similarly to the regulation of Addictive Design, **where recommender systems are used, users must be given real control, including the option to choose content curation methods that are not based on profiling without being penalised by degraded functionality**[116].

[116] Panoptykon Foundation and People vs. Big Tech, "Discussion Paper Towards algorithmic pluralism", 4 July 2025. https://panoptykon.org/sites/default/files/2025-07/towards-algorithmic-pluralism-in-the-eu-policy_pvbt-discussion-paper_04072025.pdf

This would provide consumers with a viable alternative to engagement-driven content curation, and help mitigate compulsive loops caused by algorithmic reactivity. This is especially important where design personalisation obscures the structure of manipulation; that is, where people do not realise they are being steered toward specific outcomes.

Disclosure, Testing, and Early Detection

Enforcement mechanisms that rely on transparency and *ex post* individual redress cannot cope with manipulation that is continuous, low-visibility, and calibrated through real-time data. In line with the European Parliament's stance on addictive design, the DFA should **require that digital services disclose experimentation dashboards, including A/B test outcomes and behavioural nudges used on different user segments**. These dashboards should be accessible to regulators, public interest researchers, and independent watchdogs - including the press -, and key finding should be made publicly available in clear, aggregated form to promote democratic oversight without shifting responsibility to individual users.

Regulators should also draw on **user reporting and** independent analyses of service features **to detect these patterns early**, and recognise that harm can be emotional, time-based, or compulsively financial, not only tied to product price or accuracy[117]. While these techniques are often discussed in relation to addictive design, they also constitute deceptive manipulation when used to obscure consent or to test which elements of the digital service best suppress its refusal.

Clarifying Consent Manipulation under the UCPD: General Clause, Block list, and Narrow Grey List

To promote fair consent practices, the DFA should introduce stronger protections by amending the UCPD, the CRD and UCTD. In order to support legal certainty, **deceptive designs should be defined through a general clause in the UCPD - aligned with Article 25 of the DSA - prohibiting any design, interface, or interaction pattern that is contrary to digital professional diligence or the law and that distorts or impairs a person's free and informed decision-making.** Deceptive design should refer to any digital practice that misleads, pressures, or manipulates users by exploiting cognitive, emotional, or social vulnerabilities, including through the obscuring or disguising of material information, manipulative presentation of choices, exploitative default settings, or other interface techniques that undermine autonomy and informed consent. This formulation would extend the logic of Article 5 UCPD while anchoring enforcement in the realities of digital manipulation.

Distorting consent flows through visual bias, emotional pressure, obstructed refusals, or profiling-based personalisation that adaptively steers users toward compliance should be considered an unfair practice, and thus added to the UCPD's block list.

[117] Elena Petrovskaya, Sebastian Deterding, and David I Zendle. 2022. Prevalence and Salience of Problematic Microtransactions in Top-Grossing Mobile and PC Games: A Content Analysis of User Reviews. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 560, 1–12.

Practices such as oversized 'accept' buttons, hidden 'reject' options, or consent conditioned on unnecessary tracking should be block-listed outright. So should be sequential or cumulative strategies such as confirmshaming[118], forced continuity, obstructed opt-outs, and retention through friction, which work not through a single deceptive act but by progressively escalating pressure or commitment across multiple steps. These are structurally manipulative by design, not circumstantially unfair, and their categorical prohibition is essential for legal clarity and consumer protection.

Unfair personalisation techniques must also be captured explicitly by the DFA. It should be considered unfair to personalise consent flows using:

- Inferred or sensitive traits (as defined under Article 9 GDPR or their functional equivalents, e.g. deducing distress, fatigue, or health status from usage patterns);
- Real-time psychological or emotional cues (e.g. boredom, loneliness, impulsivity); or
- Predictive modelling of compliance to increase refusal friction or intensify compulsive engagement.

Such techniques are unfair regardless of formal GDPR compliance: even if traders claim a lawful basis under data protection law, the use of profiling for manipulative consent or choice design should be considered structurally exploitative under consumer law. In this way, the DFA complements the GDPR by treating the design and commercial application of profiling as unlawful when it distorts decision-making environments.

Forms of coerced virality should also be block-listed where consent is bundled, non-granular, or presumed, similarly to patterns that manufacture the illusion of consent. Friction loops that exploit fatigue, not information imbalance, are deployed to incentivise engagement, and this justifies block-listing as a form of retention by design[119]. Equally, the use of in-game currencies without showing the real-money equivalent should be block-listed, as it obscures the cost of purchases, misleads users about the value of items, and facilitates spending without informed consent. Gamified purchase loops that hide or delay the presentation of real costs until after multiple steps are also incompatible with the transparency required by digital professional diligence, and should be prohibited as deceptive by design.

Including these practices in the block list would bring much-needed clarity to enforcers and market participants, and eliminate the legal ambiguity that has long enabled traders to exploit regulatory gaps. Unlike individual misleading claims, these patterns are embedded in system logic, repeated across services and use cases, and deployed precisely because they are effective at undermining free choice.

[118] Confirmshaiming refers to a deceptive design tactic where a service frames the option to refuse or opt out in a way that makes the user feel guilty, embarrassed, or ashamed.

[119] Retention by design refers to design strategies deliberately built to keep people engaged or prevent them from leaving a service, regardless of their intent, such as friction loops, artificial barriers to unsubscribing, or exploitative defaults.

However, certain practices might warrant context-sensitive assessment. A strictly limited grey list should apply only where the same design pattern can serve both legitimate and manipulative purposes. Its inclusion would be justified when:

- The pattern may serve a legitimate function depending on implementation;
- Enforcement requires a burden shift, but not necessarily a ban;
- There is a clearly defined threshold for unfairness and no blanket permissiveness.

Examples could include multi-step flows that add friction without preventing refusal, personalised prompts that mimic user intent but can be easily overridden, or tiered in-game currency systems where monetary value is partially obscured but disclosed transparently. Similarly, digital services that temporarily omit real-money prices at the point of purchase may warrant scrutiny if they ensure accessible and timely cost information and do not target structurally vulnerable users. Such practices sit at the edge of fairness and should trigger a presumption of unfairness unless traders can demonstrate that the design is necessary, transparent, and non-exploitative.

Such a grey list must be narrowly scoped, clearly defined, and accompanied by strong disincentives for traders to rely on it. It must not become a loophole for structurally exploitative digital service design or a shield for optimisation strategies that exploit users. The goal is not to tolerate borderline practices, but to enable accountability where outright prohibition is legally or practically complex.

CRD Reforms: Equivalence, Symmetry, and Fairness by Design

Additionally, the CRD should be amended to ensure that consent mechanisms offer genuinely equivalent options, presented with clarity and symmetry. Refusal should never be more difficult than acceptance, and traders must avoid nudges, confirmshaming, or any design tactics that steer users toward compliance. Users should also be able to revise their decisions easily, and refusal should not entail degraded service unless strictly justified. These reforms would help operationalise the GDPR's consent standards in user-facing design. Moreover, the CRD should require disclosure where digital services are personalised or optimised in ways that impair user decision-making. Finally, the CRD should establish a new right to fairness in digital service design, ensuring users can disengage without manipulation.

UCTD and the Presumption of Unfairness

Finally, **the UCTD should presume as unfair any contractual terms accepted through deceptive or addictive design flows, such as pre-ticked boxes or buried opt-outs, particularly when reinforced through testing or profiling.** These proposals aim to prevent digital service-level coercion, support meaningful user control across digital services, and help shift the burden of proof away from individuals by addressing the systemic nature of structurally exploitative design practices.

VI. Profiling and Forms of Unfair Personalisation

Executive Summary

Unfair personalisation refers to the use of profiling to adapt content, interfaces, prices, or offers in ways that distort user autonomy, reinforce structural asymmetries, or lead to discriminatory outcomes. This includes targeting based on inferred traits, nudging users toward choices that benefit the trader, or removing meaningful alternatives to personalisation.

Though often marketed as user-friendly and desirable, **profiling-based personalisation is a core tactic of commercial surveillance**[120]. It oftentimes functions as a system-level mechanism to shape user behaviour, extract value, and reinforce data power. These effects are mostly opaque, difficult to contest, and systemic in nature.

In today's digital environments, all users are exposed to the risks of unfair personalisation by default. A user's vulnerability to unfair personalisation is not only a matter of their fixed traits but is produced through the intersection of opaque profiling systems with contextual, emotional, and socio-economic factors. While everyone is exposed, those already subject to inequality frequently experience disproportionate exclusion or disadvantage.

Current Legal Gaps:

- The GDPR does not regulate how profiling shapes user environments or digital service design.
- The DSA's Article 38 only applies to VLOPs and VLOSEs, most digital services are excluded from its scope.
- Consumer law doesn't yet treat profiling-based personalisation as a form of market distortion or exploitation.

Key Policy Recommendations:

- **Default protections and user rights**
  - Guarantee a right to non-personalised services by default, unless personalisation is strictly necessary for the core function of the service;
  - Require that non-personalised options remain fully functional and accessible, without unjustified degradation of quality or restriction of essential features;
  - Introduce a visible 'erase my footprint' function, allowing users to delete tracking-based data across services.
- **Fairness-by-design in recommender systems**
  - Require that recommender systems be designed and evaluated to ensure fair and non-discriminatory treatment of users, transparency over ranking and

[120] Commercial surveillance refers to the continuous collection, processing, and monetisation of people's data by private companies, typically through profiling, behavioural tracking, and targeted optimisation. Its purpose extends to shaping behaviour and extracting economic value, often in ways that are opaque, difficult to contest, and structurally embedded in digital business models.

selection criteria, and safeguards against systematic exclusion or visibility bias.
    - Require regular assessments by digital service providers of how their personalisation logic affects agency, autonomy, exposure to manipulation, and structural disadvantage.
- **Transparency and contestability**
    - Mandate full transparency, including disclosure of what traits or data are used, what is being optimised, and how personalisation affects content, prices, and options.
    - Ensure that when service conditions or contractual offers vary based on profiling or personalisation, they remain transparent, intelligible, and open to audit and challenge by users and regulators.
- Add a general clause to the UCPD **defining unfair personalisation** as any practice that:
    - Relies on behavioural profiling, inferred emotional states, or contextual and structural vulnerabilities; and
    - Distorts user autonomy or materially impairs free decision-making.
- **Block-list the most exploitative personalisation practices**, including:
    - Personalisation based on sensitive or inferred sensitive traits, except where such processing is strictly necessary to provide a service explicitly requested by the user, complies with data protection law, and does not produce exploitative or discriminatory effects;
    - Real-time interface adaptation that exploits emotional or psychological states (e.g. boredom, fatigue, distress);
    - Personalised friction aimed at discouraging refusal, opt-out, or disengagement.
- **Introduce a grey list of presumptively unfair practices**, such as:
    - Gamified nudging (e.g. progress-based rewards linked to consent);
    - Personalised default options based on predicted compliance;
    - Personalisation based on past hesitations or refusal patterns.
- Amend the CRD to ensure **clear, upfront explanation and pre-contractual clarity** on how personalisation shapes offers, pricing, or content, and to secure users' right to easily change or disable personalisation without undue friction or loss of service quality.
- Amend the UCTD to **ensure that any contractual term or offer personalised through profiling, opaque recommender logic, or manipulative design is deemed unfair and therefore unenforceable where it exploits user vulnerabilities or conceals material information**.

By treating unfair personalisation as a structural violation of consumer rights and people's agency rather than a neutral design choice, the DFA can rebalance power, enable meaningful choice, and ensure that personalisation serves people's needs rather than solely digital service providers' profits.

'**Unfair personalisation' means any practice of adapting content, interface features, offers, pricing, or functionalities on the basis of user profiling, where such personalisation systematically undermines autonomy, deepens power asymmetries,**

or results in discriminatory or exclusionary outcomes. The DFA should define such personalisation to be unfair where **it occurs contrary to the principle of professional diligence as applied to digital services or in a way that subordinates people's rights and welfare to commercial optimisation goals**. This includes, but is not limited to, practices that:

- Personalise offers or interfaces based on inferred emotional, psychological, financial, or cognitive traits;
- Apply profiling to nudge, steer, or pressure users into decisions that primarily benefit the trader's commercial interests (e.g. maximising revenue, retention, or data extraction), rather than the user's preferences or choices;
- Create unequal access to information, services, or opportunities based on opaque or undisclosed criteria;
- Rely on behavioural tracking or inference to personalise services in ways that users cannot reasonably foresee or contest;
- Remove or obscure the option to use non-personalised versions of a service unless such personalisation is strictly technically necessary.

Personalisation systems must be understood as core infrastructure for behavioural governance. They are not benign or user-serving by default. People expect these systems to offer meaningful control, including the ability to opt out, pause, reset, or adjust what is shown and why. However, **current personalisation environments routinely obscure how decisions are made and deny users straightforward mechanisms to exercise agency**. Systems that frame engagement-based optimisation as inevitable or necessary actively undermine user autonomy and trust[121]. Moreover, AI-powered personalisation increasingly shapes user environments by optimising for engagement, conversion, or behavioural predictability. **While often presented as enhancing relevance, it can limit autonomy, reinforce market asymmetries, and normalise hidden influence, turning adaptive design into a mechanism of structural manipulation**[122].

Unfair personalisation in the context of digital services **occurs where users are subjected to differential treatment based on behavioural or inferred characteristics in a way that impairs equal access, transparency, or meaningful choice**. The industry's claim that personalisation is always beneficial to users ignores the reality that most systems are optimised not for relevance, but for extraction, persuasion, and conversion. Framing personalisation as an enhancement of choice obscures the structural harms it produces: discrimination, inequality of access, erosion of autonomy, and loss of trust in digital environments. These are not marginal side-effects, but violations of fundamental rights embedded into the everyday architecture of digital life.

---

121 Wong, Y. N., Jones, R., Das, R., & Jackson, P. (2023). Conditional trust: Citizens' council on data-driven media personalisation and public expectations of transparency and accountability. Big Data & Society, 10(2).

122 Tahir Nisar, 'The Personalisation Economy: How Is AI Affecting Businesses and Markets?' (Economics Observatory, 19 March 2025). https://www.economicsobservatory.com/the-personalisation-economy-how-is-ai-affecting-businesses-and-markets

What Unfair Personalisation Looks Like in Practice

Personalisation in practice routinely involves techniques that leverage structural asymmetries and exploit emotional or contextual vulnerabilities. A few examples across different sectors include:

- **Behavioural Pricing and Scarcity Triggers on e-commerce platforms:** A user browsing for airline tickets is shown rising prices and countdown timers based on their repeated visits, creating a false sense of urgency and pressuring them to book. Another user, profiled as less price-sensitive, is shown consistently higher prices for the same product. Such price discrimination exploits behavioural predictability and undermines the principle of equal access.
- **Ranking on streaming and content platforms:** Engagement-based personalisation is used across different types of content services, including editorial streaming platforms and user-generated content platforms. For example, Netflix and YouTube both recommend content based on inferred user preferences or behavioural patterns. While accountability mechanisms may vary, and while the risk of harmful content loops is particularly pronounced on platforms like YouTube, where user-generated content is surfaced algorithmically and in real time, the shared logic of engagement optimisation can still undermine well-being by reinforcing narrow consumption patterns and limiting exposure to diverse or intentional choices.
- **Automated Exclusion on job platforms and financial services:** Personalisation systems on job search and credit scoring platforms may systematically deprioritise or exclude users based, amongst others, on previous activity, demographic proxies, or indirect behavioural indicators. Users from marginalised backgrounds may be invisibly segmented into lower-opportunity pathways or discouraged from applying at all, without knowing how or why their options are limited. Research on crowdworking platforms has shown that digital labour systems can embed structural unfairness in how tasks and opportunities are distributed, reinforcing existing inequalities rather than creating fair access to work[123].

Why this Issue Matters

The portrayal of personalisation as technical necessity or a neutral or user-friendly feature has predominantly served to obscure the reality that hyper-personalisation is a business model choice, not a condition of service quality. Personalisation is widely promoted as a way to enhance digital services by making them more relevant, efficient, or user-friendly[124]. Whether through dynamic pricing, curated newsfeeds, product suggestions, content rankings, or adaptive services, personalisation is framed as a mechanism that improves user experience.

[123] Fieseler, C., Bucher, E. & Hoffmann, C.P. Unfairness by Design? The Perceived Fairness of Digital Labor on Crowdworking Platforms. J Bus Ethics 156, 987–1005 (2019).

[124] For example, the European Commission's 2024 Fitness Check report describes personalisation as a driver of 'consumer empowerment' despite recognising its risks [Fitness Check, 2024]

Yet this framing obscures a deeper reality: **personalisation, as currently deployed, is not simply about user convenience or technological efficiency. It is a central strategy of commercial surveillance and behavioural modulation, premised on continuous inference and targeted manipulation**. Research demonstrates that digital personalisation systems often exploit individual and group-level vulnerabilities leading to distorted decision-making to serve business interests rather than consumer rights. In one behavioural study, participants exposed to emotionally manipulative nudges were twice as likely to engage in a purchase they later regretted[125]. Such findings underscore the **need to regulate not only the outputs delivered to users, but the underlying design choices and optimisation strategies that shape how systems act on people**.

Not Technological Inevitability, but Structural Power

Many personalisation systems restrict rather than expand user choice. By continuously adjusting design and content to match predicted preferences, they confine users to behavioural paths that reinforce prior actions. This process locks individuals into increasingly narrow trajectories of engagement, structured less by interest than by system-level optimisation[126]. **Despite extensive data collection and profiling, personalisation systems frequently rely on crude segmentation techniques and opaque algorithms. Their apparent promise of "relevance" is therefore misleading: even where they achieve accuracy, the underlying optimisation still constrains autonomy and entrenches structural unfairness**.

Crucially, **personalisation is not technically necessary for digital services to function effectively**. Core features like content discovery tools, product rankings, and information retrieval, and even recommender systems, when stripped of opaque profiling, can be delivered via contextual signals, user-configured settings, or transparent filters. The portrayal of deep personalisation as a technological inevitability erases these alternative approaches. It also **conceals the fact that hyper-personalisation is not a neutral design choice, but a business model built on data extraction and asymmetrical power**.

---

125 Behavioural Insights Team, The behavioural science of online harm and manipulation – and what to do about it, March 2022. https://www.bi.team/publications/the-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it/; Kirk, H. R., Vidgen, B., Röttger, P., & Hale, S. A. (2023). Personalisation within bounds: A risk taxonomy and policy framework for the alignment of large language models with personalised feedback. arXiv preprint arXiv:2303.05453; Beer, D., Redden, J., Williamson, B., & Yuill, S. (2019). Landscape summary: Online targeting: What is online targeting, what impact does it have, and how can we maximise benefits and minimise harms?; Rezk, A. M., Simkute, A., Vines, J., Elsden, C., Evans, M., Jones, R., & Luger, E. (2024, October). User-Centric Tensions: Exploring Perceived Benefits and (Dis) comfort in Media Personalisation. In Proceedings of the 13th Nordic Conference on Human-Computer Interaction (pp. 1-13); Zanker, M., Rook, L., & Jannach, D. (2019). Measuring the impact of online personalisation: Past, present and future. International Journal of Human-Computer Studies, 131, 160-168.

126 Oscar Gandy Jr, 'Statistical Surveillance: Remote Sensing in the Digital Age' in M. Hildebrandt and S. Gutwirth (eds), Profiling the European Citizen. Cross-Disciplinary Perspectives (Springer 2008) ch 4, pp. 24–26.

This logic is particularly pronounced in the digital advertising ecosystem. While some users share information during registration, most profiling relies on data not freely given: behavioural traces, inferred signals, and cross-platform tracking. These inputs are then used to build granular profiles for behavioural targeting, targeting that is often hidden behind the language of personalisation, relevance, or user experience.

Unequal Treatment, Unequal Rights

Even when personalisation is framed as improving relevance, users may experience it as unsettling, overly intimate, or opaque[127]. What is marketed as user-centric design is in fact an asymmetrical strategy of influence. Personalisation systems do not treat users equally, nor are they designed to. Their purpose is to identify who is most persuadable, monetisable, or susceptible to certain outcomes. It is those who build and optimise these systems who benefit most, not those who interact with them.

Research of personalised news environments confirm that users frequently experience them as a loss of agency and informational autonomy: they are shaped by opaque patterns that users cannot meaningfully influence, while content is filtered through assumptions they cannot contest. Users often cannot trace how the system has profiled them, why they are being shown certain content, or how to adjust it. **This opacity sustains a power imbalance in which users are recipients of influence but lack visibility or recourse**[128].

Personalisation systems typically fail to explain the logic behind their outputs in accessible terms. Instead of offering clear justifications, they rely on technical opacity, abstract settings, or consent flows that do not reflect actual user understanding[129]. The lack of visible logic fragments user trust and shields systems from contestation. **This undermines fairness not only at the interface level but across the broader personalisation infrastructure**[130].

Structural Rights Interferences of Unfair Personalisation

The EC's Fitness Check refers to personalisation as a tool that may optimise content and improve user experience, **while acknowledging its risks of contributing to discriminatory, misleading or unfair treatment, particularly where personalisation is**

[127] Hardcastle, K., Vorster, L., & Brown, D. M. (2025). Understanding Customer Responses to AI-Driven Personalized Journeys: Impacts on the Customer Experience. Journal of Advertising, 54(2), 176–195.

[128] Anna Marie Rezk, Auste Simkute, John Vines, Chris Elsden, Michael Evans, Rhianne Jones, and Ewa Luger. 2024. User-Centric Tensions: Exploring Perceived Benefits and (Dis)comfort in Media Personalisation. In Proceedings of the 13th Nordic Conference on Human-Computer Interaction (NordiCHI '24). Association for Computing Machinery, New York, NY, USA, Article 32, 1–13.

[129] Monzer, C., Moeller, J., Helberger, N., & Eskens, S. (2020). User Perspectives on the News Personalisation Process: Agency, Trust and Utility as Building Blocks. Digital Journalism, 8(9), 1142–1162.

[130] Wong, Y. N., Jones, R., Das, R., & Jackson, P. (2023). Conditional trust: Citizens' council on data-driven media personalisation and public expectations of transparency and accountability. Big Data & Society, 10(2).

based on automated processing and profiling[131]. Yet current regulation still treats these risks as occasional or marginal. In practice, personalisation has become a structural mechanism of exploitation in the digital economy, inseparable from surveillance-based profiling and the commercial logics of engagement and monetisation.

Personalisation is not inherently harmful, but in today's digital economy, it is most often deployed in ways that are structurally unfair: It is based on opaque profiling, aimed at maximising engagement or monetisation, and designed to exploit asymmetries of knowledge, power, or vulnerability. Even when interfaces include toggles or filters, these rarely provide meaningful influence over how personalisation unfolds, nor over the data and inferences that drive it.

Rights at Stake

Personalisation in high-risk or sensitive contexts, such as mental health apps, illustrates how inferred-trait profiling can threaten rights even when systems are framed as supportive[132]. It's important to stress that **adaptive personalisation can be unfair even if it does not involve false information or forced choices**. When systems detect user susceptibility and steer decisions accordingly, they shift the balance of power away from user agency and toward behavioural extraction[133]. This erodes the freedom to make informed choices, undermines equal treatment, and entrenches structural discrimination, especially when groupings are based on inferred traits such as mood, financial distress, or addiction.

This produces a dissonance between user expectations and actual system behaviour, creating environments where 'choice' is performative and non-correction is structurally embedded[134]. A fairness-based approach to regulation must draw a clear line between personalisation that genuinely and meaningfully supports user agency, and personalisation that uses coercion, deception, or behavioural exploitation.

For example, personalised pricing based on behavioural or demographic profiling creates unequal access to goods and services, even where people are unaware that they are being treated differentially by the trader. This undermines the principle of equal treatment and erodes market trust[135]. **Even when not based on protected characteristics, personalisation systems can replicate structural inequalities** by

131 Fitness Check Report.

132 Matthews, P., & Rhodes-Maquaire, C. (2024). Personalisation and Recommendation for Mental Health Apps: A Scoping Review. Behaviour & Information Technology, 44(10), 2389–2404.

133 Strycharz, J., & Duivenvoorde, B. (2021). The exploitation of vulnerability through personalised marketing communication: are consumers protected? Internet Policy Review, 10(4).

134 Wong, Y. N., Jones, R., Das, R., & Jackson, P. (2023). Conditional trust: Citizens' council on data-driven media personalisation and public expectations of transparency and accountability. Big Data & Society, 10(2).

135 OECD, Personalised Pricing in the Digital Era (2018). https://www.oecd.org/content/dam/oecd/en/publications/reports/2018/10/personalised-pricing-in-the-digital-era_7313c12d/db4d9c9c-en.pdf

drawing on proxies such as device type, browsing history, or location, resulting in discriminatory outcomes even in the absence of *prima facie* intent[136].

Personalisation does not simply tailor content to interests: it also shapes users' identities and behaviour over time. Platform algorithms reinforce narrow norms by promoting certain aesthetics, language, and forms of self-expression, particularly through likes, filters, and feedback loops. This is especially pronounced in younger users' experiences but applies to all users exposed to curated social visibility. Such patterns of inference-driven curation normalise conformity and suppress diversity, subtly modulating how users present themselves and interact with others in digital spaces[137].

The business logic of personalisation produces structural risks to rights as well as individual impacts. Its effects are not evenly distributed. Treatment varies depending on the user's perceived financial value, behavioural profile, or inferred vulnerabilities. Adaptive profiling used to personalise content or services may unintentionally expose users to content that is misaligned with their needs or wellbeing. Without safeguards, these systems entrench unequal access to information and services, threatening fairness, equality, and the ability to exercise rights effectively[138].

Systemic and Collective Effects

The structural nature of personalisation makes its risks cumulative. The risks depend on, and reinforce, the underlying infrastructure of surveillance: constant data collection, opaque inferences, and the commodification of attention. Even in cases where advertising is not the immediate goal, tracking enables the type of differentiated influence that personalisation depends on. The logic of unfair personalisation cannot be separated from that of unfair tracking: one generates the raw material, the other exploits it.

Many personalisation systems segment users dynamically into groups. These groupings are neither stable nor legally defined, but they are central to how influence is optimised. When personalisation is based on inferred traits, such as mood, personality, financial distress, or addiction, it curtails diversity and narrows civic space. These impacts are difficult to contest. Users rarely know which group they have been sorted into, what traits were inferred, or what logic determined their treatment. This raises serious concerns for fairness, just and equal treatment, accountability, and regulatory oversight. How can a person contest discriminatory profiling they are unaware of? **In addition to consumer rights, this undermines democratic freedoms: the right to receive diverse information, the capacity to deliberate, and the conditions for meaningful political participation.**

---

136 *Ibid.*

137 Catherine Pescott, 'Children, Young People and Online Harm: An Overview', in Faith Gordon and Daniel Thomas (eds), Children, Young People and Online Harms (Bristol University Press 2024), ch 3.

138 Information Commissioner's Office (ICO), Profiling for Content Delivery and Service Personalisation (UK ICO 2022). https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/how-to-use-our-guidance-for-standard-one-best-interests-of-the-child/best-interests-framework/profiling-for-content-delivery

Consent, even when formally obtained, offers no safeguard against such structural threats to fundamental rights. As discussed in Chapter V on Deceptive Design above, unfair personalisation can produce exclusion, inequality, or compulsive behaviour even when users have technically agreed to it. Some users are nudged toward repeated engagement or spending; others - along with their content or opportunities - are made invisible through algorithmic downranking or exclusion. Consent cannot address these systemic effects. Harms linked to unfair personalisation are not limited to exposure to specific content. These dynamics cannot be addressed through consent, because they stem from how systems are structured to adaptively steer attention, shape perception, and reinforce engagement. Such mechanisms disproportionately affect groups already subject to structural disadvantage, undermining autonomy, equality, and the effective exercise of consumer rights[139].

Profiling-based personalisation is frequently deployed to promote harmful and addictive products such as unhealthy food, alcohol, and gambling. These personalisation strategies are especially difficult to detect and regulate when embedded in influencer content, livestreaming environments, or advergames[140], and when optimised by machine learning systems that update targeting logic in real time. In such environments, users are not only unaware of how their data are used, but also face highly asymmetric commercial pressure[141].

Unfair personalisation not only distorts immediate user choice but also reinforces behavioural patterns by feeding inferred preferences back into the digital service design, creating self-reinforcing loops that limit autonomy, entrench bias, and reduce the possibility of alternative or genuinely free choices. This creates environments where users increasingly act in accordance with system predictions to avoid friction or loss of access, thereby enabling a form of spontaneous normalisation. **Over time, adaptive personalisation reshapes user behaviour to align with inferred norms, producing conformity rather than empowerment**[142]. Beyond personal harm, profiling-based personalisation can therefore curtail democratic freedoms. It compromises the capacity of individuals to encounter diverse viewpoints, make unanticipated choices, or act outside system-defined expectations. These systems not only interfere with the freedom from manipulation, but also undermine the conditions under which positive freedom – the capacity to act – can be exercised meaningfully[143].

---

139 Mansfield, Karen L et al. From social media to artificial intelligence: improving research on digital harms in youth. The Lancet Child & Adolescent Health, Volume 9, Issue 3, 194 – 204.

140 Advergames are video games created primarily to advertise a product, brand, or service, often blending gameplay with marketing messages.

141 Digital Futures for Children, Children's Online Marketing Harms: Roundtable Reflections (Digital Futures for Children 2024). https://www.digital-futures-for-children.net/events/marketing-harms

142 Mireille Hildebrandt, 'Profiling and the Rule of Law' in M. Hildebrandt and S. Gutwirth (eds), Profiling the European Citizen. Cross-Disciplinary Perspectives (Springer 2008) ch 6, pp. 45–46.

143 Mireille Hildebrandt, 'Defining Profiling: A New Type of Knowledge?' in M. Hildebrandt and S. Gutwirth (eds), Profiling the European Citizen. Cross-Disciplinary Perspectives (Springer 2008) ch 15, pp. 23–24, 41.

When personalisation is designed to optimise for engagement rather than autonomy or civic relevance, it distorts users' informational environments. This weakens their capacity to exercise reasoned judgement, especially in contexts where political content, public issues, or civic participation are involved. Algorithmic personalisation restructures how individuals access information, producing fragmented environments tailored to behavioural profiles rather than collective deliberation. This reduces shared frames of reference and narrows users' exposure to diverse viewpoints[144]. One study by the Mozilla Foundation found that 71% of the videos flagged as 'regrettable' were recommended by YouTube's algorithm, rather than actively searched for, illustrating how recommender systems shape user experience and content exposure in ways that are opaque, non-consensual, and potentially harmful to democratic discourse[145].

The risks linked to unfair personalisation often cannot be traced to a single interaction. Instead, they operate cumulatively, reshaping user environments and decision-making over time[146]. These structural effects are difficult to capture under traditional enforcement models that require evidence of individualised harm or rights violation. Firms that control the infrastructure of personalisation may use it to self-preference or exploit existing consumer profiles across multiple services. This reinforces market concentration and reduces pluralism in users' decision-making environments[147].

Ultimately, the question is not whether personalisation can be done well, but who decides what gets personalised, for whom, and to what end. **When content, options, or interactions are adapted based on inferred traits, it is service providers who shape the informational landscape, not individuals or collectives.** This is not a question of convenience or optimisation, but of power: over perception, behaviour, and consumer choice. A fairness-by-design, and thus rights-based, approach is necessary to address these systemic risks and ensure that personalisation serves consumer rights, people's dignity, and democratic participation, rather than undermining them.

---

144 European Parliamentary Research Service, Key Social Media Risks to Democracy: Risks from Surveillance, Personalisation, Disinformation, Moderation and Microtargeting (EPRS, European Parliament 2021). https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/698845/EPRS_IDA(2021)698845_EN.pdf

145 Mozilla Foundation, YouTube Regrets, July 2021. https://www.mozillafoundation.org/en/youtube/findings

146 Santos, C., Morozovaite, V., & De Conca, S. (2025). No harm no foul: how harms caused by dark patterns are conceptualised and tackled under EU data protection, consumer and competition laws. Information & Communications Technology Law, 1–47.

147 Department for Business, Energy and Industrial Strategy (UK), *Algorithms: How They Can Reduce Competition and Harm Consumers* (2021). https://www.gov.uk/government/publications/ algorithms-how-they-can-reduce-competition-and-harm-consumers

Why Existing Rules are Not Working[148]

Perceived fairness in personalisation depends on whether users understand how it works and can meaningfully intervene. Without this, trust in the system collapses[149]. **Unfair personalisation exploits power asymmetries in ways that are neither transparently disclosed nor meaningfully contestable under existing EU law.**

Many personalisation practices fall into regulatory gaps because they do not neatly match the definitions of deception or coercion under consumer law, nor do they trigger harm thresholds under data protection law[150]. Yet the impact of these practices is comparable: they steer choices, exploit asymmetries of knowledge and power, and erode user agency. This legal ambiguity highlights the need for specific rules targeting personalisation-based manipulation.

While the GDPR provides a vital legal basis for assessing the legitimacy of data processing, it is largely focused on personal data and individual rights. It does not adequately address how profiling and inference are used to dynamically adapt digital environments, nor does it regulate the effects of personalisation on autonomy, access, or fairness, especially when no clearly identifiable harm occurs at the level of a single user. **The GDPR was not designed to regulate the commercial logic of tracking-based personalisation as a structural mode of behavioural governance and market manipulation.**

Crucially, **the GDPR does not prohibit profiling *per se*.** It allows it under various legal bases, including consent or legitimate interest, and only offers specific protections where automated decision-making produces legal or similarly significant effects. But most personalisation systems operate below that threshold, subtly nudging, excluding, or manipulating users without triggering the safeguards of Article 22 GDPR. This creates a structural gap: the law focuses on the presence of data and decisional impact, while unfair personalisation often unfolds invisibly, through service adaptation, content filtering, or emotional inference.

The DSA introduced, for the first time in EU law, a baseline requirement that VLOPs and VLOSEs offer at least one recommender system option that is not based on profiling (Article 38). This marks an important step toward user control over personalisation. However, **the measure is limited in both scope and ambition.** It applies only to a small subset of services, leaving the vast majority of profiling-based personalisation systems unregulated. Moreover, Article 38 DSA does not assess how content is personalised in the profiling-based default option that most users are likely to be exposed to, nor does it address the risks that arise when profiling is used to exploit psychological traits, financial distress, or addictive behaviours. As such,

148 Zardiashvili, Aleksandre and Sears, Alan M., Targeted Advertising and Consumer Protection Law in the EU (2022). Vanderbilt Journal of Transnational Law, Vol. 56, No. 3, 2023.

149 Hardcastle, K., Vorster, L., & Brown, D. M. (2025). Understanding Customer Responses to AI-Driven Personalized Journeys: Impacts on the Customer Experience. Journal of Advertising, 54(2), 176–195.

150 Santos, C., Morozovaite, V., & De Conca, S. (2025). No harm no foul: how harms caused by dark patterns are conceptualised and tackled under EU data protection, consumer and competition laws. Information & Communications Technology Law, 1–47.

personalisation continues to operate in opaque and structurally asymmetrical ways across most of the digital economy.

Consumer law, by contrast, is expressly concerned with the conditions under which consumers make decisions, including how digital infrastructures systemically distribute exposure to manipulation and entrench power imbalances, and how market fairness is undermined. It theoretically provides a legal and conceptual basis for addressing practices that distort user autonomy, particularly when structural design features create persistent asymmetries in visibility, friction, and influence.

However, **it does not currently recognise unfair personalisation as a form of market distortion, even though it systematically distorts decision-making environments, undermines autonomy, and reinforces exclusion**. The UCPD does not address adaptive interface logic, nor does it establish red lines against the structural use of profiling to engineer friction, visibility, or influence. As a result, practices like dynamic pricing based on inferred income or willingness to pay, nudging based on inferred psychological states, or differential access to information remain largely unchallenged, despite their cumulative impact on people's rights and economic conditions.

This legal absence allows the tracking economy to frame deep personalisation as a neutral or beneficial feature, masking the underlying power dynamics. Profiling systems are optimised to identify who is persuadable, distractible, or vulnerable; yet, no existing rule addresses the (un)fairness of that logic. The GDPR regulates data use, but not what is made visible or invisible as a result of that data. The DSA regulates transparency for platforms, but not the design of personalised interaction itself. And consumer law fails to prohibit personalisation practices that prey on insecurity, exclusion, or compulsion.

The DFA is needed to fill this regulatory vacuum. Without clear rules that define when personalisation becomes exploitative, and who bears the burden of justification, harmful personalisation will continue to operate beneath the threshold of enforcement - cumulative, opaque, and unaccountable.

Proposed Policy Changes to Address Unfair Personalisation

To reduce structural incentives for addictive and manipulative design, regulators should move beyond design tweaks and tackle the underlying business model, including by banning surveillance-based advertising, which profits from maximising engagement at any cost. Eliminating this model is a necessary first step: without the constant drive to monetise attention through advertising, many of the most harmful forms of profiling and manipulation lose their commercial rationale. Yet personalisation remains a powerful tool for exploitation beyond ads, from discriminatory pricing to nudging in gaming or e-commerce. It should be considered unfair where such personalisation distorts decision-making environments or impairs user autonomy by dynamically shaping visibility, friction, or perceived relevance,

thereby reinforcing compulsive engagement, constraining refusal, or reducing meaningful control.

If digital fairness is to be meaningful, the DFA must include limits on what may be personalised, how, and for what purpose, and require that any personalisation empowers, rather than exploits, user agency. **Digital fairness cannot be achieved through choice architecture alone, but requires structural red lines, default-off settings, and a rebalancing of power between users, digital services, and the systems that mediate their interactions**.

Effective oversight of personalisation systems requires enforceable obligations on traders, not reliance on individual vigilance – which particularly harms those who are less confident in the digital environment. The burden of understanding and managing complex profiling and optimisation mechanisms cannot rest on users, particularly when those mechanisms are designed to operate invisibly. Fairness must be built into the design, operation, and governance of these systems, not simply gestured at through user-facing disclosures or consent artefacts[151].

Unfair Personalisation Begins with Structurally Exploitative Design

As explored in Chapter V on deceptive design, **structurally exploitative consent design plays a key role in enabling unfair personalisation**. The same design features that distort user choice are often used to manufacture the illusion of valid consent, giving a veneer of legitimacy to personalisation practices that would otherwise lack a lawful or fair basis[152]. **The use of personal data to adapt content, pricing, or choices to specific user profiles can create structurally exploitative environments, especially when users are unaware of the underlying logic.** This is not simply a matter of insufficient transparency or consent; it reflects a design architecture that systematically exploits behavioural predictability, nudging users toward pre-determined outcomes without meaningful control. The recommendations above therefore also apply, and further underscore the need for structural safeguards that go beyond individual consent and information provision.

Consumer Law Can Provide Structural Safeguards

These safeguards should be integrated into the existing EU consumer law framework, with the DFA providing the necessary modernisation to ensure it serves as the

[151] Competition and Markets Authority, Financial Conduct Authority, Information Commissioner's Office and Ofcom, The Benefits and Harms of Algorithms: A Shared Perspective from the Four Digital Regulators (UK Government 2022) https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/the-benefits-and-harms-of-algorithms-a-shared-perspective-from-the-four-digital-regulators; Wong, Y. N., Jones, R., Das, R., & Jackson, P. (2023). Conditional trust: Citizens' council on data-driven media personalisation and public expectations of transparency and accountability. Big Data & Society, 10(2).

[152] Santos, C., Morozovaite, V., & De Conca, S. (2025). No harm no foul: how harms caused by dark patterns are conceptualised and tackled under EU data protection, consumer and competition laws. Information & Communications Technology Law, 1–47.

appropriate legal base for addressing the economic, psychological, and emotional exploitation of consumers even where data protection rights are not directly infringed.

Make Data Deletion Meaningful: The 'Erase My Footprint' Function

The DFA should also **introduce an accessible 'erase my digital footprint' function for all users: It should require a standardised, visible, and easy-to-use deletion function that enables individuals to remove personal data and online traces from digital services they previously used**. This measure should complement the GDPR's right to erasure by ensuring practical, user-friendly implementation and an obligation on traders to guarantee full and irreversible removal, without requiring legal expertise or disproportionate effort. A simplified mechanism is especially important for children and other structurally vulnerable users, but should be available to everyone to ensure effective control over one's digital presence. **While the GDPR's right to erasure provides the legal foundation, it remains procedurally difficult to exercise and substantively narrow in scope.** The DFA should therefore introduce a standardised 'erase my digital footprint' function to make this right accessible, and truly effective for all users, especially children and structurally vulnerable groups.

A Right to Know: Transparency for All Forms of Personalisation

Traders should be required to document the purpose, logic, and impact of personalisation systems, including how relevance is determined, and which data are used. These transparency and accountability obligations must apply before deployment and form part of any Behavioural Design Impact Assessment (BDIA) or audit of fairness in digital service design, as expanded upon in Chapter VIII below.

In addition to audit and enforcement transparency obligations, **people should have a clear right to know when personalisation is used, what personal data or traits were relied upon, what optimisation goals were pursued (e.g. engagement, conversion, retention), and what effect this has on the content, interface, or price they see**. This right must apply to both static and dynamic personalisation systems, and cover personalised services, rankings, pricing, and attention architectures. Information should be presented clearly, in accessible language, and without requiring expert knowledge. To implement the user-facing right to know about personalisation, the DFA should:

- Amend Article 7 of the UCPD to clarify that concealing personalisation constitutes a misleading omission;
- Amend Article 6 of the CRD to expand the scope of personalised pricing disclosures to include other personalisation effects (e.g. content, interface); and
- Introduce a new standalone right to personalisation transparency, applicable to both static and dynamic systems, mirroring Article 27 DSA.

While transparency obligations alone have limited impact, especially when buried in legalistic disclosures, they become meaningful when paired with strong defaults (see below) and procedural safeguards. The proposed right to personalisation transparency is not intended to empower individual scrutiny in isolation, but to anchor meaningful

opt-in: **no personalisation should occur unless the user is clearly informed of what it entails and explicitly chooses it.** This ensures that transparency functions as a precondition for informed agency, not as a substitute for it.

A Right to Non-Personalised Digital Services

Today, users mostly cannot opt out of personalisation or meaningfully influence how it works. This undermines agency, particularly when commercial offers are personalised by default and without clear alternatives. Given the cumulative evidence that most people cannot meaningfully evaluate or contest the implications of behavioural profiling or consent to personalisation under current conditions, **a right to access non-personalised services must be the default.** People should have the **right to access non-personalised services** unless personalisation is strictly necessary for basic functionality or explicitly requested by the user.

This principle should also apply to recommender systems, which should also be non-personalised by default, with profiling-based personalisation offered only through meaningful opt-in. The need for such safeguards is particularly urgent in light of findings that many consumers do not understand the trade-offs of 'free' services that operate through the extraction of personal data, and that legal protections in such contexts remain inconsistent[153].

The DFA should therefore recognise users' right to access non-personalised versions of services unless such personalisation is strictly necessary, building on the safeguard established by Article 38 DSA for VLOPs and VLOSEs and extending it to all digital services, regardless of size. The CRD should enshrine this right in consumer contract law, ensuring its enforceability at the point of service provision. Traders must offer a fully functional, non-personalised version of any digital service as a default or at least as a clear, easily accessible alternative, without conditioning access solely on acceptance of profiling or payment as specified in Chapter V above regarding 'Consent or Pay' models.

As mentioned below, personalisation based on tracking, inference, or profiling should not be permitted unless the trader can demonstrate that it does not systematically distort user autonomy, reinforce asymmetries, or produce cumulative exposure to manipulation. **Where personalisation is permitted, traders must clearly disclose its use, including the logic and criteria applied, and offer users an accessible and equivalent non-personalised alternative.**

Tackle Unfair Personalisation Under the UCPD

To curb exploitative personalisation practices, the DFA should also clarify that **personalisation based on tracking, inference, and/or profiling is presumed to constitute an unfair commercial practice under the UCPD** unless the trader can demonstrate that it is transparent, necessary for the service requested, and does not

[153] OECD , "Consumer vulnerability in the digital age", OECD Digital Economy Papers, No. 355 2023. https://www.oecd.org/en/publications/consumer-vulnerability-in-the-digital-age_4d013cc5-en.htm

distort user autonomy. This applies regardless of whether a specific vulnerability is targeted: profiling-based personalisation systematically creates asymmetries of visibility, influence, and friction, which undermine fairness by design.

It should be considered unfair to personalise interfaces, prices, or offers in ways that adapt dynamically to behavioural patterns, inferred emotional states, or environmental cues, when such personalisation systematically reinforces conditions of overexposure, compulsive engagement, or constrained choice. In such cases, the design logic itself contributes to users' progressive exposure to manipulation, rather than merely exploiting pre-existing vulnerability.

The DFA should draw a clear line on two levels:

- Within the GDPR framework: personalisation based on special categories of data or functionally equivalent inferences (e.g. political views, religion, health status, sexual orientation, racial or ethnic origin) should be presumed unfair under consumer law when used for commercial or behavioural targeting purposes, reflecting their recognition under Article 9 GDPR.
- Beyond Article 9: unfairness must also cover profiling that exploits dynamically inferred psychological states, contextual conditions, or structural disadvantage (such as fatigue, impulsivity, distress, compulsive play, or low digital literacy). These do not always fall under Article 9, but they still create systemically induced vulnerability and distort decision-making environments.

Personalisation practices that rely on tracking, inference, and behavioural profiling routinely distort decision-making environments, limit user agency, and reinforce structural asymmetries. **While not all personalisation should be made unlawful *per se*, the vast majority of commercial personalisation today is deployed to maximise engagement, extract data, or manipulate choices, particularly in the context of advertising.** This includes immersive digital environments such as gaming platforms, where personalisation interacts with in-game currencies, engagement loops, or targeted offers to distort users' awareness of cost, time, or meaningful choice.

As such, the DFA should treat personalisation as a high-risk practice and apply a tiered regulatory response.

First, **the most exploitative forms of personalisation should be block-listed outright in an amended Annex I of the UCPD**. This includes:

- Personalisation based on tracking, inference, and/or profiling that distorts decision-making environments or impairs user autonomy by dynamically shaping visibility, friction, or perceived relevance. It should be considered unfair to personalise content, prices, or interfaces based on behavioural profiling, inferred emotional states, or predictive modelling of compliance, particularly where these systems intensify compulsive engagement or systematically reduce the possibility of refusal or disengagement. Designs that differentially expose people to manipulation, whether through distress loops, attention capture, or constrained alternatives, should be treated as structurally exploitative and trigger heightened regulatory scrutiny;

- Profiling or personalisation that targets emotional distress, impulsivity, or structural vulnerability to influence behaviour or spending;
- Real-time interface or pricing adaptation based on detected psychological states;
- Any system designed to increase friction for refusal, opt-out, or disengagement through personalised pathways;
- Behavioural personalisation that exploits compulsive play or spending tendencies, particularly in gamified systems where design logic mimics user intention while undermining autonomy (e.g. 'just one more click' dynamics).

Second, **the DFA should introduce a grey list of presumptively unfair personalisation practices to the UCPD**. These would include:

- Personalised content, offer, or pricing strategies based on profiling unrelated to the service's core function;
- Customised promotions targeted at individuals identified through behavioural signals as impulsive, emotionally distressed, or highly monetisable, including 'VIP' targeting in games and apps;
- Personalisation systems that simulate urgency or scarcity, especially when such cues are dynamically presented to individual users based on inferred susceptibility to pressure.

These grey-listed practices should be presumed unfair unless the trader can demonstrate, with verifiable documentation, that the system does not rely on behavioural exploitation, supports user understanding and agency; and does not result in individual or collective disadvantage or cumulative harm. **The presumption of unfairness should apply particularly in immersive or closed-loop ecosystems** such as mobile games, recommender-driven apps, or voice interfaces, where users are less able to verify what they are being offered or why.

Reform the UCTD: Ensure Fair Terms of Service, Transparent Design, and Genuine User Control

The UCTD should deem **terms of service unfair when they allow traders to unilaterally impose personalisation without transparency, user control, or alternatives**. These proposals aim to rebalance power, protect vulnerable users, and make personalisation an option, never a hidden default.

These improvements to existing consumer law would not only give legal force to the principle of digital fairness but also **support enforcement coherence, reduce fragmentation and provide clear obligations for traders**. They would ensure that the personalisation practices that are most harmful to users are treated as regulatory red lines, not as optional features managed through consent alone. Traditional metrics such as engagement or click-through rates fail to capture how personalisation shapes long-term autonomy, decision quality, or systemic discrimination.

Where profiling and personalisation shape what people see, experience, or can choose, fairness-by-design must also reflect algorithmic justice[154]. This means going beyond narrow questions of transparency or accuracy to examine how algorithmic systems allocate visibility, opportunity, and risk across society. Algorithmic justice requires actively assessing whether design and recommender systems distribute attention, opportunity, or harm in discriminatory ways. In this regard, traders must provide users with an accessible, non-personalised alternative to recommender systems. This option must be meaningful - not merely a degraded experience - and must not limit access to the services' essential features. This ensures that opting out does not amount to self-exclusion from the service.

[154] University of St.Gallen, The Human Error Project. Civil Society's Struggle Against Algorithmic Injustice in Europe, 4th March 2024 https://thehumanerrorproject.ch/wp-content/uploads/2024/03/Human-Error-Project_Research-Report-II_Civil-Society_March.pdf

VII. A Modern and Effective Enforcement Mechanism

Executive Summary

Key Recommendations:

- **Enforcement Reform**:
  - Equip enforcers with real-time oversight tools, audit powers, and access to internal testing data.
  - Allow remedies that address structural aspects of manipulation, as well as proactive investigations, not just removal of features.
  - Introduce presumptions of unfairness for high-risk design patterns to reverse the burden of proof.
  - Avoid GDPR-style one-stop-shop models: prefer decentralised, coordinated enforcement with EU-level capacity to act in systemic cases, understood as practices structurally embedded in service design or business models, affecting large user groups across jurisdictions.
- **Cross-Regime Coordination**: Enable consumer, data protection, and competition regulators, amongst others, to share key documentation, breaking down silos in enforcement.
- **Collective Redress**: Support and clarify the use of the Representative Actions Directive (RAD) for manipulation cases, enabling cross-border claims and group-level remedies.

In short, **the DFA must offer not just clearer rules, but a robust regulatory model that empowers enforcement bodies to detect and act** on systemic manipulation, shifting responsibility from individuals to the systems that shape digital life.

Regulation without enforcement is architecture without foundations. Even the most sophisticated fairness obligations will fail if enforcement remains fragmented, reactive, or dependent on individual complaints. Digital exploitation today operates at a structural level. Yet EU consumer law still relies on enforcement tools designed for discrete, one-off infringements. To be credible, the DFA must therefore be accompanied by a modern, systemic enforcement architecture capable of detecting and correcting manipulation embedded in digital service design. This requires coordinated powers, real-time oversight, and the ability to impose structural remedies that address exploitation at its source, rather than chasing symptoms after the fact.

From Reactive to Structural Enforcement

The DFA should introduce a new layer of enforceable rights and regulatory tools, capable of confronting the business models and design strategies that drive digital exploitation. At the same time, it should reinforce and complement the existing consumer law framework and the digital rulebook, **enabling coherent and synergistic enforcement across legal regimes**.

To that end, the DFA should:

- Equip enforcement authorities with strong, harmonised investigative powers, including the ability to impose design-level or systemic remedies;
- Enable enforcement bodies not only to remove individual interface features but to address systemic design strategies that generate manipulation by default.

A modernised CPC Network[155]

The DFA should **ensure decentralised, rapid enforcement through coordinated action across regulators**. A one-stop-shop model - similar to that under the GDPR - should be explicitly avoided, given the enforcement bottlenecks and other enforcement gaps it has created. Instead, any authority in the CPC Network that receives a complaint should be able to act, with binding mechanisms for joint investigations and remedies in complex or cross-border cases. To prevent paralysis or under-enforcement, the European Commission should be empowered to act directly in systemic cases, following the model of the DSA and DMA.

To deliver effective and consistent enforcement, **the DFA must thus be accompanied by a fit-for-purpose institutional design**. This means going beyond a mere update of the UCPD, CRD, and UCTD, whose enforcement mechanisms are external to the texts themselves. In practice, enforcement relies on the Consumer Protection Cooperation (CPC) Regulation and fragmented national systems, structures that have long proven inadequate in the face of complex and systemic risks and harms.

Despite the intention of the CPC Network to coordinate cross-border enforcement, it remains ill-equipped to address the scale and automation of today's structurally exploitative practices. Many national authorities operate with limited mandates, insufficient resources, or lack strategic independence, undermining their ability to investigate and sanction dominant digital players. The CPC framework focuses on halting individual infringements rather than addressing the structural design choices and business models that generate harm. In practice, coordination remains *ad hoc*, with no binding obligations to act or prioritise emerging risks, and few mechanisms for joint investigations or structural remedies. Transparency is limited, and the absence of formal engagement with civil society further weakens accountability.

[155] EDRi has consistently opposed disproportionate or discretionary enforcement powers, such as the ability to shut down websites without judicial safeguards. Our call for stronger coordination and systemic enforcement in the DFA must be understood within a rights-based framework: one that enables accountability for structural threats to rights while upholding due process, transparency, and fundamental rights. Strengthening enforcement must not come at the expense of procedural guarantees.

The DFA must address these shortcomings directly. It should **establish binding coordination and joint action mechanisms within the CPC Network, require EU-wide minimum standards for national authorities, and empower a central EU body to lead systemic investigations where necessary**. A dual-track enforcement model, inspired by the DSA, could ensure that cross-border unfair practices - which in practice often amount to violations of fundamental rights - are not left to under-resourced national authorities alone. Without standalone enforcement provisions or significant reform of the CPC framework, the DFA risks reproducing the same fragmented and reactive enforcement model that has failed to tackle systemic manipulation to date.

Equip Enforcers with Real-Time Oversight and Evidentiary Leverage

To meaningfully address structurally exploitative design and other systemic threats to rights, enforcement authorities must be **empowered with real-time oversight tools and updated evidentiary standards**. This includes, as mentioned above, mandatory auditability of design systems, access to internal A/B testing results, and the ability to compel disclosure of design rationales, optimisation metrics, and behavioural targets. Some digital markets - such as finance, education, or health - entail greater risks of manipulation or harm. The DFA should ensure that enforcement reflects sector-specific risks and applies proportionate scrutiny where stakes are higher.

Regulators should also be supported with **cross-border investigative capacity, including joint enforcement mechanisms and shared evidence repositories, to tackle structurally exploitative practices that rapidly adapt and proliferate across jurisdictions**.

An **accessible and user-friendly reporting system is also needed to support enforcement**. Such a tool should allow individuals to flag problematic websites or services, describe structurally exploitative elements, and indicate how their rights or choices were undermined. To be effective, it must be integrated into national and EU-level enforcement workflows so that patterns of abuse can be identified and addressed, especially in cross-border contexts where unfair practices linked to consumer rights violations are difficult to trace.

Effective enforcement of the DFA also depends on mechanisms that reflect the reality of how people experience digital disempowerment and harms which ultimately disproportionately limit their fundamental rights and freedoms. Regulatory responses should be informed by input from affected communities, with processes that go beyond formal infringements to capture broader patterns of harm. This requires strengthening the role of civil society in enforcement, reducing the sole reliance on individual complaints in contexts where harm is diffuse, and enabling regulators to engage with systemic issues more proactively.

Last but not least, and also as a lesson from GDPR enforcement, to ensure meaningful deterrence and effective redress, **enforcement must move beyond reactive fines to recognise the systemic nature of digital manipulation. Penalties should reflect not only the severity of individual breaches, but also the scale, persistence, and**

structural embedding of exploitative practices within a digital service. Penalties should thus not be limited to monetary sanctions but should include design-level or systemic remedies, such as mandatory design changes, suspension of exploitative features, or limits on profiling.

In addition to monetary sanctions, the DFA should mandate **structural remedies, such as the reconfiguration or removal of structurally exploitative elements; rollback of unlawful personalisation systems or profiles; or obligations to notify and compensate affected users**. Where traders have deployed systems that undermine user autonomy or create environments that structurally expose people to manipulation, regulators should be empowered to demand design reconfiguration as part of the corrective action. **Public disclosure of such enforcement decisions, including required design changes, should be promoted to increase transparency, create industry-wide deterrence, and support regulatory convergence across sectors**.

Breaking Silos: Coordinated, Cross-Regulatory Enforcement

Digital fairness must be understood as a cross-cutting legal principle that draws from and reinforces consumer protection, data protection, digital, and competition law. It should provide a unified framework for addressing systemic imbalances in digital markets, particularly where commercial practices exploit attention, personal data, or behavioural vulnerabilities to manipulate outcomes.

Effective enforcement also requires systematic cooperation and information-sharing across regulatory regimes. Empirical data from multiple jurisdictions show that market incentives do not self-correct structurally exploitative design. In competitive environments, traders that attempt to respect user autonomy often suffer lower engagement or conversion metrics, leading to a 'race to the bottom' dynamic[156]. This underscores the need for regulatory baselines that eliminate exploitative strategies across the board.

Structural threats to rights identified in this paper span multiple legal domains, yet enforcement remains siloed. Authorities often lack legal pathways to share or access relevant documentation. For instance, profiling reports required under Article 15(4) of the DMA contain critical information about the logic and goals of personalisation systems but are inaccessible to consumer protection authorities. Similarly, Data Protection Authorities may hold DPIAs or audit results that could support consumer enforcement but cannot share them due to confidentiality rules.

To close these gaps, **the DFA should explicitly enable cross-regulatory cooperation - not only information exchange, but also joint investigations and coordinated enforcement actions - between consumer, data protection, competition, and digital regulators, and potentially other regulators and enforcement networks**. Authorities must be able to share and use relevant documentation, including (but not limited to)

[156] Behavioural Insights Team, The behavioural science of online harm and manipulation – and what to do about it, March 2022. https://www.bi.team/publications/the-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it

profiling reports under the DMA, DPIAs under the GDPR, and algorithmic documentation under the DSA. Such cooperation should be governed by purpose limitation, proportionality, and secure exchange protocols, but broad enough to let enforcers reach the design logics and systemic incentives that drive digital unfairness.

Embedding this cross-regime enforcement model within the DFA would make it possible to tackle manipulation as both consumer harm and a structural market imbalance, bridging currently fragmented oversight. To avoid repeating the enforcement failures of past digital legislation, the European Commission should treat joint, well-resourced enforcement as a political priority, strengthening cross-border investigations, supporting under-resourced authorities, and ensuring that regulatory cooperation is mandatory rather than voluntary.

To reinforce systemic enforcement, the DFA should also facilitate coordinated EU-wide action. This includes introducing clear triggers for joint investigations, mechanisms for pooling evidence across Member States, and enabling collective redress procedures that allow affected individuals or groups to surface structural patterns of harm. Empowering collective enforcement not only enhances access to justice but also supports regulators in detecting and addressing widespread manipulation that might otherwise remain invisible at the national level.

Ensuring jurisdictional access for all individuals and civil society actors

The DFA should **build on the Representative Actions Directive (RAD) to support collective redress in cases of structurally exploitative design, exploitative personalisation, and other systemic unfair practices.** These rights violations often affect large numbers of people in diffuse and subtle ways, making individual complaints ineffective. By enabling qualified entities to bring cross-border actions and pool evidence across Member States, RAD can serve as a vital enforcement complement also for consumer law. However, this potential remains underused, and even hindered in some Member States. The EC and Member States should take active steps to support the use of RAD in the digital context through guidance, funding, and by clarifying its relevance to emerging patterns of online exploitation.

As digital services increasingly involve ongoing, personalised, and cross-border relationships, the question of where and how people can enforce their rights becomes essential. The DFA should include a **dedicated jurisdiction clause, modelled on Article 79(2) GDPR, allowing any natural person to bring legal proceedings under the DFA before the courts of their domicile.** This provision is crucial to ensure that individuals, including those with hybrid or dual-use roles, are not excluded from redress due to outdated consumer definitions in Brussels I bis Regulation (the EU's main framework governing jurisdiction and civil procedure). The clause should also cover civil society organisations and small enterprises, which often lack the resources to litigate in a trader's home jurisdiction. Including this jurisdictional right would make the DFA more

enforceable in practice, and could serve as a model for improving access to justice across the digital rulebook, including in the DSA, DMA, and AI Act.

***