Part three:
A regulatory framework for online platforms

6 The Digital Services Act and the Codes of Practices

6.1 Prologue to DSA: the legacy of the E-Commerce Directive

The Digital Services Act is a complex legal rule with multiple aims. Its primary aims include regulating intermediary services in the internal market and ensuring that the online environment is safe, predictable, trusted, facilitates innovation, and respects fundamental rights.[598] Societal impacts on the informational landscape by DSA are derived indirectly from its systemic approach. Over the past decade, it has been observed that very large online platforms and very large online search engines are capable to strongly influence the shaping of public opinion and discourse, besides online trade and safety. Negative effects on democratic processes, civil discourse and electoral processes hold the third place out of the four risks mentioned in the Recitals.[599] A free and competitive market and a free and diverse informational landscape are the two basic pillars of libertarian market economies, therefore, also of liberal democracies.[600] Fundamental rights are treated as counterweights to the predominance of pure market logic, protecting both market plurality and societal diversity with the same move.[601]

The regulation of platforms has long been on the agenda of the European Union. (And their business-to-business applications has already been

598 Article 1 (1) DSA.
599 Recital 82 DSA, Article 34 (1)c DSA.
600 Eifert, *Taming:* 993.
601 Giovanni De Gregorio and Pietro Dunn, "The European risk-based approaches: Connecting constitutional dots in the digital age," *Common Market Law Review* 59, no. 2 (2022).

regulated in 2019 for transparency and fairness.[602]) Platforms have become primary vehicles of information that is distributed to the public eye. Their market role dwarfed that of media service providers, and their impact is overshadowing television.[603] Especially the giant platforms are flagships of the disruptive transformation of the media landscape, and those that altered the information environment.

Even in the pre-platform internet age, enforcing the old legal rules on publication, let alone journalistic and ethical standards, appeared overwhelming in this new environment. The first legal cases were characterised by a coming to terms with the sheer volume of content and the rivalling interpretations regarding liabilities for content.[604] The pressing need to settle the issue of liabilities has led to the passing of rules that exempted intermediaries from the liability for content. First the United States had passed its provision CDA §230 (1996), then the EU has passed the E-Commerce Directive (hereafter: ECD) in 2000.[605] However, at this time, online platforms were not yet meaningful services. The peer-to-peer technology was only about to start its march which later changed the entire nature of how the internet was used. This technology was truly revolutionary: it opened up the internet to ordinary users, and made online communication interactive even for lay people, who had no clue about how html or other coding languages operate. Platforms are, on the one hand, a simple user-interface: like the light switch, they allow communication with the ease of moving a finger. On the other hand, their service includes more than just performing user-initiatives. Platforms added their service of aggregating and ranking content, thereby tailoring user's information consumption.

The E-Commerce Directive defined three layers of intermediaries: the mere conduit – which transmits the information; the caching provider – which transiently stores information only to make transmission more effi-

602 Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services.

603 More than 80 % of Americans get their news from digital devices, vs. 68 % from television, according to the Pew Research Center. See: Shearer, E. (2021) "More than eight-in-ten-Americans get news from digital devices". https://www.pewresearch.org/short-reads/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/.

604 ACLU v. Reno, Yahoo v. France, Godfrey v. Demon, etc.

605 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce').

cient; and the hosting provider – which hosts third party content. One of the main goals of ECD was to exempt these intermediaries from liability so that they do not obstruct the free flow of information and services through the internet.[606] In the case of hosting providers, this exemption depended on the condition that they were truly non-cognizant of the information. Should they get knowledge about illegal content, they were required to diligently remove that, or lose their immunity. The rule did not provide detailed procedural framework, but as a directive, it allowed space to Member States to pass their own detailed rules.

Platforms, which emerged around 2004 and quickly became so popular that they transformed internet economy and communication, did not fit into this structure of actors.[607] Their services combined hosting and transmitting, and with time, more and more content organising with the help of opaque algorithms.[608]

The ECD's literal interpretation disallowed such a combined interpretation of services, and therefore it could not be applied to platforms. Platforms, of course, tried to take avail of the immunity protection of Article 14 (1) ECD, which exempted intermediaries from liability for the hosted content, if they fulfilled two conditions: first, that they had no actual knowledge about the illegal activities or information that was stored through their services, and second, that upon obtaining such knowledge, they expediently removed or disabled access to that information. In the court case Loréal v. eBay,[609] the CJEU found that this exemption may only apply if a service operator did not play an active role allowing it to have knowledge of the information stored. However, it found that eBay played a sufficiently active role when it optimised the presentation of the offers for sale. Thus, the ranking activity of eBay was interpreted as an action that deprived it from immunity. Even though eBay probably did optimise by way of an algorithm rather than by a human employee who would have had the possibility to acquire actual knowledge about the violations of law

606 Rosa Julià-Barceló and K. J. Koelman, "Intermediary liability: intermediary liability in the e-commerce directive: so far so good, but it's not enough," *Computer Law & Security Review* 16, no. 4 (2000): 231–239.

607 Tim O'Reilly, "What is web 2.0. Design Patterns and Business Models for the Next Generation of Software," 09/30/2005. https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html.

608 Fuchs, C. (2011). Web 2.0, prosumption, and surveillance," *Surveillance & Society* 8, no. 3 (2011): 288–309.

609 C-324/09. Loréal v. eBay, 2011 I-06011.

(trademark law, in that specific case). Even today, it is the ranking activity that raises unsolved questions: even without direct access to the content, platforms carry out systemic governance of how content is perceived by the public. Already this first decision has signalled that there should be a third, interim type of liability structure, somewhere between liability for the content, and the full immunity of an intermediary who has a merely technical relationship to the content.[610] At about the same time, another international court procedure dealt with the issue of liability for third party content: the Estonian online journal "Delfi" claimed to be immune from liability for hate speech in their comment section. This leading online journal reported on a matter of public concern: that a certain ice route at sea that was awaited to get frozen, would remain unsuitable for longer, because a certain ferry company broke up the ice with its ships. Emotional public reactions flooded the comment section, which amounted to antisemitic hate speech targeted against the head of the ferry company. Once Delfi was notified about the hate speech, it expeditiously removed the incriminate content, the same day – six weeks after the original publication.[611]

The court of first instance found that Delfi's liability was excluded under the Estonian Information Society Services Act, which implemented ECD. It found that Delfi could not be considered the publisher of the comments, nor did it have any obligation to monitor them, and that the administration of comments was essentially of a mechanical and passive nature. However, the court of second instance quashed this judgement and sent the case back to the court of first instance for new consideration. It instructed the lower court to rely on the Obligations Act and ignore the Information Services Act. In the second procedure both the lower court and the Tallinn Court of Appeal held that Delfi was not a technical intermediary, and that its activity was not of a merely technical, automatic and passive nature. They argued that Delfi invited users to post comments and gained extra revenues from the vivid commenting section. The court observed that Delfi had indicated on its website that comments were not edited and that the posting of comments that were contrary to good practice was prohibited. The portal reserved the right to remove such comments, and it did exercise

610 See more in: Marta Maroni and Elda Brogi, „Freedom of expression and the rule of law: the debate in the context of online platform regulation," in: *Research handbook on EU media law and policy*, ed and Pier L. Parcu Elda Bogi (Cheltenham, UK/ Northampton, USA: Edward Elgar Publishing, 2021): Chapter 8.

611 Case of Delfi v. Estonia, no. 64569/09, Judgement of 16 June 2015.

this right in cases like the "Bronze Night", when a relocation of the Bronze Soldier Monument caused public unrest, and Delfi removed between 5,000 and 10,000 comments on its own initiative within a day. The court also noted that Delfi had a system where users could notify of inappropriate comments but found this insufficient. It found that Delfi should have created some other effective system to ensure rapid removal of unlawful comments, even without notice. The argumentation sounded very similar to the court decision in the Prodigy case, after which the "Good Samaritan" provision was inserted in the American Communication Decency Act § 230.[612]

In sum, Delfi was found to be the publisher of the comments, and accordingly responsible for them. Delfi submitted a complaint to the European Court of Human Rights (ECtHR) which delivered one of its most controversial decisions at the time. It found that Article 10 was not violated because the hate speech in question did not deserve protection.[613] It failed to examine the question whether Delfi was a speaker or a carrier of the content? One can only wonder why Delfi requested protection from the ECtHR, when in fact they did not regard themselves as publishers of the incriminate content? They could have asked for a preliminary decision in the court, to inquire whether their responsibility is governed by the Estonian Civil Code or the ECD. Even in this case, the outcome of the case would not have been certain, as Delfi, due to its regular moderating activity, could have been regarded by ECJ as well, as having imputed constructive knowledge.[614] And, considering the level of controversy of the topic, and the type of comments in the case (hate speech) it could have been expected to diligently remove them. Although, in contrast to eBay (in the case Loréal v. eBay), Delfi did not organise the comments, merely moderated them. Moreover, the often-cited requirement that intermediaries' activity should remain of a mere technical, automatic and passive nature, is based solely on Recitals (42, 43) of ECD, and are not found in the text. Van Eecke argued that Article 14 did not require this passive role for the protection to apply, merely that the provider does not have knowledge or control over the data

612 Stratton Oakmont Inc v Prodigy Services Co (1995) 23 Media L Rep 1794 (NY). See more in Bayer, 2007, p. 21. Liability of Internet Service Providers.

613 Case of Delfi v. Estonia, no. 64569/09, Judgement of 16 June 2015.

614 Peggy Valcke, Aleksandra Kuczerawy, and Pieter-Jan Ombelet, "Did the Romans get it right? What Delfi, Google, eBay, and UPC TeleKabel Wien have in common," *The responsibilities of online service providers*, (2017): 101–116.

which is being stored.[615] As a result of the Delfi case, commenting sections were disabled in several member states, or bound to registration.

As seen, a clarification of the status of platform services was painfully missing since 2004 until the DSA was passed. The lack of monitoring obligation and the expectation of acting as a diligent economic operator"[616] led to contradicting interpretations in courts and among scholars.[617]

During this phase, the question of responsibility became particularly pressing when disinformation, hate speech and political propaganda were found to make an impact on democratic societies in 2015–2016.[618] Reacting to this regulatory gap, several states passed laws against disinformation or hate speech. Diverging national legislation would build up obstacles within the internal market and would hamper the freedom to provide and receive services throughout the union.[619] The emerging fragmented regulation of cross-border services was another strong incentive for common EU legislation. The goal to achieve uniformity and to avoid fragmentation is reflected in several instances related to DSA, among others, choosing the instrument of a Regulation that prevents national divergences that may arise through implementation.

6.2 Aims, scope and structure of DSA: more than just services

DSA is regarded as the new framework for online services that partly replaced and partly completed the ECD. Its main mission was to regulate platform services. Like ECD, it aimed to ensure the free movement of services, of establishment, as well as the free reception of services across the borders.[620] Nevertheless, a second main aim has also been added: to set out uniform rules for a safe, predictable and trusted online environment, where fundamental rights enshrined in the Charter are effectively protected.[621]

615 Paul Van Eecke, "Online service providers and liability: A plea for a balanced approach,". *Common Market Law Review* 48, (2011): 1455–1502.
616 L'Oréal v. eBay, para. 120.
617 Valcke, Kuczerawy and Ombelet, "Did the Romans", 101–106.
618 Samuel C. Woolley, and Philip N. Howard, *Computational propaganda: Political parties, politicians, and political manipulation on social media.* (New York, NY: Oxford University Press, 2018.
619 DSA, Recital 2.
620 Article 1 (2a) DSA.
621 Article 1. (2b) DSA.

This signals that the issue of intermediary liability has grown into a pressing question of content regulation. To address the systemic significance of platforms, "uncharted regulatory waters"[622] were entered by the legislator, lowering the threshold of intervention compared to more traditional forms of regulation. Yet, the regulation seems progressive on the one hand, but also at the shoulder level on the other hand: it appears to accept the position of platforms as market regulators, and creates rules of supervision and transparency to keep track of and to contain any event that may have a systemic significance.[623]

For the purposes of this book, our focus is on the second set of aims: a safe, predictable and trusted online environment with effective protection of fundamental rights. The Regulation puts an indirect obligation on providers of intermediary services to respect the applicable fundamental rights of the users as enshrined in the Charter.[624] By imposing this obligation on providers of intermediary services, the eventual dilemma whether horizontal effect of human rights applies, becomes obsolete. Being a Regulation, the instrument is capable of imposing rights and obligations directly on legal subjects under its jurisdiction. Hence, the protecting and ensuring of fundamental rights will become a direct legal obligation under European law, without the necessity to reach out to international human rights law.[625] In fact, according to the Charter, all parts of European law must be applied in harmony with the Charter,[626] but that obligation applies to Member States and official bodies of the Member States,[627] and does not explicitly oblige private parties.[628]

With this being said, the DSA's rules are still very general on fundamental rights protection. It is left for the private parties, and to the national authorities to interpret the vague obligations. Nevertheless, Recital 41 lists a number of specific human rights to be ensured: the right to freedom of

622 Martin Eifert et al., "Taming the giants: the DMA/DSA package," *Common Market Law Review* 58, no. 4 (2021): 987–1028. at page 994.
623 Eifert, "Taming the giants", 987–1028.
624 Article 14, 34 DSA.
625 Judit Bayer, "Rights and duties of online platforms," in: *Perspectives on Platform Regulation*, ed. Judit Bayer et al. (Baden-Baden: Nomos, 2021).
626 Article 52 of the Charter of Fundamental Rights of the European Union.
627 Article 51 of the Charter of Fundamental Rights of the EU.
628 Although, see: Case C-176/12 Association de médiation sociale. " may be invoked to request the disapplication of conflicting national provisions even in proceedings between private parties".

expression and information, the right to respect for private and family life, the right to protection of personal data, the right to non-discrimination and the right to an effective remedy of the recipients of the service; the freedom to conduct a business, including the freedom of contract, of service providers; as well as the right to human dignity, the rights of the child, the right to protection of property, including intellectual property, and the right to non-discrimination of parties affected by illegal content.[629]

Thus, while DSA aims primarily to regulate market relations, it has a clear objective to ensure non-commercial values within its scope of application. It explicitly sets out the protection of the online environment as a goal, even though that is not among the explicit competences of the European Union, and refers to the Charter as a guiding pole for actions of market actors.

6.2.1 Scope

The personal scope of DSA is very similar to that of ECD: intermediary service providers.[630] The basis of defining services remains "information society services" as defined in Article 1(1)(b) of Directive (EU) 2015/1535. In contrast to the ECD, DSA gives a very brief definition of mere conduit, caching service and hosting service in its Article 3 (g), however, the wording of the provisions that exempt these service providers from the liability remained almost literally the same.

6.2.1.1 Quo Vadis, Platform?

Rather than creating a new category for platform providers, they were simply added as a subcategory of hosting providers. The definition of platforms explains that beyond storing information at the request of a recipient of the service, they also *disseminate* information at the request of the recipient of the service.[631] However, does the action "disseminate" truly reflect the content governance that platforms pursue? This definition fails to grasp the gist of the core service that online platforms perform in addition to hosting: namely, *ranking* the content, and thereby exercising a *formative*

629 Article 34 (1)b.
630 Article 2 DSA.
631 Article 3 (i) DSA.

effect on the information offer (whether it is goods, services or news, etc.). This gentle "invisible hand" in ordering the published content has been the very activity that gave rise to policy concerns. Whenever online platforms are accused of inflicting harm upon human rights or democracy, it is not *because* they disseminate content, but *how* they disseminate it. Dissemination as such is also defined separately: 'dissemination to the public' means making information available, at the request of the recipient of the service who provided the information, to a potentially unlimited number of third parties.[632]

The definition clarifies that both storage and dissemination take place "at the request of a recipient of the service". This implies that platform's ranking activity is not part of their core activity that makes them what they are, or not part of their "service." It leads us to the conclusion that ranking is really an activity that serves the interests of platforms, rather than that of users. Indeed: ranking is not made "at the request" of a user, and is therefore not part of the service, at least not an indispensable part of it. It could rather be regarded as a feature that influences the quality of their service, which makes platform services more engaging and above all, more lucrative.

At the same time, this feature was the reason for depriving eBay from the immunity in the court case of 2011.[633] Would the DSA repeat the exact same mistake as the ECD? Would a platform that does more than simply disseminate to the public at the request of the user, for example by upranking, factchecking or deprioritising the post, lose immunity for the content? Even if this seems plausible from the literal interpretation of the law, the answer is almost certainly no. Primarily, because the definition does not refer to the technical and passive, automatic processing of the information, as ECD did, and this also allows a more inclusive interpretation, even the former words are repeated in Recital 18, which withdraws the exemption from those providers which place an active role that gives them knowledge of, or control over that information. However, while the so-called Good Samaritan provision may seem relevant at first thought, it does not exclude liability for ranking. Its immunisation is limited to measures that serve

632 Article 2 (k) DSA.

633 Hoboken, J.v. (2011) Legal victory for trademark litigants over intermediary liability. EDRI. "The exemption applied only to third party data processing that is merely technical and automated, as well as passive and neutral. In the Court's view, an online market place is not passive enough if "it provides assistance which entails, in particular, optimising the presentation of the offers for sale in question or promoting them."

to detect, identify and remove or disable access to illegal content, and it does not include the activity of ranking, organising, recommending, etc.[634] Even though platforms can follow explicit requests, for example when users can select the audience of their posts, dissemination is an action that is significantly tampered with, by way of automated algorithms that organise content. Platforms do still play an active role "optimising the presentation of the offers for sale in question or promoting those offers, it must be considered not to have taken a neutral position between the customer-seller concerned and potential buyers", as objected in the case Loréal v. eBay.[635] The DSA relies on the same liability exemption based on the assumption that intermediaries play a passive, neutral role, whereas platforms engage in excessive content governance.[636]

Nevertheless, "ranking" does not necessarily constitute knowledge, because it is done by algorithms, and often even based on metadata.

Finally, omitting the act of ranking from the definitional elements widens the scope of definition to include platforms which do not employ ranking or organising, but simply list their content randomly, or by chronological order of publication, or some other user-chosen logic. Whether they apply ranking or not, they still qualify as platforms.

6.2.1.2 Who else are not platforms?

The definition of online platforms excludes those service providers whose such activity is of a minor or a purely ancillary feature of another service, or a functionality of the principal service, rather than an independent service (as it cannot be used without that other service). This effectively excludes commenting sections of journals and other hosting services from the scope of online platforms, and therefore does not impose the same obligations on them. This leaves the status of online journals like Delfi, further unsettled. They are still not platforms in regard of the user comments, and Recital 13 explicitly holds that comment sections allowing to store and disseminate information outside of editorial control do not turn online newspapers automatically into online platforms for the purpose of applying the DSA.

634 Article 7, DSA.

635 Case C-324/09, 12 July 2011 (GC) at 116.

636 Miriam C. Buiten, "The Digital Services Act From Intermediary Liability to Platform Regulation," *Journal of Intellectual Property, Information Technology and Electronic* Commerce *Law* 12, (2021): 361.

However, now they can take avail of the Good Samaritan provision, which provides that voluntary self-investigations or other measures to comply with the law, should not be deemed as excluding immunity for providers of intermediary services.[637] At the same time, for platforms, such activity is mandated by the DSA as a risk mitigation method (see later in more detail).[638]

Services that are not strictly information society services are excluded from the scope of DSA. In the light of CJEU's case law, Uber's UberPOP application qualified as a transport service, rather than an information society service.[639] In AirBnB Ireland, CJEU held that a case-by-case assessment approach needed to be applied.[640] Additionally, the amended Recital 14 of the DSA clarifies that information exchanged using interpersonal communication services, such as emails or private messaging services, are also not considered to have been disseminated to the public. The original draft wording that they fall outside the scope of the Regulation, has been removed for the final version (Recital 14 or original DSA).

6.2.2 Territorial scope

As regards territorial scope, the relevant element for EU jurisdiction is residence or place of establishment of the recipients of the service, irrespective of the establishment of the providers. Similarly to GDPR's territorial effect, any service provider which provides services to users who reside within the EU, are subject to the law and should comply with its requirements.

The country-of-origin principle generated fierce debates during the legislation process.[641] Ireland, which is currently the homeland to Apple, Google, Twitter, Microsoft and Facebook, has collected a pool of ten Member States to submit an opinion against ditching the country-of-origin princi-

637 Article 7 DSA.

638 Article 34 DSA.

639 Case C-434/15, Asociación Profesional Elite Taxi, EU:C:2017:981; Case C-320/16, Uber France, EU:C:2018:221.

640 Case C-390/18, Airbnb Ireland, para. 64. and Pieter Van Cleynenbreugel, "The Commission's digital services and markets act proposals: First step towards tougher and more directly enforced EU rules?" *Maastricht Journal of European and Comparative Law* 28, no. 5 (2021): 667–686. https://doi.org/10.1177/1023263X211030434.

641 Luca Bertuzzi, "Ireland draws a red line on country of origin principle in DSA," *Euractiv* last modified 2021 szept. 29. https://www.euractiv.com/section/digital-single-market/news/ireland-draws-a-red-line-on-country-of-origin-principle-in-dsa/.

ple. They argued that the destination principle would create disproportionate burden to SMEs to comply with 27 different jurisdiction.[642] At the same time, the past years have demonstrated that authorities in the country of establishment are not necessarily equipped with every resources to deal with the problems of all Member States.[643] This practically allowed forum-shopping for market players, who are able to establish their headquarters in any a state with the most favourable legal rules, regardless of the capacities of the national authority. This market-liberal approach has in practice resulted an enforcement bottleneck.[644] Even after equipping one authority, the platform might even change its seat again with little investment compared to the public investment to its supervision. Moreover, the authority of the establishment is less likely to be adequately responsive to public sentiments in the different cultural environment of the other Member States in which the services are received.[645] In spite of all the arguments, the final version of the DSA relies on the country-of-origin principle, following the structure known from the GDPR. Under the GDPR, the data subjects may turn to their local authorities, which is obliged to transmit the case to the authority of establishment. Under the DSA, the Digital Services Coordinator of destination may, if it has reason to suspect that a provider violates the regulation, request the DSC of establishment to assess the matter, and to take the necessary measures. The Coordinators are obliged to provide mutual assistance, and the Board will coordinate or mediate if necessary.[646] Just like under GDPR, eventual conflicts between the authorities would be solved through a system called consistency mechanism.[647] A further element of decentralisation has been implanted through the definition of what is "illegal" which opens the door for national divergences. In contrast

642 Croatia, Czechia, Estonia, Finland, Latvia, Lithuania, Luxembourg, Slovakia and Sweden signed a non-paper "on the effective supervision under the Digital Services Act." D9+. (2019) D9+ Non paper on the creation of a modern regulatory framework for the provision of online services in the EU. Warsaw. Retrieved from https://www.gov.pl/attachment/dc1d7068-caf3-4a1b-b670-0f2f568e84c4.

643 Krisztina Rozgonyi, "Negotiating new audiovisual rules for Video Sharing Platforms: proposals for a Responsive Governance Model of speech online," *Revista Catalana de Dret Públic* 61, (2020): 83–98. https://doi.org/10.2436/rcdp.i61.2020.3537.

644 Sebastian Heidebrecht, "From Market Liberalism to Public Intervention: Digital Sovereignty and Changing European Union Digital Single Market Governance," *JCMS: Journal of Common Market Studies* (2023).

645 Eifert, "Taming the giants," 1021.

646 Recital 128–129, Article 57 DSA.

647 Articles 63–67 GDPR, Article 56 DSA.

to the AVMSD, where merely the legal system of the country of origin would serve as a basis to decide on legality of content, under the DSA, any Member State's laws can serve as a basis of illegality.

This can potentially lead to some practical problems of implementation, as well as overrestriction of content. On the ground of the CJEU decision Glawischnig-Piesczek vs Facebook, we may assume that orders to remove or block access to information that had been deemed illegal, or any equivalent content to that, can have a global reach. Barata argues that DSA grants national authorities an almost discretionary power to unilaterally impose their standards on third countries.[648] At the same time, the Commission has a central role in the enforcement of DSA, in particular in assessing the compliance of those obligations that apply exclusively to VLOPs.[649]

It will remain an open question whether the ECD's remaining provisions will further oblige only those service providers who are settled in the EU – a more conservative approach to territorial scope. Given the unresolvable consumer complaints that emerge in relation to transactions with extraterritorial effect, e.g. Chinese e-commerce businesses, the more ambitious territorial approach would be more beneficial for EU consumers. On the other hand, several businesses may disable their services for EU residents, as it has happened with some of the service providers (typically: news sites) in connection with the GDPR.

6.2.3 The structure of DSA

The ECD will remain in force, with the exception of the provisions relating to liability of service providers that are absorbed by DSA. The remaining provisions – that are not regulated in the DSA – are the ones that have laid the grounds of the European online environment and have significantly contributed to the development of a trustworthy online market in the EU. These provisions require states to refrain from demanding authorisation of online services; to oblige providers of online services to publish basic information at their websites to inform the consumers, such as contact data; and to define basic rules of online contract, with the aim to protect consumers; as well as to rule out spam (through allowing the sending

648 Joan Barata, "The Digital Services Act and its Impact on the Right to Freedom of Expression: Special Focus on Risk Mitigation Obligations," *Platforma por la Libertad de Infoacion* 12 (2021).

649 Eifert, „Taming the giants,".

of unsolicited commercial communication only with prior consent of the targeted person). This set of rules created a safe environment for online commerce, and a protection for consumers in the EU that has been uncommon in other parts of the world.

DSA has three major parts: one that replaces existing liability provisions in the ECD (Articles 12–15 ECD; Articles 4–10 DSA); a part on due diligence obligations for hosting providers, including online platforms (Chapter III., Articles 11–48) and a third part on implementation and enforcement (Chapter IV, Articles 49–88). Before discussing these in detail, the next chapter provides a brief overview of DSA's structure.

The first part by and large corresponds to the original ECD provisions, but is more elaborated. Beyond the basic rules (Articles 4–15) which discuss liability rules, corresponding to the logic of ECD, new rules are added in Chapter III that add more detail and due diligence obligations (Chapter III, Section 2, Articles 16–18). These rules serve the protection of users' fundamental rights – without defining them, by reference to the Charter –, such as transparency of online services in general and of the notice-and-action mechanism in particular. The following Chapter III Section 3, in its entirety, can be regarded as a separate logical unit that applies specifically to online platforms, rather than to all hosting providers (Article 19-48). The rules provide further details of the notice and action system, including details on the procedure (exclusion of micro and small enterprises, internal complaint-handling system, trusted flaggers, measures and protection against misuse, Article 16–23). In addition, it sets out platforms' obligations after the action (out-of-court dispute settlement, notification of suspicions of criminal offences, Articles 21 and 18). Further transparency obligations apply to reporting on the notice procedures, and the placement of advertisements (Articles 24, 26), recommender systems and the online protection of minors (Article 27–28).

With this, the range of relatively concrete and compulsory material obligations of service providers comes to an end.[650] The most characteristic part of DSA, the discussion of very large online platforms and the systemic risks are addressed in Section 5. The following part elevates the legislative

650 Specific obligations of online trading platforms are relatively briefly discussed in Section 4. These are platforms that allow consumers to conclude distance contracts with traders, such as Amazon, for instance, and have obligations such as the exclusion of SMEs, traceability of traders, the obligation to provide online interfaces that are designed to comply with the legal obligations (compliance by design); and consumers' right to information about illegal products. Article 29–32 DSA.

method of due diligence responsibility to a new level: the goals are even more vaguely defined and the route shall be entirely designed by the legal subjects themselves; and this applies only to very large online platforms (VLOPs) and very large online search engines (VLOSEs; hereafter together: VLOPs). The risk management method is supposed the tackle the complex issues with the information society environment and establish co-regulation, and will be analysed in more detail below.

VLOPs are those platforms which have at least 45 million average monthly active recipients, and which are registered as such by the Commission.[651] They are required to assess any systemic risks that arise from their operation within the Union, and to take measures for the mitigation of those risks. This exercise shall be supervised by an independent audit on a yearly basis. Further transparency measures, crisis protocols, standards and codes of conducts are prescribed for VLOPs in Section 4 and 5. Finally, Chapter IV is about the system of enforcement, discussing competent authorities, new responsibilities in the form of the Digital Services Coordinators, penalties and sanctions.

In sum, the DSA aims to replace the ECD, which, however, may partially remain in force regarding some of its provisions that are not included in the DSA. A major difference to ECD is the choice of instrument: as a regulation, the DSA will be directly effective in the Member States, and apply also to market actors which are settled outside the EU. It does not leave room for member state legislation in the realm of compulsory legal rules.[652]

6.3 Regulating illegal content: transparency and fair procedure. A detailed scrutiny.

6.3.1 Liability and due diligence

The ECD rules governing intermediary liability clearly needed an update. However, its liability framework only addresses the regulation of illegal

651 Article 33 (1–2) DSA. With the changing of the Union's population, by at least 5 % to the baseline in 2020, the number needs to be adjusted, to represent 10 % of the Union's population.

652 This is confirmed by Recital 9 which sets out that the DSA fully harmonises the rules for intermediary services, and that Member States should not adopt or maintain additional national requirements in the field.

content, whereas the challenges posed by online platforms were more complex than that. Online platforms' typical added value in comparison to a hosting provider is not merely to "disseminate" the message to the public, but "tagging, indexing, providing search functionalities"[653] as well as recommending, up- or downranking and moderating. They govern which peace of content receives what level of publicity, and which peace of information reaches which specific user. As said above, their definition does not include this significant activity, which, however, is still in the central focus of DSA. The absence of this element in the definition makes it clear that content governance is optional, meaning that platforms do it voluntarily to enhance their revenues. Whether or how they govern the content that they disseminate, does not influence their immunity for third party content. In fact, content *moderation* (i.e. the removal of illegal content) and content *governance* (i.e. optimising) should be viewed as two separate issues, with two parallel regimes attached: the liability framework and the due diligence framework. However, due diligence includes the notice-and-action regime and its safeguards, and there are overlaps also at other points. For instance, the first example of a systemic risk is the dissemination of illegal content, which is supposed to be dealt with under the notice-and-action regime.[654] Further, a violation of fundamental rights would also be illegal per se (see more on this below).[655]

Regulating content moderation was easy: the liability framework of the ECD was retained with a few, albeit meaningful, differences. The same cannot be said for content governance. The legislator had several reasons to approach this complex issue cautiously, rather than proposing binding hard rules. First, the activity of content governance has been opaque and constantly changing, rendering it difficult to even define the subject of regulation. Often social media platform representatives were not fully aware what exactly their algorithms have been doing. Algorithmic content governance has been all about experimenting, trial and error, with rapid immediate feedback loops.[656] Of course, no matter how thrilling this experimentation may have been for those in charge of its design, it has been

653 Buiten, "The Digital Services," 371.
654 Article 34 (1)a DSA.
655 Barata, "The Digital Services," 18.
656 Bibal et al., "Legal requirements on explainability in machine learning," *Artificial Intelligence and Law* 29, (2021): 149–169, available at: https://doi.org/10.1007/s10506-020-09270-4.

grossly unethical, lacking moral and legal considerations about implications for individuals, minorities, societies, and political processes.[657] Second, besides the obvious interference by platforms, the effects would not have been achieved without the contribution of users, through their liking, sharing, posting and commenting actions. Third, imposing hard regulations on dealing with otherwise lawful content would constitute a constraint on freedom of expression. At the same time, if platforms tamper excessively with their optimisation, that may also infringe upon the rights of users, among others, their right to free expression.

6.3.2 Immunity, as a constraint on liberty

Intermediaries enjoy a conditional exemption from liability for third party content. This is less than the unconditional American rule of CDA §230 which holds that "*No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider*." As an exception to the exception, DSA allows that providers may be liable for illegal information, if they played an active role in such a manner as to possess knowledge of, or control over, that information.[658] The precise content of this "active role" is not settled. Apparently, a mere indexing, maintaining of a search function and recommending information on the basis of the profiles or preferences of users is not a sufficient ground for considering that provider having 'specific' knowledge of illegal activities carried out on that platform or of illegal content stored on it.[659] There are merely a few cases when exemption from liability is excluded: where the recipient of the service operates under the authority or control of the hosting service provider, or when an online platform presents transaction-related information in such a way that misleads consumers to believe that the information was provided by that platform, or by the traders under their authority or control.[660]

657 Zuboff, *The age of surveillance capitalism*. See also: Philip N. Howard, Samuel Woolley and Ryan Calo, "Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration," *Journal of information technology & politics* 15, no. 2 (2018): 81–93.

658 Recital 18, DSA.

659 Recital 22, DSA.

660 Recital 23, DSA.

Besides, the Commission also made it clear that the liability framework should not be regarded as the main tool to tackle the complex problems with harmful online content, content governance and manipulation.[661] Therefore, liability for content is not applied as a sanction or as a threat on platforms to incentivise their due diligence. Instead, the DSA moves the questions of responsibility[662] away from the area of liability for content, into the realm of administrative regulation. While losing liability for hosting third party content if they fail their due diligence obligations may seem as a simple, and therefore, tempting regulatory solution,[663] it would put intermediaries into the role of speakers. This would bestow on them more than just obligations: it comes with rights, especially the right of free speech. As speakers, they would have the right to express their opinion, to discriminate, to represent their own agenda, like newspapers or audiovisual channels do. This would be a step back from a plural public discourse in more than one way. First, it would make platforms the most powerful and influential content providers ever. Second, their empowerment would take place to the detriment of the actual content providers, whose content would be appropriated by platforms, their right to freedom of expression would thereby be curbed.[664]

In sum, responsibilities and liberties go together, and the reduction of intermediaries' responsibilities also keep them confined to their activity of neutral and technical transmission. To avoid liability, they are bound to transmit third party content authentically. This contributes to keeping their activities within boundaries and prevent that they completely outgrow the traditional media system.[665]

Obviously, even if platforms do not alter third party content, and do not assume responsibility for them, their role of organising content still renders them unparallel impact on the public discourse. It is doubtful whether the due diligence/risk management approach will suffice to tackle this, as discussed later.

---

661 Recital 27, DSA.

662 For a distinction between liability, responsibility and accountability, see Hartmann, S. *Perspectives of Platform Regulation* (Baden-Baden: Nomos, 2021).

663 Buiten, "The Digital Services".

664 Kate Klonick, "The new governors: The people, rules, and processes governing online speech," *Harvard Law Review* 131 (2017): 1598.; Jack M. Balkin, "Free speech is a triangle," *Columbia Law Review* 118 (2018): 2011.

665 Judit Bayer, "Between Anarchy and Censorship. Public discourse and the duties of social media. CEPS Papers in Liberty and Security in Europe," (2019).

This regime raises some concerns in at least two aspects. First, whether courts will be able to give a consistent interpretation of the liability rules. These rules are not clearer than those under the ECD, because the filtering, sorting, optimising, up-and downranking, recommending and moderating activity is still, somewhat hypocritically, presented as if they were included into a passive and neutral, technical "dissemination".[666] Second, whether the due diligence obligations, which are meant to tame the listed activities, can be adequately enforced in absence of clear and hard obligations. As discussed below, DSA is completed with a set of co-regulatory codices, which are supposed to fill the legal principles with more concrete content. In any case, interpretation and enforcement of this soft legal package will need strong enforcement authorities. The similar "Duty of Care" regime in the UK will be enforced by Ofcom, which has centralised and established regulatory power in contrast to the scattered network of envisaged Digital Services Coordinators across the EU.[667] This is partially balanced by the Commission's powers in enforcing the Act, and supervising VLOPs and VLOSEs.

6.3.3 The liability framework

Following the red thread throughout the digital legislative package of the EU, the legislator made efforts to balance the power asymmetry between service providers and users.[668] As a condition for exemption from liability for third party content, DSA sets out the requirement of notice and

666 Recital 18. "The exemptions from liability established in this Regulation should not apply where, instead of confining itself to providing the services neutrally by a merely technical and automatic processing of the information provided by the recipient of the service, the provider of intermediary services plays an active role of such a kind as to give it knowledge of, or control over, that information."

667 Lorna Woods, "Introducing the Systems Approach and the Statutory Duty of Care," in: *Perspectives on Platform Regulation*, ed. Bayer et al. (Baden-Baden: Nomos, 2021).

668 Benjamin Wagner et al., "Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act." *Barcelona: ACM Conference on Fairness, Accountability, and Transparency* (2020) See also: Amélie Heldt, „Reading between the lines and the numbers: an analysis of the first NetzDG reports." *Internet Policy Review* 8, no. 2 (2019) https://policyreview.info/articles/analysis/reading-between-lines-and-numbers-analysis-first-netzdg-reports.

action,[669] but with a significant set of procedural rules that are meant to protect fundamental rights, primarily freedom of expression and procedural rights.[670] Hosting service providers should put in place easily accessible and user-friendly mechanisms to allow any individual or entity to notify them, by electronic means, of an information that they consider to be illegal content. Providers should also provide a template to ensure that substantial elements are included in the notice, so that it can establish actual knowledge beyond doubt. Obviously, providers will still not be in the position to pass a well-based judgement in several instances of notified content, as content that is deemed illegal can be very diverse. Some could be declared as manifestly illegal doubtlessly in the first moment of encountering them (e.g. child pornography), others may need careful balancing by a court (e.g. defamation) or even courts.[671] Private content moderation is bound to stay below the constitutional requirements of freedom of expression restrictions. The system of due diligence with its safeguards around the notice and action process intends to soothe the negative effects of what is often called "outsourced censorship" while still reap the advantages of platforms' first-hand intervention against illegal content.

Clarifying the previously ambiguous situation, the DSA explicitly empowers providers to also remove content that is not illegal but contrary to their terms of services (TOS).[672] The TOS is treated as a contract between the user and the platform, although this is never spelled out. Requirements towards the TOS provide transparency and protection to the users, whereas they restrict the contractual freedom of service providers. The set of requirements apply not only to VLOPs, but to all platforms, which may put small enterprises to a disadvantageous situation.[673] DSA provides that the TOS must respect fundamental rights, as described in the Charter. While the Charter is not directly applicable to private parties, a European Regulation is competent to impose this requirement and refer to the Charter as

669 Article 16 DSA.

670 Judit Bayer, "Procedural rights as safeguard for human rights in platform regulation," *Policy&Internet* 1–17. Online first, 25 May 2022. https://doi.org/10.1002/poi3.298.

671 Barata, "The Digital Services," 16.

672 Article 15 (1)b-c-d, 17 (3) e, 20 (1) DSA.

673 Alexander Peukert, (2022). Zu Risiken und Nebenwirkungen des Gesetzes über digitale Dienste (Digital Services Act). *KritV Kritische Vierteljahresschrift für Gesetzgebung und Rechtswissenschaft* 105, no. 1 (2022): 57–82., 7–13.

a point of reference.[674] Through incorporating the obligations to respect fundamental rights into DSA, the legislator has transformed horizontal human rights into a new, vertical protection formed by directly applicable law.

All the safeguards that apply to the action related to illegal content, also apply to actions applied on content not in harmony with the TOS. Through this, the legislator has acknowledged that the voluntary content-moderation of platforms may also have negative consequences to freedom of expression.[675] These guarantees will also close the loophole that TOS would offer to evade the procedural safeguards, as it happened with the original version of NetzDG.[676]

The same principle is followed in the provisions which pay respect to the rights of both sides: the providers of the incriminate content and the notifiers, whose rights may have been infringed by the content. Both must be informed about the action that follows the notice, and both can start an internal complaint procedure, or an out-of-court dispute resolution.

To serve justice to hosting providers, the so-called "good Samaritan" rule is incorporated, which exempts service providers from liability if they voluntarily carry out investigations or apply other measures aimed at detecting, identifying, removing or disabling access to illegal content.[677] In absence of this exemption, service providers were in fact more interested in remaining passive, to avoid that they are made liable for illegal content, when they offer content moderation but oversee some illegal content, like it happened in the Delfi case. It should be noted that this liability exemption system departs from the system set out and applied by the Copyright DSM Directive, under which platforms are responsible for copyright-violating content even if they were uploaded by third parties.[678]

During the legislative procedure, several amendments were suggested by the European Parliament (IMCO), many of them were finally not included in the text. It may be interesting to see what were those ambitions that finally were not realised. First, it was suggested that providers should ensure

674 Article 14 (4) DSA.

675 Kalbhenn, J., & Hemmert-Halswick, M. "EU-weite Vorgaben für die Content-Moderation in sozialen Netzwerken," *ZUM* 65, no. 3 (2021): 184. at p. 189.

676 Judit Bayer, "Procedural rights as safeguard for human rights in platform regulation," *Policy&Internet* 1–17. Online first, 25 May 2022. https://doi.org/10.1002/poi3.298.

677 Article 7, DSA.

678 Directive 2019/790 on Copyright and Related Rights in the Digital Single Market.

that such voluntary investigations and measures against illegal and TOS-conflicting content have adequate safeguards, such as human oversight, documentation or other, to ensure and demonstrate that they are accurate, non-discriminatory, proportionate, transparent and do not lead to over-removal of content. Where algorithms are used for such voluntary policing, providers should have made best effort to limit false positives. Second, the exemption retained from the ECD that providers are not obliged to monitor illegal content and illegal *activity,*[679] was suggested to be further specified so that to embrace both *de iure* and *de facto*, as well as by both automated or non-automated means, and extending on the behaviour of natural persons. This amendment, again, was not passed in the final round. It might have collided with the intellectual property regulation[680] which prescribes compulsory pre-screening of content for intellectual property rights violations.[681] Furthermore, an amendment would have prescribed the absence of obligation to use automated tools for content moderation and their right to use end-to-end encryption techniques.[682] It also would have prohibited Member States that they oblige providers to limit anonymous use, and to retain personal data indiscriminately. The absence of this rule may have significance in future, because without it, it remains technically possible for Member States to prescribe providers that they require identification of their users, or to retain personal data.

6.3.4 Due diligence

As said, the due diligence obligations include obligations that are not in direct correlation with the liability. If these due diligence obligations are violated, it would not result in the provider becoming liable for third party content. However, it can still result in receiving a draconian fine from the

679 Article 7–8 DSA.

680 Article 17 of Copyright Directive, https://eur-lex.europa.eu/eli/dir/2019/790/oj.

681 Quintais, João and Schwemer, Sebastian Felix, "The Interplay between the Digital Services Act and Sector Regulation: How Special is Copyright?" (January 28, 2022). forthcoming in European Journal of Risk Regulation 2022, Available at SSRN: https://ssrn.com/abstract=3841606 or http://dx.doi.org/10.2139/ssrn.3841606. See also: EDRI (2018) Commission claims that general monitoring is not general monitoring. https://edri.org/our-work/commission-claims-that-general-monitoring-is-not-general-monitoring/.

682 Article 7 (1a-1b) DSA.

Commission,[683] fine or periodic penalty payment from the Digital Services Coordinator which may also issue orders and interim measures.[684]

Due diligence obligations extend on all kinds of intermediary service providers including platforms, in a pyramid-like progressive manner (Figure 1.) The first section of Chapter III sets out due diligence obligations for all providers of intermediary services in order to ensure smooth communication with authorities, and with users. Section 2 narrows the scope of provisions down on hosting services including online platforms, section 3 to only online platforms, section 4 to online trading platforms, and section 5 to VLOPs.

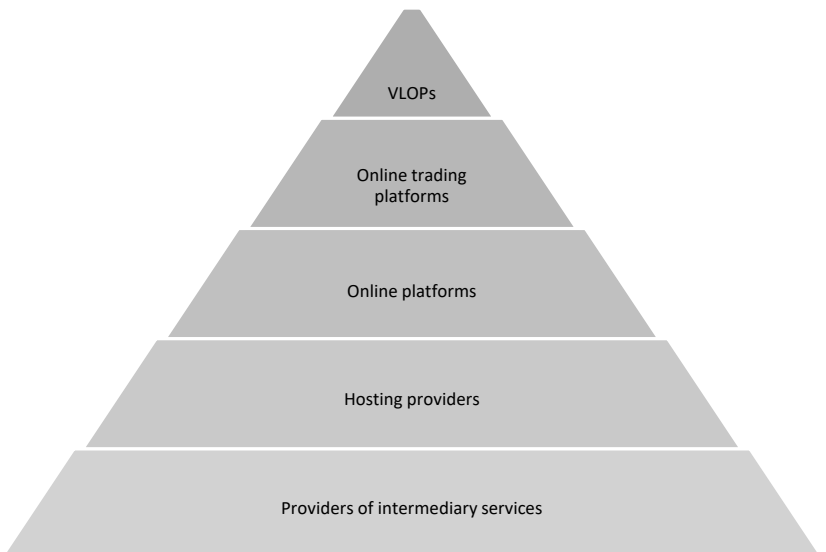

Figure 1. The structure of legal subjects of DSA from the general to the more specific.

The obligations under the due diligence category are diverse. The most basic set of obligations is the requirement for transparency, which is also the most typical regulatory tool applied throughout the DSA. The extended transparency rules are meant to address the notorious opacity of content governance and the diffuse harms they are suspected to cause. The required

683 Article 74 DSA.
684 Article 51 DSA.

reports are expected to deliver sufficient data that has been so far missing, empowering not only the regulatory, but also the researcher community and civil society, which may contribute with its analysis and reflections. Transparency towards the users reflects the hope of developing a more literate and conscientious user base. Below, these transparency obligations are discussed very briefly.

In addition to the general opacity problem, some online platforms were found unapproachable, non-responding to authority's requests.[685] This should be different if all providers are obliged to designate one single point of contact who can be contacted directly by electronic means, and to specify the selected language(s) (which must be one of the official languages of the EU), in addition to "a language broadly understood by the largest possible number of Union citizens" which is likely to be English.[686]

Intermediaries that are not established within the Union, but which offer services there, must designate a legal representative in one of the Member States, who should be possibly held liable for non-compliance with the DSA, besides the liability carried by the providers themselves. This person's name, postal address, email address and telephone number shall be publicly and easily accessible, and be notified to the DSC.[687] One legal representative may represent more than one provider, which makes compliance easier for smaller companies. For users, the means of communication shall include other manners than solely automated tools. This would provide a great relief for users in comparison to previous access possibilities restricted to a chatbot. However, the measure will need an extra investment into human resources from online providers, which might be a burden on smaller enterprises – as this section, too, applies to all intermediaries and not only to very large ones.[688]

685 Tony Zamparutti et al., "Developing a handbook on good practice in countering disinformation at local and regional level," European Committee of the Regions, Commission for Citizenship, Governance, Institutional and External Affairs. CIVEX 2022. doi: 10.2863/066582.
686 Article 11 DSA.
687 Article 13 DSA.
688 Article 12 DSA.

6.3.4.1 Platforms' terms of services

Terms and conditions have become more formalised, and intermediaries are liable on these, similarly as on the terms of a contract, as held by German courts. Failure to act in harmony with their terms and conditions would establish their civil liability. One of the most basic requirements towards all intermediaries is to make their terms and conditions transparent to all users. This should include all information relating to content moderation and their internal complaint handling system, in an accessible and user-friendly manner, and VLOPs should, in addition, provide a brief and machine-readable summary. It is allowed to unilaterally change the terms, as long as the users are informed.[689]

German courts have repeatedly judged that the TOS must respect freedom of expression,[690] and other human right principles,[691] and treat users as equal parties.[692] The German Civil Code – like that of several other nations' – includes clear limitations on the content of General Terms and Conditions,[693] including that unilateral amendment of the terms is invalid.[694] Twitter's TOS was found unlawfully saying that they are entitled to revise their TOS from time to time, as a vague condition empowering the platform with a blank slate.[695] As a European Regulation, DSA's rule that enables the unilateral changing of the TOS prevails over the national rule in regard of the specific personal scope of online intermediaries. However, the DSA also incorporated several of these principles: the terms need to pay due regard to the rights and legitimate interests of all parties involved, including fundamental rights, particularly freedom of expression, freedom and pluralism of the media and other fundamental rights and freedoms

689 Article 14 DSA.

690 LG Frankfurt am Main, 14.05.2018 – 2–03 O 182/18, MMR 2018, 545.

691 OLG Karlsruhe, 25.06.2018 – 15 W 86/18, NJW 2018, 3110; LG Heidelberg, 28.8.2018 – 1 O 71/18, MMR 2018, 773.

692 BGB [German Civil Code] (87th edition, 2021), § 241 para. 2. The Court held that the TOS violated the principle of good faith when it stated that the platform may remove any content. Additionally, the fact that Facebook alone decided whether a post violated its guidelines was contrary to the Civil Code, which provided for equal rights of the contracting parties.

693 BGB, §305–310.

694 BGB, §308, no. 4–5.

695 LG Dresden, 12. 11. 2019 – 1a O 1056/19, MMR 2020, 247; OLG Dresden, 07.04.2020 – 4 U 2805/19, MMR 2020, 626.

in the Charter.[696] The same rights are also subject to the risk assessment exercise to be applied by VLOPs (see below in more detail).[697] Authorities are required to achieve a fair balance between these rights concerned, when exercising the powers set out in this Regulation.[698] What exactly this obligation should comprise, is not elaborated, but the interpretations open a wide horizon. The German media regulation requires that social media platforms refrain from systemic discrimination of media content carried through their services.[699] This is one of the first mentioning of the obligation to respect 'freedom and pluralism of the media' as a fundamental right in a legislative document.

6.3.4.2 Transparency reporting obligations

Similar to the previous requirement for transparency of TOS, the level of transparency in reporting obligations has also risen considerably during the legislative process, in comparison to the initial draft. The last amendments by the IMCO were largely accepted in this regard and added to enhance the transparency of services.

All intermediary service providers are obliged to yearly publish their easily accessible, comprehensible and machine-readable reports on their content moderation activities. Micro or small enterprises are excepted, unless they are also VLOPs – while this may sound surprising, as the two categories have different logic for their definition, it is not impossible.[700] An enterprise is considered as micro or small, if it employs fewer than 10 or 50 persons (respectively), and its annual turnover or its balance sheet does not exceed EUR 2 or 10 million. Whereas the status of being "very large" is defined per number of users.

The reports need to include a wide range of information, different for each subcategory of intermediaries. First of all, all providers of intermediary services must report on the number of orders that they received from authorities to remove illegal content or to provide information on users, categorised according to the type of illegal content, the issuing Member State, and the median time needed to get the order done. The same scope

696 Article 14 (4) DSA.
697 Article 34 (1)b DSA.
698 Recital 153 DSA.
699 § 94 MStV.
700 Article 19 DSA.

of actors should also report on the number of complaints that they received through their internal complaint-handling systems. Providers of hosting services are required to include the number of notices to remove content, distinguishing those that were submitted by trusted flaggers, and categorised by the type of illegal content. The report should include information on any action taken upon the notices and whether those were taken on the basis of the law or the terms and conditions, further, the number of those notices processed using automated means and the median time needed for taking action. Providers of online platforms should, in addition, include the basis for such complaints, the decisions taken, and the median time for taking action, as well as the number of instances when these decisions were reversed.[701]

Content moderation initiatives, including those with automated tools, must be reported by all providers of intermediary services. This shall also include the number and type of measures that affected the availability, visibility and accessibility of user content – in other words: up- and downranking, or deprioritising of third party content. Up- and downranking is the least visible method of governing speech, because rather than removing the information, it merely pushes the information further down on the "long tail",[702] to decrease the chances that users would encounter them. This is exclusively done through algorithms, the transparency of which faces several technical difficulties (discussed below in more detail).[703] Reports on content moderation should include the measures taken to provide training and assistance to the persons in charge of content moderation.[704] This responds to the obligation of employing moderators who are aware of the local or national circumstances.[705] Further, if automated means were used for the purpose of content moderation, their qualitative description should be included with several details, among others a specification of the precise purposes, indicators of the accuracy and the error rate.[706]

701 Article 15 (1)(a, b, d) DSA.
702 Chris Anderson, *The Long Tail. Why the Future of Business Is Selling Less of More* (New York, NY: Hachette Books, 2008).
703 Joan Donovan, Why social media can't keep moderating content in the shadows," *MIT Technology Review* November 6, 2020. https://www.technologyreview.com/2020/11/06/1011769/social-media-moderation-transparency-censorship/.
704 Article 15 (1)(c) DSA.
705 Recital 87 DSA.
706 Article 15 (1)(e) DSA.

The transparency report is regularly published by the Commission, and available for the public, with special regard to researchers and NGOs. This transparency is expected to enable scrutiny over the content moderation practices of online platforms, and provide information about the spread of illegal content online, even if most of those are removed.[707] It remains to be seen, how informative in reality these formalised quantitative reports will be.

6.3.4.3 Transparency reporting for online platforms

Following the pyramid-like structure, online platforms have further reporting burderns, such as the number of disputes submitted to the out-of court dispute settlement bodies, and their outcomes, with the median time needed for completing the disputes, as well as the ratio of those where the provider implemented the decision. Also the number of suspensions of user accounts, and the complaint processing of notorious complainants should be included. Platforms and search engines both should publish information on their monthly average active users, and submit the information to the responsible DSC and the Commission, in order to calculate whether they count as "very large".[708]

Transparency of recommender systems will be ensured by the terms and conditions which must set out the main parameters of those, and any options that are available for users to modify or influence those main parameters. These main parameters shall explain why certain information is recommended, but at least the most important criteria, and the relative importance of those.[709] As these algorithms are of crucial importance for the governance of the public discourse, they will be discussed later in more detail.

6.3.4.4 Further obligations for very large online platforms

VLOPs have further specific transparency obligations regarding advertisements and content moderation. If they present ads, they shall publish an ad repository on their online interface (which can be a website, a mobile

707 Article 40, Recital 66 DSA.
708 Article 24 (1–4) DSA.
709 Article 27 DSA.

app or whatever the future brings). This repository needs to be searchable through multicriteria queries, be reliable, and include not only the information that had previously been required for political ads, but also the name of the product, service or brand and the subject matter of the advertisement. The "classical" information includes the person on whose behalf the ad is presented (usually called advertiser), the person who paid for the ad, the period during which the ad was presented, the main parameters of targeting if it was targeted, including parameters used to exclude one or more groups of the population, user-generated ads, the total number of users reached, and numbers for the group(s) of users that were targeted, broken down by Member State.[710]

Even with the archive openly accessible, the average user is unlikely to browse the archive to learn about the background of ads. Researchers and advocacy organisations are expected to carry out this task, in order to monitor for discrimination or deception.

Reporting on content moderation shall be done with double frequency for VLOPS, i.e. two months after notification about their status and every six months thereafter, at least. Besides the statistical information that is required generally from all intermediaries (Article 15, see above), VLOPs also need to add the human resources that they dedicated to content moderation, broken down by each applicable official language of the Member States. The report should also contain the qualifications, the training and the support given to the content moderators, as well as their linguistic expertise, in a way that enables to control compliance that they are able to adequately respond to notices, and to handle internal complaints.[711] Automated means used for content moderation have to be transparent on indicators of accuracy (and possible error rate) and related information, broken down by each official language of the Member States.[712] This information may point out weak points in the content moderation algorithms and AI applications, as well as eventual understaffing.

The transparency obligations encompass the sharing of the audit report, along with the audit implementation report, a report on the results of the risk assessment, on the mitigation measures, and information about consultations conducted, if any. After three months of sending these to

710 Article 39 DSA.
711 Article 42 (2)a read together with Articles 16, 20 and 22 DSA.
712 Article 42 (2)c read together with Article 15 (1)e, DSA.

the Digital Services Coordinator and the Commission, they also must be published (with confidential information removed).[713]

6.3.4.5 Scrutiny of the shared data

These extensive transparency measures will certainly deliver ample information for research, improvement of services, and improvement of policies. However, transparency can lead to a meaningful change only if the delivered information is processed, analysed and acted upon. It cannot be expected from transparency alone to raise user awareness or induce users to make more conscious decisions during their use of the online platform or search services.

Digital Services Coordinators will function as special hubs of this information. Beyond accessing data by VLOPs for the purpose of monitoring and compliance, they can request access on behalf of "vetted researchers" for the purpose of conducting research that contributes to the detection, identification and understanding of systemic risks in the Union. The "vetted researcher" status is granted to researchers who fulfil the conditions by the Digital Services Coordinators, however, only for specific projects. Platforms must give vetted researchers access to their data based on DSA, the Code of Conduct against Disinformation and a Code of Conduct developed by EDMO – this specifies platform-to-researcher data access in compliance with GDPR.[714] The latter also designed an independent intermediary body that could vet researchers and research proposals, and evaluate the codebooks and the datasets that platforms make available.[715]

In this chapter, only the most relevant transparency requirements were introduced, as briefly as possible. Transparency requirements bind all types of providers, as these are the foundation level of the pyramid of obligations. The most prominent and recurrent systemic risks, as well as best practices on how to react on them will be identified and collected by the Board on a yearly basis, broken down by Member States, based on the reports by VLOPS and on other sources.[716] The sheer amount of the delivered data will generate considerable analytical tasks requiring both human and

713 Article 42 (5) DSA.
714 Article 40 DSA.
715 Mathias Vermeulen, "Researcher Access to Platform Data: European Developments," *Journal of Online Trust and Safety* 1, no. 4 (2022) https://doi.org/10.54501/jots.v1i4.84.
716 Article 35 (2) DSA.

computational resources. Drawing conclusions on the basis of the data will be a further task after several years' trends are visible. Ensuring the objectivity of the data analysis is likely to present further challenges, due to a limited pool of independent experts, as large parts of the pertinent intellectual capital are engaged with either platforms, policy, or advocacy organisations.

6.3.5 Rights of the users in the notice-and-takedown procedure

The notice-and-takedown procedure is nothing new, but furnishing it with procedural guarantees is a novelty. Procedural fairness is a human right in itself, and also a tool to protect other human rights that are the subject matter of the procedure at hand, in this case, freedom of expression and its conflicting rights. This package of provisions is another example how the theoretically horizontal relationship between platforms and users, all private actors, is acknowledged as a de facto vertical relationship where users' rights are protected by law as if they would be in relation to a public authority.[717]

The providers' exemption from liability for content is not affected by these rules, as everything beyond the actual removing or disabling of an illegal material belongs into the realm of due diligence obligations. Should they omit one or more of these requirements, they can be sanctioned with the available administrative tools, but not become responsible for the content itself.

Content moderation has a broader definition than just removal, and thus spills over to the field of content governance. It includes measures taken that affect the availability, visibility, and accessibility of that illegal content or that information, such as demotion (downranking), demonetisation,[718] disabling of access, as well as the termination or suspension of a recipient's account.[719] Some procedural rules also apply when the content moderation

717 Judit Bayer, "Procedural rights as safeguard for human rights in platform regulation," *Policy&Internet* 1–17. Online first, 25 May 2022 https://doi.org/10.1002/poi3.298.

718 Demonetising, i.e. depriving the content of the possibility to feature advertisements, or simply terminating or suspending the respective payment has been applied as a method to disincentivising the publishing of harmful material, whereas preserving the fundamental rights to publish them.

719 Article 3 (t) DSA.

takes place on the basis of the platform TOS, or even "regardless of why or how it was imposed", rather than only in case of (assumed) illegality.[720] This provides additional safeguards and transparency for content moderation, and in some cases, to content governance, protecting the fundamental informational rights of users. Moreover, notifications are required to reach a certain level of preciseness and substantiation as well as to protect the free expression rights of the content provider to minimise unjustified content removal.[721]

As an exception, no reasoning must be given in case of intentional manipulation of the service or "deceptive high-volume commercial content", such as inauthentic use by bots, fake accounts, or large disinformation schemes. Even though the reasoning is not required in these cases, the users in question would still retain the right to effective remedy before national court.[722] It should be noted that the obligations to provide reasoning apply only if the relevant electronic contact details are known to the provider,[723] that is, only for registered users.

The obligations apply to providers of hosting services, which includes platform providers. They must put in place easily accessible and user-friendly mechanisms to allow notification of illegal content exclusively by electronic means. This suggests an online template which ensures that the notice includes sufficiently precise and complete information.[724] The reasoning should identify the nature of the restriction and its territorial scope and duration. The duration had been a crucial point in the Facebook Oversight Board's landmark decision on the suspension of Donald Trump: Facebook imposed an indefinite suspension on the user (then incumbent President of the United States), which was not among the possible measures. A suspension had to be either terminal (then called 'termination') or temporal for a fixed time period, therefore the Oversight Board found the suspension violating Facebook's own terms of services (the suspension was approved on the other accounts).[725] In response to the decision, Facebook

720 Article 17 (2) DSA.
721 Article 16 DSA. On unjustified content removal, see: Bayer, VUW, 2007.
722 Recital 55. DSA.
723 Article 17 (2) DSA.
724 Article 16 DSA.
725 FOB: Oversight Board Upholds Former President Trump's Suspension Finds Facebook Failed To Impose Proper Penalty. https://www.oversightboard.com/news/226612455899839-oversight-board-upholds-former-president-trump-s-suspension-finds-facebook-failed-to-impose-proper-penalty/.

suspended Trump's account for two years, with a reinstatement depending on the fulfilment of conditions.[726]

Besides, the statement of reasons must contain further details on whether the decision was taken on the basis of a notice or on voluntary investigations, whether automated means (e.g. algorithms) were used taking the decision or its detection or identification; the legal or contractual ground of the restriction; and the possibilities for redress, in a clear and user-friendly form, in particular with reference to the internal complaint-handling mechanisms, the out-of-court dispute settlement and also to judicial remedy. Whenever the restriction is required by an authority, such information does not need to be given to the user.[727]

This was the last meaningful protective measure among those that apply to all hosting providers as the further measures are specific to platforms.

6.4 Specific obligations for online platforms

Liability and due diligence rules discussed so far apply to all providers of hosting services. However, the main goal of DSA was to regulate online platforms, which have grown to dominate online commerce and communication, as key actors of aggregation and distribution, and those which facilitate the interaction of other actors. In the model of Luhmann's system theory, platforms can be viewed as key nodes within the system networks that are capable of defining relationship of several other actors.[728]

All communication platforms have a strong potential in vitalising communication. They enable a fulfilment of the human potential at all levels, by establishing a myriad version of connections between people, groups and associations, whether for private, political, social or commercial purposes. On the downside, human behaviour has its negative features, and the networked environment provides increased opportunities for malicious activities, too. Those, who take the effort and investment to leverage the full potential of platform communication, can be more successful in trans-

726 Nick Clegg, "In Response to Oversight Board, Trump Suspended for Two Years; Will Only Be Reinstated if Conditions Permit," Meta June 4, 2021. https://about.fb.com/news/2021/06/facebook-response-to-oversight-board-recommendations-trump/.

727 Article 17 DSA.

728 Niklas Luhmann, *Social systems* (Stanford, CA: Stanford University Press, 1995).

mitting their message than their competitors whether pursuing commercial or political goals. Social media platforms have been viewed as having a major influence on democratic processes.[729] Search engines are less in the spotlight: however, they are not less influential in forming the public information landscape. The conflict over the compatibility of their public tasks and private goals presents a similarly fiendish media policy conflict.[730]

The obligations for platforms specifically are diverse: part of them is related to their content moderation activity (internal complaint-handling system, out-of-court dispute settlement, trusted flaggers), others to their content governance (measures and protections against misuse, online interface design and organisation, advertising, recommender system transparency, protection of minors).[731]

6.4.1 Alternative dispute resolution

DSA sets out two procedures to discuss controversies around content moderation: the internal complaint-handling system and the out-of-court dispute settlement. These are designed to ensure users' right to remedy, as it is included in the international covenants[732] and Article 47 of the EU Charter of Fundamental Rights. Interestingly, this right is traditionally ensured only against authorities, primarily in a criminal procedure.[733] Contractual relations do not typically require such an exceptional protection, as they are normally between parties of equal standing. However, there are some areas where despite the contractual nature, the parties are not equal in their positions and law has intervened to protect the weaker parties.

729 Pablo Barberá, "Social media, echo chambers, and political polarization," in *Social media and democracy: The state of the field, prospects for reform*, ed. Nathaniel Persily and Joshua A. Tucker (Cambridge: Cambridge University Press, 2020).

730 Boris P. Paal, "Vielfaltsicherung im Suchmaschinensekto,"*Zeitschrift für Rechtspolitik*, (2015): 34–38. at p. 35.

731 Transparency reporting obligations specific to online platforms have been discussed above under Chapter 6.3, therefore they are not discussed here again.

732 Article 10 of the Universal Declaration on Human Rights (UDHR), Article 6 of the European Convention on Human Rights (ECHR).

733 Article 6 ECHR extends this right explicitly to both criminal and civil procedures, and ECHR practice has extended it also to commercial and administrative law, and beyond courts also to administrative authorities (Georgiadis v. Greece, § 34; Bochan v. Ukraine (no. 2) [GC], § 43; Naït-Liman v. Switzerland [GC], § 106, see also ECHR Guide, 2020).

In consumer protection, labour law, and other sectoral regulations such as banking or telecommunication.[734] These classical fields also represent situations where the theoretically horizontal relationship between the private parties becomes de facto vertical, and regulation interferes to correct the power asymmetry. This is even more so with platforms, because with the *mandated* removal of illegal content, platforms perform a function that has originally been reserved for authorities. They are practically the lengthened hand of the state administration.[735] Therefore, all safeguards for the restriction of users' fundamental rights should be applicable, as if platforms were an authority,[736] including the possibility of complaint and judicial review.[737] However, the sheer volume of content and the complaints that their moderation entails is without precedent in human legal history. Courts would be severely overloaded if they had to discuss all user complaints.[738] Moreover, a large proportion of the content in question is restricted not on the grounds of illegality, but of conflicting with the platform terms and conditions.[739] Legal theory has still not satisfactorily clarified the extent of platforms' freedom in defining their terms and conditions,[740] or whether users have the right to publish their perfectly lawful and unharmful content through online intermediaries.[741] Whereas, in the US, First Amendment has

734 Michael J. Trebilcock, *The limits of freedom of contract* (Harvard University Press, 1997).

735 HRW, Human Rights Watch. 2018. "Germany: Flawed Social Media Law." HRW. February 14. https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law.

736 Rikke Frank Jørgensen and Lumi Zuleta, "Private governance of freedom of expression on social media platforms: EU content regulation through the lens of human rights standards," *Nordicom Review* 41, no. 1 (2020): 51–67, https://doi.org/10.2478/nor-2020-0003.

737 Judit Bayer, "Procedural rights as safeguard for human rights in platform regulation," *Policy&Internet* 1–17. Online first, 25 May 2022 https://doi.org/10.1002/poi3.298.

738 Anumeha Chaturvedi, "Facebook Draws Ire of its Own Oversight Board in First Transparency Report," *The Economic Times* October 21, 2021. https://economictimes.indiatimes.com/tech/technology/facebook-draws-ire-of-its-own-oversight-board-in-first-transparency-report/articleshow/87188735.cms?from=mdr.

739 Rolf Schwartmann and Robin L. Mühlenbeck, " NetzDG und das virtuelle Hausrecht sozialer Netzwerke," *ZRP*, 170. (2020).

740 LG Dresden, 12. 11. 2019 – 1a O 1056/19, MMR 2020, 247; OLG Dresden, 07.04.2020 – 4 U 2805/19, MMR 2020, 626.

741 LG Frankfurt am Main, 14.05.2018 – 2–03 O 182/18, MMR 2018, 545. See more in: Bayer, *Rights and duties*.

been understood to protect intermediaries' right to moderate content.[742] In any case, such decisions raise different questions from typical contractual questions, because conflicting rights need to be balanced. This requires an individual, case-by-case decision-making. The claims for cost-effectiveness and for a speedy decision also supported the need to work out alternative dispute-resolution solutions.

DSA created two instances to solve disputes. The first one, the internal complaint-handling system should be established by online platforms, so as to be open for complaints regarding all their content restriction actions (including downranking and demonetising), for at least six months after informing the user about the decision. Complaining should be possible electronically and free of charge, the system should be easily accessible, user-friendly and enable and facilitate the submission of precise and substantiated complaints. The decision within this system should take place under qualified human supervision and not merely through automated means.[743]

The second level is the out-of-court dispute settlement. The way it is defined in the DSA resembles a form of mediation service that does not deliver binding decisions. Therefore, the judicial route remains open to both parties. However, the idea of the dispute settlement may come from the intent to save time and costs, and to have a better PR. The Facebook Oversight Board is generally regarded as an example of such a body. Even though it has been created by Facebook (now META), after the selection of the first set of co-presidents, the body elects its own members and jurors. Its funding has been established through a Trust that is supposed to finance the Board in at least two consecutive three-year terms. The question remains, why would users be interested in initiating this procedure, when its results are not binding? Still, DSA goes to great lengths to ensure that the independence, credibility and fairness of such bodies are guaranteed through regular certification by DSCs. To acquire the certificate, the bodies need to demonstrate their independence meaning that they are completely independent from both online platforms and users, including user associations; and that their members are remunerated in a way that is unrelated to the outcome of the procedure. To prove their credibility, they need to have the necessary expertise in at least one particular area of illegal content, or

742 Barata, "The Digital Services".
743 Article 20 DSA.

apply and enforce the terms and conditions of at least one type of online platform (where "type" remains undefined, but we can assume that it refers to the function and customer basis of the platform); and that they are capable of settling disputes swiftly, cost-efficiently and effectively in at least one of the official languages of the institutions of the Union (this is different from the official languages of the EU: it merely includes English, French and German). To demonstrate fairness, they need to offer easily accessible electronic means for submission; and have clear and fair rules of procedure that comply with the applicable law, including the DSA. The procedure may be initiated by any user, and both parties must engage in good faith with the procedure, but users pay only a nominal fee, and need not reimburse the platforms' costs if they lose. However, platforms must pay the procedural costs if they lose, and reimburse the user's any reasonable expenses.[744]

The certificate is renewable after five years and the list of certified bodies is publicly available on a dedicated website by the Commission. Digital Services Coordinators have to generate a biannual report identifying problem areas, best practices and developing recommendations. A comparative evaluation of the data and the reports is likely to yield interesting insights on whether this institution contributes to a crystallisation of the new norms of the public discourse.

6.4.2 Trusted flaggers

Notifications submitted by trusted flaggers are prioritised by the platforms and are processed and decided in an accelerated procedure.[745] This bestows considerable power and responsibility on these organisations, they can also be regarded as gatekeepers, because they greatly impact what content can stay online and what will be removed. At the same time, false or inaccurate notifications are a burden to platforms, and also threaten users' freedom of expression. Abusive notification practice has also been widely documented,[746] for example, the Church of Scientology has habitually asked for

744 Article 21 (3–5) DSA.
745 Article 22 (1) DSA.
746 Stephen McLeod Blythe, "Freedom of speech and the DMCA: abuse of the notification and takedown process," *European Intellectual Property Review* 41, no. 2 (2019): 70–88.

removal of articles critical of their organisation.[747] Theoretically, the possibility offers an easy way to reduce visibility of political opponents and rival artists, while at the same time, illegal material that harms the most vulnerable social groups may remain underreported.[748] Therefore, well-functioning trusted flaggers can provide a great benefit to content moderation and the common interest. Nevertheless, their independence and expertise is of crucial importance. Trusted flaggers can now be officially recognised by the DSC if they demonstrate that they have specific competence and expertise to detect, identify and notify illegal content; that they are independent from any online platform, and that they perform their notifying activities diligently, accurately and objectively. If a trusted flagger has submitted a significant number of imprecise, inaccurate or inadequately substantiated (unfounded) notices, platforms are entitled to notify the DSC with a documentation and explanation of the statements. The DSC can then open an investigation during which the trusted flagger status is suspended. In case the entity no longer meets the necessary criteria, the status can be revoked.[749]

To tackle abusive noticing in another way, online platforms are entitled to suspend processing of notices and complaints by users who frequently submit manifestly unfounded notices or complaints.[750] Before deciding on this, the provider must assess carefully the absolute and relative number of manifestly unfounded notices, their gravity, and the intention of the user if that can be identified, and give prior warning. The suspension can last for a reasonable period of time which is not closer identified.

Trusted flaggers have been active in relation to the Code of conduct against illegal hate speech. The Code's effectiveness and success were measured by the level of removal rate as a response to the notifications.[751] However, in absence of qualitative information on the merit of decisions, the success rate of removals may merely signal that the cooperation be-

747 See more in: Judit Bayer, "Liability of Internet Service Providers for Third Party Content (2007). A comparative analysis with policy recommendations," *VUW Law Review Special Edition* Wellington, New Zealand, (2007): 1–109.

748 David Paul, "Online Abuse Not Being Reported by 1 in 4 UK Women," *Digitnews* 27 July 2022 https://www.digit.fyi/online-abuse-not-being-reported-by-1-in-4-uk-women/.

749 Article 22 (2) and (6–7) DSA.

750 Article 23 DSA.

751 Factsheet – 7th monitoring round of the Code of Conduct. https://commission.europa.eu/system/files/2022-12/Factsheet%20-%207th%20monitoring%20round%20of%20the%20Code%20of%20Conduct.pdf.

tween the trusted flaggers and the respective platform have improved. In other words, as the trusted flaggers become institutionalised (and perhaps monopolised), the notification practice may become characteristic of the entity and objectivity may suffer. A diversity of flaggers would better ensure the protection of freedom of expression and of content pluralism.

6.4.3 Measures related to content governance

Content governance is the opaque field where platforms govern with their "invisible hand".[752] Even if certain instances of interferences are known, their exact impact cannot be proven. The link between cause and effect – i.e. between the platform action and a social effect – cannot be established with certainty. For this reason, hard regulation would be both impractical and disproportionate. The obligations that fall in this field, are endeavours into unknown waters, provisions which had little or no regulatory history. They come the closest to having significance from the perspective of the rational discourse. Three such measures can be identified which apply to all online platforms, and relate to content governance: refraining from dark patterns, the separation of advertising from organic content, and transparency of recommendation algorithms.

First, what is widely called as "dark patterns", is described by DSA as practices that distort or impair users' ability to make autonomous and informed decisions. For example, by making an option less visible, more time-consuming or less easily accessible than another, also known as deception or nudging of users, via the structure, design or functionalities of an online interface.[753] It includes the prohibition of making the cancellation of a service more cumbersome than signing up to it,[754] and other techniques of forcing users gently to become or remain customers through nudging or deception, including default settings and exploitative design choices, such as giving more prominence to certain choices through visual, auditory or other components.[755] The prohibition extends to requesting the user repeat-

752 Tarleton Gillespie, "Regulation of and by Platforms," in *The Sage Handbook of Social Media,* ed. Jean Burgess, Alice Marwick, and Thomas Poell (London: Sage Publications Ltd., 2018).
753 Amendment 202 of IMCO, proposing Article 13 a, now in Article 25 and Recital 67 DSA.
754 Article 25 DSA.
755 Recital 67.

edly to make a choice, when that choice has already been made, especially by pop-ups. (A similar prohibition has been incorporated in the DMA which prohibits asking for consent repeatedly within a year if a decision had been already made.[756]) The further specifications are to be defined in Commission guidelines.[757]

The second and the third measures are yet other transparency provisions, albeit the most relevant ones for the public discourse, as they directly affect content organisation. Distinguishing advertisements from organic content has been a standard for media ethics at least in the liberal media systems.[758] Advertisements have served as the driver of platform communication and form the cornerstone of their business model. Personalised advertising has been coupled with personalised content offer, and thus interfered with the public information landscape, as discussed in the Introduction.[759] Platforms must ensure that each advertisement is identified as such, moreover, users shall be able to immediately identify without ambiguity on whose behalf the ad is presented, and who paid for the advertisement (which are not necessarily identical).[760]

Similarly, the user should be given information about how he or she was selected as an audience for the ad (what were the main parameters used to determine the recipient) and how to change those parameters.[761] This information provides immediate feedback to the user about what personal data about him or her is available for the platform, and how he or she is profiled. Special category of data should not be used at all for targeting, and minors should not be targeted either.[762] However, it is not always known whether a user is of minor age, and platforms are not required to process extra personal data in order to ascertain this.

Advertising transparency is crucial, as it indirectly affects the logic of content organisation through recommender systems. Although the cornerstone of content governance, it is addressed rather gently. Online platforms that use recommender systems shall set out clearly in their terms and

756 Article 5 (1–2) DMA.
757 Article 25 (3) DSA.
758 Daniel C. Hallin and Paolo Mancini. 2004. Comparing Media Systems : Three Models of Media and Politics. Cambridge: Cambridge University Press. Article 26 DSA.
759 Zuboff, The age of surveillance capitalism.
760 Article 26 (1) DSA.
761 Article 26 (1)d DSA.
762 Article 26 (3), 28(2) DSA.

conditions the main parameters for these. If there are any options for the recipients, then also how to modify or influence those main parameters, which should at least include the most significant criteria in determining the recommended content; and the reasons for the relative importance of those criteria. If they offer options to choose from, then they should also offer a direct and easy tool to make that choice, at the same page where the ranking takes place.[763] Only VLOPs are obliged to offer more than one option, however, the Strengthened Code of Practice on Disinformation stipulates that relevant signatories commit to offer options.[764] On transparency of the content recommending algorithms, read more in Chapter 9 on AI regulation.

6.5 Very large online platforms' due diligence obligations

The giant companies that have determined public communication in our last decade are plainly called by the DSA "very large". During the legislative process, very large online search engines have also been included along with very large online platforms, because public information, knowledge and perception is determined at least as much by online search engines such as Google. These may collect even more information about their users and possess an astonishing social power. For simplicity, in the following, both very large online platforms and very large online search engines will be understood under the term VLOPs.

Online platform providers count as "very large" if the number of their average monthly active users within the EU reaches 45 million. They needed to provide information about this number to the Commission by 17 February 2023 for the first time, and at least every six months thereafter, taking the average of the past six months. The updated list is published in the Official Journal. The Commission is entitled to adjust this number through delegated acts, in case the EU population increases or decreases by at least 5 % in relation to the latest adjustment (the baseline is 2020). In any case, the number should correspond to 10 % of the Union's population rounded up or down to millions.

Rather than liability, VLOPs hold "responsibility" in the theoretical meaning of the word: while they are not liable for the content that they car-

763 Article 27 DSA.
764 Commitment 19, Strengthened Code of Practice on Disinformation, 2022.

ry, they bear responsibility for how they govern the content and activities through their platform services.[765] While these terms have sometimes been used interchangeably,[766] there are significant differences between them: legal liability arises from a violation of law,[767] whereas, responsibility is the prior positive obligation to take the necessary measures with due diligence to prevent certain harms.[768] DSA refrains from defining those measures, and the autonomy that is thereby granted to platform providers becomes a core component of the responsibility scheme. This due diligence scheme is enhanced into a risk assessment and risk mitigation framework and culminates in co-regulation. The system is topped up with a frame of "accountability" through the auditing exercise and the potential of ensuing sanctions.[769]

6.5.1 The system of risk-management and co-regulation

Risk-regulation has already been an important part of European regulation in the past decades, especially under the aegis of the Digital Single Market Strategy,[770] and in particular in data and artificial intelligence. Alemanno called risk assessment the privileged methodological tool for regulating risk in Europe.[771] De Gregorio and Dunn call the system of DSA a hybrid system, as one halfway between the GDPR and the AI Act, where the GDPR is supposed to follow a bottom-up perspective, and the AI Act a top-bottom perspective in defining the risk categories, in the evaluation of risk, and

765 Judit Bayer et al., "The fight against disinformation, and the right to freedom of expression – an update." A study requested by the European Parliament's Committee on Civil Liberties Justice and Home Affairs. (2021).

766 EPRS, Liability of Online Platforms, (2021): 24.

767 Jaani Riordan, "A Theoretical Taxonomy of intermediary Liability," in *The Oxford Handbook of Online Intermediary Liability*,ed. Frosio (2020): 58.

768 John Naughton, "Platform Power and Responsibility in the Attention Economy," in *Digital Dominance – The Power of Google, Amazon, Facebook, and Apple*, ed. Moore and Damian Tambini (2018).

769 Article 37 'Independent audit', DSA.

770 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A Digital Single Market Strategy for Europe, COM(2015)192 final.

771 Alberto Alemanno, "Regulating the European Risk Society," in *Better Business Regulation in a Risk Society,* ed. Alberto Alemanno at al. (Springer, 2013): 53., cited by De Gregorio, and Dunn, supra note.

in the mitigation of the risk.[772] The DSA defines the categories of service providers in a top-down manner (hosting service, platform, very large platform), but the levels of risk are to be defined by the services themselves on a scale, rather than in a binary logic of prohibited and non-prohibited actions. This creates a matrix-like system where each service provider is supposed to find the individual position, depending on its size, type of services and the self-assessed risk that it carries. The subjects matters of this part in DSA are mainly related to the content-governance function, and a smaller part to the methods of content-moderation (to be adapted to deal with the systemic risks).[773] Co-regulation is incorporated as a form of risk-regulation.[774]

The idea of the autonomy-responsibility scheme, completed with co-regulation, is based on the observation that hard legal regulation faces considerable obstacles from several angles. First of all, a considerable part of the content that was perceived as causing societal problems, is not literally unlawful. These contents are protected by the right to freedom of expression and belong into the "lawful but awful" category that is often called "harmful content".[775] Even soft measures that would impose concrete obligations on providers would remain questionable, because the cause-and-effect link between action (or function) and the societal harm cannot be established with certainty. Especially not between one certain action and individual effect, although there are statistical correlations.[776] Critiques call attention to the vagueness of the systemic risks, which include all kinds of "intentional manipulations" which need neither be illegal nor violate terms and conditions.[777] It also includes the term "any negative effects", which are inevitable by-products of conflicting rights and should not be regarded as an absolute

772 De Gregorio and Dunn, "Risk-Based Approaches," 4.

773 Article 34 (2)b DSA, Article 35 (1)c DSA.

774 Giovanni De Gregorio and Pietro Dunn, "The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age," *Common Market Law Review*, 59, no. 2 (2022): 473–500.

775 Daphne Keller, (2022) "Lawful but Awful? Control over Legal Speech by Platforms, Governments, and Internet Users," *Univ. Chi. Law Rev. blog* 2022. See also: Judit Bayer et al., "The fight against disinformation and the right to freedom of expression," *Study* 05–07–2021.

776 Hannah Ruschemeier, "Kollektive Grundrechtseinwirkungen," *RW Rechtswissenschaft* 11, no. 4 (2021): 450–473.

777 Alexander Peukert, "Five Reasons to be Sceptical About the DSA" *Verfassungsblog* 31 August 2021 <https://verfassungsblog.de/power-dsa-dma-04/> accessed on 12 September 2021.

cause for limitation. For example, reporting on matters of public interest is often likely to exercise a negative effect on certain public figures.[778]

Nevertheless, it was also observed that the functioning of platforms can exercise a dramatic effect on societies even in the absence of a malicious intention by the operators of these companies. This led to the conclusion that a systemic risk arises from these functions, and therefore these intermediaries – platforms, and especially the giant ones – should be required to assess and mitigate their own risks diligently.

6.5.2 The risk assessment in DSA

VLOPs thus enjoy the autonomy to define the specific systemic risks that they take responsibility for. This, however, does not happen entirely freely: the DSA names four types of systemic risks.[779] This structure corresponds to risk-management structures in other fields, which usually offer a set of methodologies, templates and processes in order to support rational decisions on potential future threats.[780]

The first of such risks is the dissemination of illegal content itself, whereas the three others are defined through the negative effects – whether actual or foreseeable – that are due to the service or any related system thereof. The list of relevant human rights at risk have been extended during the legislative period to also include human dignity, personal data, and the freedom and pluralism of the media, besides private and family life, freedom of expression and information, non-discrimination, the rights of the child, and a high-level of consumer protection, all as enshrined in the Charter.[781] As third, actual or foreseeable negative effects on civic discourse and electoral processes, as well as public security must be considered, and

778 Alexander Peukert, „Zu Risiken und Nebenwirkungen des Gesetzes über digitale Dienste (Digital Services Act)" („On the Risks and Side-Effects of the Digital Services Act (DSA)" Forthcoming, Kritische Vierteljahresschrift für Gesetzgebung und Rechtsprechung (KritV)/*Critical Quarterly for Legislation and Law*, March, 28. 2022, SSRN: https://ssrn.com/abstract=4068354 or http://dx.doi.org/10.2139/ssrn.4068354; Barata, "The Digital Services"; https://libertadinformacion.cc/wp-content/uploads/2021/06/DSA-AND-ITS-IMPACT-ON-FREEDOM-OF-EXPRESSION-JOAN-BARATA-PDLI.pdf.

779 Article 34 DSA.

780 Raphaël Gellert, *The Risk-Based Approach to Data Protection* (Oxford, UK: Oxford University Press, 2020): 27.

781 Article 34 (1)b, DSA.

fourth, the actual or foreseeable negative effects in relation to gender-based violence, to the protection of public health, minors, and serious negative consequences to the person's physical and mental well-being should be considered as systemic risks.[782]

Intentional manipulation of services has been devoted a separate paragraph, presenting it as a method that can influence the risks listed. It includes inauthentic use (e.g. fake profiles), automated exploitation of the service (e.g. social bots), and the amplification and viral dissemination of illegal content and information that is incompatible with platform terms and conditions. Emphasis is laid on assessing such misuse with attention to regional or linguistic aspects in the various Member states. In smaller language areas, giant platforms demonstrated a limited readiness to respond to deficiencies and needs.[783] While the biggest language areas (Germany, Spain, France, Ireland and Italy) benefited from enhanced policy actions and specific responses to fight Covid-related disinformation, smaller and less developed language areas (Slovenia, Slovakia, Bulgaria, Estonia, Malta and Croatia) received poor service in this respect and a low level of country-specific responses.[784]

When doing the assessment, VLOPs need to evaluate whether and how particular aspects of their systems manifest any of the systemic risks. This includes examining the design of their recommender systems or any other relevant algorithmic systems, their content moderation systems, their terms and conditions, their ad selection and presenting systems, as well as their data related practices.[785]

Once identified, VLOPs are obliged to mitigate the risk with reasonable, proportionate and effective measures that are tailored to the specific risks of their systems. The bulk of such mitigation will be found in the Code of Conducts (see later below). The Act also sets out a list (a-k) of measure types, enumerating activities that *may* be included among

782 Article 34 (1)d, DSA.

783 Zamparutti et al. *Developing a handbook on good practice in countering disinformation at local and regional level.* (European Union: European Committee of the Regions, Commission for Citizenship, Governance, Institutional and External Affairs. CIVEX, 2022) doi: 10.2863/066582.

784 Trisha Meyer, Alexandre Alaphilippe, and Claire Pershan, *The good, the bad and the ugly: how platforms are prioritising some EU member states in their COVID-19 disinformation responses* (European Unio: EU Disinfo Lab, 2021) https://www.disinfo.eu/publications/the-good-the-bad-and-the-ugly-how-platforms-are-prioritising-some-eu-member-states-in-their-covid-19-disinformation-responses/.

785 Article 34 (2) DSA.

the measurements. While the descriptions seem on the surface relatively detailed, the listed activity types define merely the areas where adjustments are expected, and reiterate obligations that had already been set out in previous sections of the Act. Among others, to adapt the design, features or functioning of their services, including their online interfaces, their terms and conditions, enforcement, content moderation processes, relevant decision-making processes and the dedicated resources for content moderation. This requirement adds weight to the already existing obligation of expeditiously removing illegal content (Article 16, applicable to all hosting providers).

This regulatory technique creates layers of liability and layers of responsibility: all hosting providers are *liable* to remove illegal content that they know of, otherwise, they can be held liable for the content itself. VLOPs, on the other hand, are *responsible* for *how* they fulfil this obligation, and also for reducing the likelihood of carrying illegal content – the latter is expressed by identifying illegal content as a systemic risk.[786] Similarly iterative are the requirements of testing and adapting their algorithmic systems including their recommender systems, adopting targeted measures aimed at limiting or adjusting the presentation of advertisements, initiating or adjusting cooperation with trusted flaggers and the implementation of the decisions of out-of-court dispute settlement bodies, as well as providing more information to users.[787] Other measures include initiating or adjusting cooperation with other online platforms or search engines through the codes of conduct and the crisis protocols; age verification and parental control tools, and finally, ensuring that deep fakes are prominently marked and that users are able to label those when they upload them.[788]

The significance of this list lies in the logic of co-regulation: sanctions and penalties can only be imposed in response to a breach of the Act. The preventive responsibility framework grants VLOPs autonomy in achieving the listed goals. Nevertheless, within the context of supervision, monitoring, and auditing, compliance with the listed actions is crucial in demonstrating due diligence. After all, participation in the co-regulatory codes remains voluntary (see more below).[789]

786 Article 35 (1) a, b, c DSA.
787 Article 53 (1) d, e, g, i. DSA.
788 Article 35 (1) h, j, k.
789 Article 45 DSA: Codes of Conduct "The Commission and the Board shall encourage and facilitate the drawing up of voluntary codes of conduct".

6.6 Self- and co-regulation as part of the legal regime

The various codes of conducts, envisioned by the DSA, would function as an extension of the law, a flexible augmentation that reaches beyond the possibilities of legal regulation.

Price and Verhulst[790] distinguished four types of self-regulation, based on the roles that state authorities play in their creation and enforcement. The authors stated that the perceived need by the industry to regulate an issue may either root in a threat of legal regulation, or a societal demand for increased responsibility, or economic factors.[791] In the light of this typology, the first self-regulation in the US was due to a combination of the drivers behind *coerced* and *voluntary* self-regulation, primarily based on the election influencing scandals in the US which was followed by hearings in the Congress of the leaders of Twitter, Google and Facebook. The actions were prompted partly by the looming governmental intervention, and partly by the need to satisfy users' needs and interests, with regard to the double markets: both consumers and advertisers.[792]

Self-regulation had traditionally meant not merely a code that a commercial or industrial interest group created, but also a body that dealt with complaints and takes decisions.[793] Even if the decisions taken by a private body were soft tools, and compliance was voluntary, they could deliver manifest expressions of non-compliance and entail exclusion from the group or deprivation of certain benefits of membership. Online intermediaries have not yet established such a body which would represent their

790 Monroe Price and Stefaan Verhulst, *The Concept of Self-regulation and the Internet* (The Hague: Kluwer Law International; Frederick, MD: Sold and distributed in North, Central and South America by Aspen Publishers, 2000).

791 Gibeon and Bollmann cite Price-Verhulst, p.4.: Gaia Gibeon, MPP 2022 Hanna-Sophie Bollmann, (2022) The spread of hacked materials on Twitter: A threat to democracy? A case study of the 2017 Macron Leaks. MPP 2022. https://opus4.kobv.de/opus4-hsog/frontdoor/deliver/index/docId/4493/file/Twitter_Regulation_France.pdf.

792 Gibeon and Bollmann, 2022.

793 Stefaan G. Verhulst and Monroe E. Price, "In Search of the Self: Charting the Course of Self-Regulation on the Internet in a Global Environment" (March 1, 2000). Available at SSRN: https://ssrn.com/abstract=216111 or http://dx.doi.org/10.2139/ssrn.216111.

industry, unlike advertisers, for example.[794] However, the recent decades opened up the definition to include a wide range of various industry standards.[795]

The European platform self-regulation codes have been top-down initiatives from the European Commission. The first examples of induced self-regulation in the field of online platform-regulation were the Code of Conduct to tackle illegal online hate speech in 2016 and the Code of Practice against Disinformation in 2018. The Code of Conduct to tackle illegal online hate speech is monitored by the European Commission without considerable consequences.

	Mandated Self-Regulation	Sanctioned Self-Regulation	Coerced Self-Regulation	Voluntary Self-Regulation
Government:	...formulates framework on basis of which industry should self-regulate	...approves/disapproves of self-regulation industry has developed	...threatens to enforce binding regulation if industry does not self-regulate	... has no formal relationship to the self-regulation

Figure 2. Price and Verhulst's typology, source: Gibeon-Bollmann, 2022.

Following the structure of the Price-Verhulst's division, the first Code of Practice was a type of "*Mandated Self-Regulation*", where the framework was formulated by the authorities, whereas the Strengthened Code of Practice (2022) corresponds to the type of "*Sanctioned Self-Regulation*", where the Commission can approve or disapprove of the industry self-regulation. The first Code of Practice, too, was developed together by the European Commission and the industry stakeholders. The Strengthened Code was based on the Guidance issued by the Commission,[796] and polished through negotiations between an honest broker and an independent consultant,[797] and industry stakeholders. This time, the Commission also has the authori-

794 Théophile Megali, "Digital Platforms as Members of Meta-Organizations: A Case Study of the Online Advertising Market," *M@n@gement*, 25, no. 2 (2022): 10–26. https://www.cairn.info/revue--2022-2-page-10.htm.

795 Robert Gorwa, "The platform governance triangle: conceptualising the informal regulation of online content," *Internet Policy Review*, 8, no. 2 (2019) https://doi.org/10.14763/2019.2.1407.

796 EC (2021) Guidance on Strengthening the Code of Practice on Disinformation, (COM(2021) 262 final, 26 May 2021, https://digital-strategy.ec.europa.eu/en/library/guidance-strengthening-code-practice-disinformation.

797 Oreste Pollicino, a Constitutional Law professor of the Bocconi University as honest broker and Valdani, Vicari and Associates (VVA) an independent consultant.

ty to oversee and approve or disapprove the self/co-regulative actions, i.e. the compliance with the Code. This was not the case with the previous Code, which was therefore regarded as toothless.

Additionally, DSA requires that all codes precisely define the required measures, objectives and key performance indicators, aiming to produce a more impactful code than the previous ones. Nevertheless, the first Code delivered the important benefit of organising the industry actors, structuring problem areas and suggesting possible solutions, effectively serving as a pilot to the eventual current co-regulation outlined by DSA.

Finally, there is another characteristic feature to the codes under DSA: they are explicitly to serve all stakeholders, including the needs of citizens at the Union level. Reference to the public interest objective is another distinctive factor between this new form of co-regulation and traditional forms of self-regulation. Serving the industry interests was often by enforcing ethical standards, complying with the laws, and attending the rights of consumers.[798] Still, reference to citizens has not been typical before outside the European legislator. In the context of EU law, exportation of public interest policy measures into co-regulation is a necessary compromise because of the limited legislative competences which extend to the protection of the internal market interests. This is another aspect of what I call as "augmentation of legal policy". Critics also refer to this phenomenon as "competence creep", for example, it is listed as type 4) of competence creep by Sacha Garben.[799] Regulating the behaviour of industry actors with the public interest objective would normally be a competence of the national legislators. However, national regulation is unlikely to be successful in this heavily globalised industry, especially with its asymmetric power relations, where giant corporations can simply ignore the requests of states with lesser economic power and a smaller market, in particular smaller language areas.

6.6.1 The hidden traps of auditing

The due diligence obligations and especially the risk-management obligations of VLOPs become meaningful mainly because of the annual audit that

798 Check out, for example, IAB Member Code of Conduct. https://www.iab.com/iab-member-code-of-conduct/.

799 Competence Creep Revisited, Sacha Garben, *Journal of Common Market Studies,* First published: 14 September 2017 https://doi.org/10.1111/jcms.12643.

will review and report on their compliance with Chapter III of the law, and with the Codes. Auditing organisations will have to be allowed access to all necessary information and the premises of VLOPs, receive responses to their questions, both oral or written, and they should receive all necessary cooperation and assistance from the audited organisation.[800]

The auditing must conclude with an audit report – it is the responsibility of VLOPs to ensure that this occurs. The audit report should include, among other elements, an opinion on whether the provider complied with the obligations and commitments. The possible outcomes are: positive, positive with comments, or negative. If the outcome is other than "positive", the report must make operational recommendations on the specific measures to achieve compliance with a deadline. The providers must adopt an audit implementation report within one month, which discusses how they will implement the operational recommendations. Should they not implement those, they must give the reasons and set out alternative measures that they have taken to achieve compliance. Both the audit report and the audit implementation report must be submitted to the responsible Digital Services Coordinator and the Commission without undue delay upon completion and be publicized within three months thereafter. Confidential information or information that might cause significant vulnerabilities for the platform security, undermine public security or harm recipients, may be removed from the publicized reports, but the DSC and the Commission still should receive the complete version.[801] The Commission may appoint independent external experts and auditors, to monitor platforms outside the normal course of auditing, and request the audit organisation – like any other entity, for that matter, – to provide information on the VLOPs.[802] If the Commission found on the basis of its investigations that the VLOP did not comply with the Act, with an interim measure, or with a commitment made binding specifically by the Commission, it may adopt a non-compliance decision.[803] Ultimately, the Commission may impose fines on VLOPs up to 6 % of their total global turnover.

The text of DSA makes no explicit reference on whether a negative audit report can lead to a Commission investigation and then sanctions. The Recitals discuss the matter of positive and negative reports, however, leave

800 Article 37 DSA.
801 Article 42 DSA.
802 Article 67 (1) DSA.
803 Articles 73, read together with 70 and 71 DSA.

this question in shadow. Thus, there seems to be no direct relationship between a negative audit report and the sanctions, because the Commission may pursue investigation on its own initiative, or following a request of the DSC.[804] An audit report may only indirectly lead to Commission action – whether an investigative action, or directly a non-compliance decision.[805]

According to the Recitals, a 'positive opinion' should be given where all evidence shows that the VLOP complies with the obligations laid down by DSA or, where applicable, any commitments it has undertaken *pursuant to a code of conduct or crisis protocol*, in particular by identifying, evaluating and mitigating the systemic risks posed by its system and services, and a 'negative opinion' should be given where the auditor considers that the VLOP does not comply with this Regulation or its undertaken commitments. The commitments mentioned here are distinct from "binding Commitments" defined above, that are specifically made binding by the Commission in the course of its investigative action.[806]

Besides the positive and negative opinion, an audit report can also contain a "positive opinion with comments", in which the auditor includes remarks that "do not have a substantial effect on the outcome of the audit".[807] It must be concluded that if the non-compliance with the Act or with the codes is deemed as substantial by the auditor, then it must give a negative opinion. However, this absence of clarity puts a considerable burden on the auditor: its opinion might or might not trigger the Commission to pass a non-compliance decision and eventually impose a fine. On the positive side, auditing should have a consequence, and the codes should flow into the sanctioning regime, otherwise we could not call it co-regulation. On the negative side, there are some theoretical and practical challenges that this deficiency raises or aggravates.

The theoretical problem emerges with the degradation of the risk-management and due-diligence framework into a binary positive/negative decision. The advantage of the autonomy of risk-management system, and the soft rules in the Codes, would have been to provide an auditing result that lies on a scale,[808] and this complexity is lost with the positive-negative

804 Article 65 (2) DSA. " If there is a suspicion that a VLOP has not complied with its obligations or systematically violated any other provision of the Code in a manner that seriously affects users."
805 Articles 66–73 DSA.
806 Recital 93 and Article 71 DSA.
807 Recital 93 DSA.
808 De Gregorio and Dunn, "Risk-Based Approaches,".

nature of the audit report. The purpose of auditing should be providing a qualitative and quantitative analysis of the measures deployed by VLOPs to fulfil their obligations, concluding indispensably with an assessment. This assessment should, however, still reflect the complexity of the obligations, and not be reduced to a yes/no answer.

The practical challenges have existed even without this extreme burden on auditors. The primary risk would be the high concentration of the auditors' market and their dependence on VLOPs. Only VLOPs need to care for their audit, which means that a few, but very powerful actors must be regularly audited. To minimise the risk of audit capture, DSA has ordered that auditors must not have provided non-audit services to the service provider 12 months prior and after the auditing, and must not have provided audit services to that provider during a period longer than ten years.[809] A limitation of this kind is obviously necessary, however, it is debatable whether the time spans are reasonable, and how effective the restrictions will be. Currently, 19 companies are identified as VLOPs and VLOSEs[810] but not all of them signed the Strengthened Code of Practice on Disinformation,[811] and Twitter even unsigned the Code in May 2023.[812] The market of audit-bound platforms will remain rather small, and the ten-year ban will further reduce it for auditing firms. Research has found that auditing exercise can actually benefit from somewhat longer periods of ongoing client relationship, and that the most fraud occurs during the first year of auditing a new client.[813] Accessing the relevant data, as well as adequately processing and analysing those, improves with experience. Taking the other limit under scrutiny, the one-year ban for providing "other services" appears rather disproportionately short. This allows that a company for digital solutions can develop an algorithmic system for a VLOP, and

809 In both cases, the very large online search engine and any legal person connected to that provider are also included. Article 37 (3) DSA.

810 https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses.

811 On the current status of signatories see: Molly Killeen, "Code of Practice on Disinformation signatories regroup with AI focus," *Euroactiv*, last modified Jun 6. 2023. https://www.euractiv.com/section/platforms/news/code-of-practice-on-disinformation-signatories-regroup-with-ai-focus/.

812 Natasha Lomas, "Elon Musk takes Twitter out of the EU's Disinformation Code of Practice," *Techcrunch* Last modified May 27, 2023. https://tcrn.ch/43tQ8ml.

813 Johann Laux, Sandra Wachter, and Brent Mittelstadt, „Taming the few: Platform regulation, independent audits, and the risks of capture created by the DMA and DSA," *Computer Law & Security Review*, 43, Nov. 2021, 105613.

a year later already be eligible to provide auditing services on the same system.

Further difficulties are suspected because of the lack of a benchmark for the quality of auditing services. The numerous requirements in the Act and in the codes need to be interpreted in the following years, striving to establish, how much effort is enough to demonstrate due diligence? The Commission may partially fill in this gap by adopting delegated Acts laying down detailed rules for the audits, in particular the necessary procedural rules, methodologies and reporting templates.[814] In addition, relevant European and international standardisation bodies are expected to set up voluntary standards among others, for auditing.[815]

Even with the help of detailed templates and standardisation, the binary judgement that auditors must conclude to, will reduce their possibilities for weighing, and leave those auditing firms with an oversized responsibility for the consequences of their reports. In light of this burden, the ten years' ban may even appear as a relief, giving ample time to run a Commission non-compliance procedure, a fine and a recovery after such a negative experience, before the same auditor company even comes into the pool as a potential business partner again. However, given the huge market power of VLOPs, an audit firm that produced a negative audit report may be less popular with other VLOPs, too. Lost business opportunities won't be filled in by smaller platforms, as they are not obligated to engage with auditing. The highly concentrated market on the demand-side (VLOPs) is likely to cause concentration on the supply-side (auditors) and capture appears hardly avoidable, despite the rules. Beyond their proven expertise, auditors must also prove objectivity and professional ethics, based on appropriate professional codes and standards.[816]

The two main areas of auditing are not equal in terms of their consequences: non-compliance with the law, including due diligence obligations, may entail a fine. These comprise the notice-and-takedown regime, the transparency obligations, and the many other due diligence obligations that we have discussed above. When it comes to the assessment and mitigation of systemic risks, the obligations become more complex, and their fulfilment can be assessed on a scale, rather than with a binary answer. Systemic risks are not defined, merely examples are given. This allows the term to

814 Article 37 (7) DSA.
815 Article 44 (1) e. DSA.
816 Article 37 (3) b, c. DSA.

remain open-ended, ready to absorb new risk areas with time. Should new significant systemic risks emerge, the Commission may invite providers to generate new codes of conduct.[817] However, non-compliance with the commitments in the codes becomes risky only if it reaches the level of systemic infringement of the risk mitigation obligation.[818] On the mitigation of risks, the VLOPs separately report to the DSCs, which provide a summary report to the Commission. Therefore, the Commission will have parallel sources of information about the behaviour of VLOPs on which it can base its eventual non-compliance decision.

6.6.2 The incentives, execution and objectives of the codes under DSA

Interestingly, no explicit rule obliges service providers to participate in the drafting of, or to sign the codes of conduct. The Act merely says that "the Commission and the Board shall encourage and facilitate the drawing up of voluntary codes of conduct", and that the Commission "may invite" industry actors to participate in the drawing up of codes of conduct.[819]

Penalties can be levied only for non-compliance with the provisions of the Act, with interim measures, other orders and with binding commitments.[820] The obligation is to mitigate the risks, which requires the platform to "put in place reasonable, proportionate and effective mitigation measures". Even if these adjectives ("reasonable, proportionate and effective") are still not objectively definable, at least they have a considerable legal history having been used in various fields, whether separately or together.[821] The relative vagueness of the expectations makes reference to the recommended actions meaningful (Article 35(1)a-k). If the listed actions remain unattended by a VLOP, they are less likely to prove that they complied with their duty to apply reasonable, proportionate and effective measures.

817 Article 45 (2) DSA.
818 Article 35. DSA.
819 Article 45 DSA.
820 Article 51, 52, 74 DSA.
821 For instance, used together in the Regulation (EU) 2017/1128 of the EP and of the Council of 14 June 2017
on cross-border portability of online content services in the internal market. See also in the Council of Europe Recommendation CM/Rec(2016)3 on human rights and business.

6.6.3 The soft power of the codes

The new codes under DSA need to contain concrete measures, key performance indicators, and aim to ensure that the participants report regularly on the measures they take. Their drafting and regular review is facilitated and encouraged by the Board and the Commission, which assess, regularly monitor, and evaluate the achievement of the objectives, and publish their conclusions.[822] The soft power of the Commission and the Board is expressed in their right to "invite the signatories [...] to take the necessary action", and their scrutiny on the application of the Code.[823] Still, the Code remains voluntary and for example Twitter withdrew its signature from the Code of Practice on Disinformation.[824]

Specific codes of conducts are the Code of Practice on Disinformation strengthened in 2022, the Code of Conduct against Illegal Hate Speech that had been issued in 2016, and planned to be reviewed after 2023; the codes of conduct for online advertising[825] and the codes of conduct for accessibility, to address the needs of persons with disabilities. Both need to be developed by 18 February 2025.[826]

The codes of conduct on online advertising are supposed to provide transparency throughout the value chain of advertising, and not only of VLOPs, but of a wider scope of participants: general online platforms, providers of online advertising intermediary services, and other actors involved in the programmatic advertising value chain, or user representation organisations.[827] Civil society organisations and relevant authorities are expected to participate in preparing the code. Search engines are included only if they are very large. Important transparency obligations have been made compulsory, such as information on the advertiser (on whose behalf the ad is presented), the sponsor (who paid for the advertisement), and the

822 Article 45 DSA.
823 As Recital 103 DSA says "The public oversight should not impair the voluntary nature of the codes and the freedom of the interested parties to decide, whether to participate".
824 Ewa Krukowska, "Twitter Withdraws From EU Disinformation Code, Commissioner Says," *Time* May 27, 2023. https://time.com/6283183/twitter-withdraws-from-eu-disinformation-code-commissioner-says/.
825 Article 46 DSA.
826 Article 47 DSA.
827 Article 46 DSA.

targeting criteria for all providers,[828] and additional requirements apply for VLOPs, which should include the content, the period of advertising and the total number of users reached (see in Chapter 6.5). The Code should, in addition, require adding and facilitating the transmission of meaningful information on the monetisation of data. The preamble of DSA implies that other, previous self-regulatory codes, such as the Product Safety Pledge, the Memorandum of understanding on the sale of counterfeit goods on the internet, the Code of conduct on countering illegal hate speech online should also be adapted to the new, stricter requirements of the DSA.[829]

6.6.4 The Strengthened Code of Practice on Disinformation

The Strengthened Code of Practice on Disinformation (herafter: Code) have been substantially extended compared to its original form in 2018. It became inclusive of other themes beyond disinformation, such as political and issue advertising, integrity of services, empowering users, as well as co-operation with researchers and the fact-checking community. All platforms may join the Code by signing, without regard to their size and field of activity, as well as fact-checker and researcher organisations, players from the advertising ecosystem, and civil society organisations.[830]

As described above, the systemic problems of platform operation have arisen from online behaviour that was not strictly illegal. Instead, it has been the systemic and large-scale impact, which negatively affected the societal functions of the information environment. The phenomena of the "information disorder"[831] is based on a diverse range of services and applications that are flexibly and rapidly changing. The rigidity of legal regulation would not be appropriate to keep up with their fluid nature in defining counter-actions.

The Strengthened Code of Practice starts with an explicit statement that signing up to all commitments that are relevant and pertinent to their service should be considered as a possible risk mitigation measure under

828 Article 26 (1) DSA.
829 Recital 106 DSA.
830 Point (o). Preamble.
831 Claire Wardle and Hossein Derakhshan, (2018) *Information disorder: Toward an interdisciplinary framework for research and policymaking* Strasbourg: Council of Europe, 2017) file:///D:/Letoltesek/162317GBR_Report%20desinformation.pdf.

the DSA.[832] Thus, platforms can expect that they fulfil their legal obligation of risk mitigation if they comply with their undertaken commitments.

The previous Code of Practice tackling Disinformation (2018) was criticised for being vague, for absence of standards for its evaluation and reporting, for lack of oversight on compliance, lack of sanctions for non-compliance, and for absence of independent data to check the reports by platforms themselves.[833] Indeed, ERGA has found in its cooperation with national regulators that platforms have reported their achievements in a better light than in reality.[834]

As a great difference to the previous Code of Practice, each Commitment in the Strengthened Code is followed by various (facultatively applicable) measures, which are followed by Qualitative Reporting Elements (QREs) and Service Level Indicators (SLIs). The signatories can pick and choose from the list of measures that they decide to implement. Each QRE or SLI defines clearly to which measure it is applicable. Concrete formulation of the measures, QREs and SLIs is the interest of all stakeholders, to eliminate the risk arising of legal insecurity.

A Task-force should regularly adapt the measures to the changing technological, societal, market and legislative developments, as necessary. The signatories commit themselves to participate in the Task Force, which consists of the signatories' representatives, and that of the ERGA, EDMO, External Action Service (EEAS) and is chaired by the European Commission, and may invite experts as observers and to support its work. It will also work in subgroups and workstreams, and exchange information on trends, tactics, techniques and procedures of disinformation, or otherwise employed by malicious actors. Moreover, it will establish a risk assessment methodology and a rapid response system for special situations like elections or crises, including cooperating and coordinating their work in these situations.[835]

About one third of the Code is devoted to procedural commitments related to monitoring, compliance and sustaining the Code: to set up and maintain the Transparency Centre, to participate in the permanent Task

832 I. Preamble, (j) of the Strengthened Code of Practice.

833 Elda Brogi and Konrad Bleyer-Simon, (2021) "Disinformation in the Perspective of Media Pluralism in Europe – the role of platforms, in *Perspectives on Platform Regulation,* ed. Judit Bayer et. al. (Baden-Baden, Nomos Verlagsgesellschaft mbH & Co. KG, 2021): 531–548.

834 Ibid. at 537–538.

835 Chapter IX. of the Strengthened Code of Practice.

Force and to actively be committed to implement the Code, to cooperate with the Commission in crisis situations, by reporting on their actions, and in particular, to be audited at their own expense for their compliance with the commitments undertaken in the Code.

6.6.5 The Code's content: Reordering the information landscape

Disinformation is intentionally created, strategically distributed, false or misleading information that has a hidden agenda: political or economic incentive that is not obvious from the content.[836] Misinformation is false or misleading information without the intentional and strategic element. The distinction of truth and falsity presents a deep semantic, philosophical and legal problem, therefore, falsity alone cannot and should not be the basis of any regulation. The other characteristics of disinformation also cannot be promptly established without careful examination, balancing, and knowledge of the context. Therefore, focusing solely on the content would not only be restrictive of the freedom of expression principle, but it would also not be successful.

Therefore, the Code avoids focusing directly on the content and instead presents a comprehensive system addressing key elements of the platform communication environment. Although removal of disinformation is sideways mentioned, it is not in the centre of the regulation. Where mentioned, it is limited to "harmful" disinformation. In fact, disinformation or misinformation are not even defined in the Code – that is left over to signatories. The elements of this communication environment can be listed into two groups: key actors, and behavioural subsystems. Key actors are the users, the researchers and fact-checkers, who are to be empowered by platforms, in order to balance the informational asymmetries. The key subsystems of this informational environment are the advertising ecosystem, and the "impermissible manipulative behaviours and practices".

In the context of advertising, signatories should strive to deprive disinformation from its funding (a.k.a. defunding, or demonetising disinformation), in other words, to avoid placement of advertising next to disinformation. Further recommended are various brand safety actions, counting on

836 Wardle and Derakhshan, *Information disorder*; Bayer et al., *Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States* (Strasbourg: European Union, 2019).

the interest of advertisers to have their ads displayed next to trustworthy, good quality content, rather than one that would be disapproved by their clients. The actions are extended to the entire value chain of advertising including not only advertisers but also online e-payment services, e-commerce platforms, crowd-funding or donation systems.[837]

The rules on political advertising (Chapter III of the Code) replicate the rules in the draft Regulation on Political Advertising Transparency.[838] The purpose is partly to apply those rules already before the Regulation steps into force, and partly to develop detailed know-hows and procedures for the deployment of the legal principles. (See the Chapter 8 on political advertising).

In the context of manipulative behaviour, (Chapter IV of the Code) signatories are expected to agree on a cross-service common understanding of impermissible manipulative behaviours, actors and services, which allows to tackle more generally disinformation, misinformation and manipulation.[839] This requires Task-Force action to create a list of shared terminology and to keep it regularly updated. While this list should be reviewed and updated, the Code offers a "starting kit", including fake accounts, bot-driven amplification, hack-and-leak, impersonation, malicious deep fakes, purchase of fake engagements, and various activities of paid trolling: the common element is inauthentic behaviour. Deepfakes and other AI-generated and manipulated content should at least be transparent.[840] The latter (AI generated content) may gain more importance in the future with ever newer chatbots on the market that are able to generate convincing, but potentially baseless statements on any question, as demonstrated by ChatGPT[841] and the disinformation generated and disseminated by X's chatbot.[842]

837 Chapter II, Commitment 1.

838 Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising. COM/2021/731 final.

839 Chapter IV, Commitment 14. of the Strengthened Code of Practice on Disinformation.

840 Commitments 15 and 16.

841 Ethan Mollick, "ChatGPT Is a Tipping Point for AI," *Harvard Business Review* December 14, 2022 https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai.

842 James Thomas, "No, Iran has not started attacking Tel Aviv." *Euronews*, April 11, 2024. https://www.euronews.com/my-europe/2024/04/11/no-iran-has-not-started-attacking-tel-aviv.

The second part of the Code is concerned with empowering the key actors. The empowerment of users includes enhancing media literacy, critical thinking, safe design of the architecture, among many others. The measure "safe design" applies a different vector for empowerment and may become one of the most far-reaching measures. This entrusts even more power on platforms: instead of requiring neutral transmission, on the contrary, it requires them to improve the *prominence* of authoritative information and to *reduce the prominence* of disinformation.[843] Furthermore, signatories also commit to develop policies that aim at prohibiting, *downranking* or not recommending harmful, false, or misleading information.[844] These measures explicitly empower platforms to structure the public discourse, taking on themselves the responsibility to follow ethical standards similar to editorial ones. This activity is likely to have massive implications for the structuring of the public discourse. On the one hand, platforms are already influencing this, but with a goal of increasing engagement and maximising revenues.[845] On the other hand, the entrusted power without the requirement of viewpoint neutrality creates a significant risk. With the commitment of prioritising trustworthy information and deprioritise disinformation, signatories undertake partially public-service-like obligations. Through these requirements, platforms and search engines would be taking responsibility for their real activity, that constitutes the core of their services: organising and ranking content. The addressed policies focus mainly on content management, although content regulation (moderation) is also part of the set.

This public service obligation signals a key turning point in the regulatory attitude, because it builds on platform autonomy. Using the philosophy of eastern martial arts, the law does not fight against platform power: it acknowledged this power and attached responsibility to it, thereby imposing even more power on them.

Commitment 18 of the Code is an implied acknowledgement of the fact that platforms are *not neutral* intermediaries, that they indeed *form* the public discourse and public opinion with their algorithmic governance.

843 18.1. SCOP.

844 18.2. SCOP.

845 The means to achieve this end were maximising user engagement, and prioritising content that attracted the most users, which paved the way for the attention-exploitative, manipulative techniques and disinformation that ultimately caused considerable social and political changes. Hoffmann-Riem, *Recht im Sog*.

Not even limits are set against this activity, other than transparency requirements. The few safeguards are limited to fairness towards users, and are placed among the general rules of DSA. Providers must inform users even about demoting their content, with a clear and specific statement of reasons.[846] Whether this is practically possible in regard of each and every downranked content, is highly questionable. Even though individual attention to each content piece is not expected under the SCOP, because providers only need to take action on actors that *persistently* violate their policies, and they need not react to sporadic events. At the same time, algorithmic content governance is likely to sweep in sporadic events as well. Besides, providers must include in their transparency reporting the number and type of each measure that affected even the visibility of user content.[847]

Beyond these rules on procedural fairness, providers are free to decide what they interpret as harmful and what as trustworthy or "authoritative". The DSA merely prescribes that their terms of services (TOS) must have "due regard to the rights and legitimate interests of all parties involved, including the fundamental rights of the recipients of the service, such as the freedom of expression, freedom and pluralism of the media, and other fundamental rights and freedoms as enshrined in the Charter."[848] Nevertheless, platforms are not left to define these standards by themselves but in cooperation and consultation with the Task-force which also includes experts from EDMO and other organisations, plus platforms may also delegate their own experts. For instance, defining what authoritative content is, may be a precarious task especially in times of global political tensions when several states in and outside the EU are captured by populistic governments that pursue propaganda and where authorities are the source of disinformation.[849] Further commitments that involve cooperating with the researchers' community and the fact-checking community are likely to inform platforms in this respect too, among others.[850] Finally, neither the DSA, nor the

846 Article 17 DSA.
847 Article 15 (1) c. DSA.
848 Article 14 (4) DSA and Recital (47).
849 Samuel C. Woolley and Philip N. Howard, *Computational propaganda worldwide: Executive summary* Working Paper 2017.11. Oxford, UK: Project on Computational Propaganda. demtech.oii.ox.ac.uk. (2017); Erin Kristin Jenne, András Bozóki, and Péter Visnovitz. "Antisemitic Tropes, Fifth-Columnism, and "Soros-Bashing"," in *Enemies Within: The Global Politics of Fifth Columns* (Oxford University Press, 2022): 45.
850 Chapter VI.-VII, Commitments 26–33.

SCOP rule out systematic content-based discrimination, or discrimination between users, provided that it is not part of the TOS. From a theoretical point of view, the regulator has evaded important constitutional dilemmas. First, by placing this public-service-like obligation into the Code, rather than the law, the regulator evaded the "freedom of expression conundrum" posed by disinformation, which cannot be directly tackled because it is not illegal at the level of the content. Second, by outsourcing this as a platform obligation, it seemingly removed the problem from being a constitutional question into an administrative and compliance question.[851]

The second large subsystem that is tackled in the SCOP is user empowerment. This aims to elevate users in a position that is closer to the platforms' playing field, and support them in making informed choices, in their informational behaviour. Besides the humanitarian perspective, this is also based on the observation that social media content, being an interactive service, is formed to a large extent by users themselves. Not only is a large part of the content provided by users, but they influence the algorithmic ranking through their likes, shares and other reactions. Therefore, spreading disinformation (and hate speech, etc.) is partly users' responsibility. Legal and policy discussions work with different models of the user community: on the one hand, users are regarded as rational beings who take conscious decisions, on the other hand, as vulnerable victims of a malfunctioning system. The first perspective with the picture of inherently good and rational people was the underlying preconception of the early internet optimists,[852] who saw the internet as a place where governmental intervention is unwelcome. A more moderate view of reasonable and conscious audience is still popular.[853] At the same time, empirical research also suggests that certain

851 Jack M. Balkin, (2023) "Free Speech Versus the First Amendment," *UCLA Law Review, Forthcoming – Yale Law & Economics Research Paper Forthcoming* Apr 19, 2023: 19.

852 John P. Barlow, "A Declaration of Independence of the Cyberspace," *Electronic Frontier Foundation* Feb. 8, 1996, www.eff.org/cyberspace-independence.

853 Natali Helberger, Kari Karppinen, and Lucio D'Acunto, "Exposure diversity as a design principle for recommender systems," *Information Communication and Society* 21, no. 2 (2018): 191–207. DOI: 10.1080/1369118X.2016.1271900., Judith Möller et al., "Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity," *Information Communication and Society* 21, no. 7 (2018): 959–977. DOI: 10.1080/1369118X.2018.1444076, Philip Michael Napoli "Exposure Diversity Reconsidered," *Journal of Information Policy* 1, no. 2 (2011): 246–259. Available at: https://www.jstor.org/stable/10.5325/jinfopoli.1.2011.0246 Natali Helberger, Katharina Kleinen-von Königslöw, and Rob van der

traits make people more likely to share disinformation.[854] Users' cognitive predisposition and human vulnerabilities are also diverse and bring about a variety of patterns.[855] Ample research discusses users' susceptibility and divergent attitudes towards disinformation.[856] In reality, the user community is the opposite of homogeneous: it includes political, business, and criminal actors. The anomalies in the information environment should be viewed as various conflicts between the different users' interests. Users' predisposition and concluding action has been amplified by the possibilities of technology, and will be further increased with the advent of generative and general purpose AI applications. However, the ethical perspective whether users have a moral responsibility has been yet under-researched.[857]

Noll, "Regulating the new information intermediaries as gatekeepers of information diversity", *info* 17, no. 6 (2015): 50–71. https://doi.org/10.1108/info-05-2015-0034.

854 Nir Grinberg at al., „Fake news on Twitter during the 2016 U.S. presidential election," *Science,* 363, no. 6425 (2019): 374–378. doi: 10.1126/science.aau2706 – on how older and more politically right-leaning people are more likely to share disinformation. See also Jay J. Van Bavel et al., "Political Psychology in the Digital (mis)Information age: A Model of News Belief and Sharing," *Social Issues and Policy Review,* 15, no. 1 (2021): 84–113. doi:10.1111/sipr.12077 – found that those who perceive society to be more polarised, are more likely to share disinformation.

855 Angela Anthony and Richard Moulding, "Breaking the news: Belief in fake news and conspiracist beliefs," *Australian Journal of Psychology* 71, no. 2 (2019): 154–162. doi: 10.1111/ajpy.12233, Roland Imhoff and Pia Karoline Lamberty, "How paranoid are conspiracy believers? Toward a more fine-grained understanding of the connect and disconnect between paranoia and belief in conspiracy theories," *European Journal of Social Psychology* 48, no. 7 (2018): 909–926. doi: 10.1002/ejsp.2494; See: Lea-Johanna Klebba and Stephan Winter, (2021, January 22), "Selecting and sharing news in an "infodemic": The influence of ideological, trust- and science-related beliefs on (fake) news usage in the COVID-19 crisis," Preprint retrieved from https://doi.org/10.31234/osf.io/dbghp, last modified February 8, 2021. See: Pierre, J.M. (2020). "Mistrust and Misinformation: A Two-Component, Socio-Epistemic Model of Belief in Conspiracy Theories," *Journal of Social and Political Psychology*, 8, no. 2 (2020): 617–641, doi: 10.5964/jspp.v8i2.1362.

856 Corinne Tan, "Regulating disinformation on Twitter and Facebook," *Griffith Law Review* 31, no. 4 (2022) DOI: 10.1080/10383441.2022.2138140. Judit Bayer et al., "Disinformation and Propaganda: Impact on the Functioning of the Rule of Law and Democratic Processes in the EU and Its Member States – 2021 Update: Study" (Strasbourg: European Parliament, 2021).

857 Elements of the moral responsibility for contributing to collective harm are discussed by Rainer Mühlhoff and Hannah Ruschemeier, "Predictive Analytics Und DSGVO: Ethische Und Rechtliche Implikationen," in *Telemedicus – Recht Der Informationsgesellschaft, Tagungsband Zur Sommerkonferenz 2022*, Frankfurt am Main: Gräfe and Telemedicus, 2022): 38–67.

The discussed policy requires service providers to support users in taking rational decisions, with commitments such as empowering users with tools to assess the provenance, edit history, and authenticity of the digital content, better equipping users to identify disinformation, and to make more informed decisions when they encounter online misinformation or disinformation. These provide useful defensive tools for users without imposing on them obligations to use those tools. Obligations are imposed merely on service providers, which may include nudging users who try to share misinformation, and many other actions including cooperating with fact-checkers, adding labels, and ratings. However, the tools are not exclusively defensive: for instance, flagging harmful content is a power that may be misused and potentially cause harm to other users.[858] Platforms are warned to take steps that the tool remains duly protected from abuse (both human and machine-based), such as mass-flagging to silence other voices.[859]

6.6.6 Interim summary on DSA and the Code of Practice on Disinformation

The Code's regulatory style reflects a few important characteristics. First, the obligations carry some resemblance to public service obligations. They reflect the recognition that the information-ordering function of platforms is so dominant and of such a universal importance for social discourse, that it should not be restricted, rather, qualitative expectations are set. Second, they aim at empowering the participants in the network who are in a power asymmetry against platforms' power: uses, fact-checkers, researchers, and a wide industrial cooperation including experts and other organisations. Third, it strives for slight correction of excesses by the data-driven advertising economy.

Further, the Code builds on the cooperation of all stakeholders, involving users, researchers, civil society and authorities; and allows industry actors to adapt their policies to the practical and technical possibilities, in order to lay the ground for a better governance of publicly available content.

The aim is high: to achieve a new balance in the public discourse that is so necessary for democracy. However, the insecurities are manyfold: the

858 Commitment 23.
859 Commitment 23.2.

recommendations might be only superficially applied; they might lead to overcensorship; new services may emerge with new communication structures. The current speed of technological development especially with the advent of generative AI, is unprecedented, therefore, public communication is bound to further change. The Code aims at power relations and not content, but even power relations may change as generative AI puts yet new actors – in this case, content providers – into the market. In any case, the DSA and with it the SCOP resembles more a research project than a regulation: it builds on the collection of a vast amount of information about the regulated services, and presumes a continuous cooperation between the actors, discussion, negotiations, and assessments. With the assessment of compliance, it projects an ongoing power game between the Commission and giant platforms. Nevertheless, other stakeholders, for example advertisers are also forced to think about their strategies and enter into dialogues with the other actors. As a further tactical step, the legislative package creates a dense network of bodies for standard-setting, scrutinising, or supervising. These hedge the risks that the entrustment of platforms carries by softly guarding their every move.

6.7 The Code of Conduct tackling illegal hate speech

The first induced self-regulatory instrument that had been drawn up between the Commission and the largest online platforms (Facebook, Microsoft, Twitter and YouTube, initially) had been the Code of conduct on countering illegal hate speech online in 2016.[860]

This relatively simple and brief document (3 pages) includes a commitment by the signatories to review notifications, remove illegal hate speech, and clarify this in their Terms of Services. They are to rely on the EU Framework Decision against racism and xenophobia[861] in this respect, and have a dedicated team to review notifications, possibly in less than 24 hours. This is the document that originally introduced the concept of trusted flaggers (there: "trusted reporters") who are to provide high quality notices. Additionally, it contains some general aims to promote counter-

860 Code of conduct on countering illegal hate speech online, 30 June 2016. https://commission.europa.eu/document/551c44da-baae-4692-9e7d-52d20c04e0e2_en.

861 Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law.

narratives, new ideas and support educational programs that encourage critical thinking, in collaboration with the Commission. It includes an agreement for regular assessment of the commitments. The Commission has been monitoring the implementation of the Code annually. The results have indicated an 'improvement' in the initial years followed by a relapse in 2022. However, the monitoring measured only how swiftly the platforms complied with the notifications and the rates of removal – the more removal, the 'better' the enforcement was deemed.[862] It never scrutinized the merit of the decisions or the content of the notifications. The number of false positives remains therefore unknown. The ratio of the removed illegal content to the remaining, undetected or unremoved illegal content was not measured. Therefore, the monitoring has not delivered information on whether the entire system is appropriate to deal with online illegal hate speech.

The limitations of this Code are obvious by today. This served mainly to cover the regulatory loophole of online platforms whose status was ambiguous under the ECD. The DSA has built on the experiences collected through the Code and covers illegal hate speech now.

A further limitation is that the Code, being based on the Framework decision, includes only racism and xenophobia and does not include gender-based hate speech or sexual orientation-based one, which makes 15,5 % of all grounds of hatred reported in 2022 (pars with Anti-Gypsism and Xenophobia at 16,8 and 16,3 %).[863]

To tackle these areas, the European institutions have been preparing new legislation. A proposal for a Directive on violence against women has been published in March 2022.[864] Among violent crimes, also the sharing of non-consensual sexual images, cyber stalking, cyber harassment and cyber incitement are regulated. In 2021, the Commission, following an EP

862 EC (2023) EU Code of Conduct against online hate speech: latest evaluation shows slowdown in progress. (Press Release) https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7109.

863 Countering illegal hate speech online, 7th evaluation. of the Code of Conduct. https://commission.europa.eu/system/files/2022-12/Factsheet%20-%207th%20monitoring%20round%20of%20the%20Code%20of%20Conduct.pdf.

864 Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence. COM/2022/105 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0105.

Resolution in this regard,[865] published a Communication on the plan to extend the list of EU crimes to hate speech and hate crime.[866] This initiative aims to reach a Council decision that would extend the current list of EU crimes in Article 83 (1) of the Treaty on the Functioning of the European Union to hate crimes and hate speech. This is needed in order to pass then a directive that would harmonise the definition and the sanctions across the Member States. Importantly, the new definition is meant to include sex, sexual orientation, age and disability as protected characteristics, which are currently protected under the Charter of the Fundamental Rights (Article 21), but not by the Framework Decision.[867] A Treaty amendment would require a unanimous decision in the Council which is unlikely due to unresolvable disagreement on the values to be protected.

6.8 Crisis protocols

The DSA envisaged that certain voluntary crisis protocols are drawn up to ensure prominence of official information in crisis situations. The specific extraordinary situations that are addressed here are strictly related to public security or public health.[868] The crisis protocol may only be initiated by the Commission, upon recommendation by the Board, and completes the compulsory crisis response mechanism of VLOPs, which may apply in any crisis situation. Under the crisis response mechanism, the Commission may, if recommended by the Board, require VLOPs to do certain individual measures in order to address the crisis.[869]

The voluntary crisis protocols are meant to coordinate a rapid, collective and cross-border response, for example, when online platforms are misused for the rapid spread of illegal content or disinformation, or where the need arises for rapid dissemination of reliable information. The official in-

865 European Parliament resolution of 16 September 2021 with recommendations to the Commission on identifying gender-based violence as a new area of crime listed in Article 83(1) TFEU.

866 COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL. A more inclusive and protective Europe: extending the list of EU crimes to hate speech and hate crime, COM/2021/777 final. 9. December 2021.

867 Study to support the preparation of the European Commission's initiative to extend the list of EU crimes (2021).

868 Article 48 DSA.

869 Article 36 DSA, Crisis response mechanism.

formation may be provided by Member States, the EU authorities or other relevant bodies. The crisis protocols may be activated only for a limited period of time, and the measures should also remain limited to what is strictly necessary; they should not amount to a general obligation of monitoring.[870] Platforms should provide for the prominent display of official information on the crisis situation, designate a point of contact for crisis management, and if necessary, increase resources to ensure compliance with the DSA, in particular Articles 16 (notice-and-action), 20 (internal complaint-handling), 22 (trusted flaggers), 23 (measures and protection against misuse) and 35 (mitigation of risks).[871]

What exactly counts as a crisis situation, who and in what procedure shall determine its occurrence, its end and when the crisis protocol should be is activated, is to be defined in the protocols themselves. These should also include safeguards to address any negative effects on fundamental rights, in particular the freedom of expression and information, and the right to non-discrimination.[872] The concept is somewhat similar to that of the Audiovisual Media Services Directive which provides that Member States shall ensure that emergency information are broadcast through audiovisual media services natural disaster situations, even in a manner which is accessible to persons with disabilities.[873]

6.9 Summary on the DSA and its regulatory structure

The DSA updated and solidified the European common grounds for digital services. In comparison to the DMA, which applies to gatekeepers, the DSA applies to a wider scope of actors: all service providers, including mere conduits, search engines, and all hosting providers.[874] At the top of the pyramid-like scope of subjects to the DSA sit VLOPs (very large online platforms and very large online search engines). DSA opted not to create a new category for platform services, but to define them as a subcategory of hosting service providers. The dilemma of regulating content while preserv-

870 Recital 108. DSA.
871 Article 48 (2) DSA.
872 Article 48 (3) DSA.
873 Article 7a AVMS Directive.
874 Paal/Kieß: Digitale Plattformen im DSA-E, DMA-E und § 19a GWB(ZfDR 2022, 1), p. 14.

ing freedom of expression, was addressed by constructing a large part of the regulatory material based on semi-voluntary co-regulatory codes. The Strengthened Code of Practice is a 'first of its kind', and likely a pioneer of a range of new codes, whose oversight is regulated in the Act. This mixture of self-regulation and legal oversight may raise the DSA to the level of a new "constitution" of the digital space, on the one hand. On the other hand, there are concerns that compliance may fall short of the expectations.[875]

The regulatory structure of DSA appears to be based on the recognition that digital space is a moving target: by the time a piece of regulation steps into force, it may be outdated, as it happened with ECD. For this reason, DSA's regulation is processed-based: it has created processes which would:

a) generate a continuous dialogue between authorities and service providers, that would allow a steady adaptation of the requirements to the changing circumstances,
b) deliver vast quantities of data about the functioning of service providers, which, if processed and analysed, may lead to additional conclusions and point at new regulatory needs or solutions,
c) involves users to a higher extent, building on their conscious and informed decision-making.

In fact, rather than eliminating a problem, DSA structured the problem and assigned new tasks to all parties and stakeholders: service providers, authorities, users, civil society actors, researchers, and the Commission. Provided that all parties perform their duties at a high level, this cooperation can lead to an equilibrium between the interests and rights of the stakeholders and citizens, and serve society as a whole. In this perspective, the DSA can be seen as a societal endeavour, or a collaborative research programme in which collective participation is essential. Therefore, the DSA should not be regarded as the end of a regulatory process, rather a beginning of a long and intricate journey. Through this journey, we can hope to get more precise information about how online intermediaries impact society, politics and the economy.

875 Kuhlmann/Trute: Die Regulierung von Desinformationen und rechtswidrigen Inhalten nach dem neuen Digital Services Act (GSZ 2022, 115) p. 122.

6.10 The relationship of AVMSD, ECD and DSA

DSA further elaborated the co-regulatory model that the AVMSD introduced, making the model more consistent and sophisticated. A significant novelty is the relatively clarified position of platforms within the chain of intermediaries. The regulation of video-sharing platforms under the AVMSD was not in harmony with the ECD's definitions of personal scope, because the definitional elements that ECD used for establishing the liability categories, did not fit platforms. The legal status of platforms remained ambiguous until the DSA defined them as a subcategory of hosting service providers. AVMSD avoided to touch upon this issue: it defined video-sharing platforms by circumscribing the type of service they provided, without reflecting on their status under the ECD. The baseline category that AVMSD used has been "a service as defined by Articles 56 and 57 of the Treaty on the Functioning of the European Union" rather than an "information society service".[876] It merely acknowledged the ECD rules by adding that the measures to be applied by video-sharing platforms shall not lead to any ex-ante control measures or upload filtering which would be contrary to Article 15 of the ECD.[877] The second meaningful difference is the choice of legal instrument: a regulation. A directive would have been incapable of achieving the same level of harmonisation that was aimed at by the DSA.[878]

The country-of-origin principle, which serves as the jurisdictional model of the AVMSD, the ECD and the GDPR, has similarly been adopted in

876 Article 1. point 1.b(aa): AVMSD: "... the principal purpose of the service or of a dissociable section thereof or an essential functionality of the service is devoted to providing programmes, user-generated videos, or both, to the general public, for which the video-sharing platform provider does not have editorial responsibility, in order to inform, entertain or educate, by means of electronic communications networks within the meaning of point (a) of Article 2 of Directive 2002/21/EC and the organisation of which is determined by the video-sharing platform provider, including by automatic means or algorithms in particular by displaying, tagging and sequencing."

877 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce').

878 Ivana Kostovska and Sally Broughton Micova, "Platforms Within the AVMSD Regulatory Architecture: a VSPs Governance Model" *TPRC49: The 49th Research Conference on Communication, Information and Internet Policy* 2, 2021. https://ssrn.com/abstract=3898142 or http://dx.doi.org/10.2139/ssrn.3898142.

the DSA. According to this principle, Member States have jurisdiction over providers based on the location of their main establishment.[879] This principle is held as indispensable to protect the European single market, however, its practical implications cause considerable difficulties.[880] As Rozgonyi suggests, national authorities (especially the most effected Irish Broadcasting Authority) ought to adapt to the challenges with voluntary coordination and other proactive measures.[881] The suggested "Responsive Governance Model" would rely on the grounds created by Section 4a of the AVMSD that envisages a co-regulatory mechanism. According to Rozgonyi, the code of conduct of the video-sharing platforms should ensure that illegal content is assessed according to the different national legal standards applicable in the Member States. Careful distinction between illegal and harmful content and procedural safeguards were also recommended, which ideas have been adopted in the DSA.

The AVMSD remains intact,[882] and will persist as *lex specialis* alongside the DSA and the ECD which serve as a *lex generalis.* Providers of online platforms such as YouTube will be covered by all three regulations. It is worth noting that the AVMSD does not distinguish between very large video-sharing platforms and smaller ones. However, the envisaged code of conducts may make such a distinction, if deemed necessary. Further questions may emerge related to the classification of a platform as a video-sharing platform. A platform may qualify as a video-sharing platform even if videos are "an essential functionality of the service". For example, on Facebook, while videos are neither the principal purpose, nor a dissociable section of the service, they could still be regarded as an essential functionality.

The regulatory object of AVMSD has also been addressed by the European Media Freedom Act (EMFA, discussed above in Chapter 5). Key differences to AVMSD are that EMFA would extend its scope to all media

879 Article 56 DSA.

880 https://blogs.lse.ac.uk/medialse/2016/07/04/caution-loose-cornerstone-the-country-of-origin-principle-under-pressure/.

881 Krisztina Rozgonyi, "Negotiating new audiovisual rules for Video Sharing Platforms: proposals for a Responsive Governance Model of speech online," Revista Catalana de Dret Públic, 61, (2020): 83–98. https://doi.org/10.2436/rcdp.i61.2020.3537.

882 Recital (9) DSA. (This Regulation should complement, yet not affect the application of rules resulting from other acts of Union law regulating certain aspects of the provision of intermediary services, in particular [...] Directive 2010/13/EU of the European Parliament and of the Council as amended).

outlets including print and online; the topics it embraces extend to media independence, platform-media relationship, state advertising and media concentration.

7 Regulation of the Digital Market

7.1 The regulatory philosophy of the Digital Markets Act

The organic development in the market of platform services resulted in monopolistic markets, dominated by a few companies whose services are only partly comparable. The sheer size of the ominous actors alone predisposes them to possess a significant market power. Competition law appeared insufficient to deal with this new complex situation.[883] First, ex-post regulation was seen as less efficient to achieve the desired result in the business environment. Second, the tools of competition law did not appear to offer an optimal solution in most cases. Especially in the case of global companies, even defining market position with classic competition law would pose a disproportionate challenge and demand extensive market research resources. Further, the "zero-price-problem"[884] prevents classic competition law from adequately addressing the unfairness of a service that is seemingly offered for free. Additionally, there are situations when the gatekeeper's dominant position cannot be realistically, or should not be, altered in the near future.[885] Breaking up large platform companies would not solve the real problems of value extraction and remain insufficient for addressing the systemic problems.[886] Moreover, inter-platform competition cannot be regarded unambiguously as a goal, because the value of the service – pri-

883 Heiko Richter, et al., "To Break Up or Regulate Big Tech? Avenues to Constrain Private Power in the DSA/DMA Package," *Max Planck Institute for Innovation & Competition Research Paper No. (2021): 21–25*
last modified Nov 4, 2021.

884 Manuel Wörsdörfer, "Big Tech and Antitrust: An Ordoliberal Analysis," *Philosophy and Technology* 35, no. 65 (2022). https://doi.org/10.1007/s13347-022-00556-w.

885 Thomas Stuart, "Too little too late? An exploration and analysis of the inadequacies of antitrust law when regulating GAFAM data-driven mergers and the potential legal remedies available in the age of Big Data," *European Competition Journal*, 17, no. 2 (2021): 407–436.

886 Mariana Mazzucato, Josh Entsminger, and Rainer Kattel, "Reshaping platform-driven digital markets" in: *Regulating big tech: Policy responses to digital dominance*, ed. Martin Moore and Damian Tambini (New York, NY: Oxford University Press, 2021): 17–34.

marily for end users – grows with the level of market share, although, this feature can be best balanced with interoperability.[887]

Characteristics like the network effect, the economies of scale, and the benefits from collected data have limited the contestability of the market ecosystems.[888] The lack of contestability for a certain service can enable a gatekeeper to engage in unfair practices, which further reduce the chances for other actors to contest the dominant position.[889] As in the case of abuse of market dominance, contestability and fairness are intertwined. Contestability, while related to market concentration, is not identical with it.[890] According to the contestable market theory, even monopolistic actors "behave better" in markets where entry barriers are low, because they are aware that their market position may be contested.[891] Whereas even in an oligopolistic market, if contestability is low, the players more easily engage in unfair practices. Contestability of a market assesses an undertakings' ability to overcome entry barriers, to expand and to challenge the gatekeeper. By shifting the focus from the characteristics of the big companies to those of the market, the regulator can avoid a potentially fruitless debate on defining the dominant position (whether the affected companies have any market competitors, what are comparable services, and so forth).

The idea of contestability means that smaller providers receive favourable treatment under DMA: they are not required to open up their services, but can demand the big ones to do so. At the same time, DMA may inadvertently and indirectly privilege gatekeepers, by striking out the effect

887 As Cornils explains, large providers are better placed to moderate content, and form an easier contact point for authorities to maintain cooperation. Matthias Cornils, "Designing platform governance: A normative perspective on needs, strategies, and tools to regulate intermediaries" *Algorithm Watch*, May 26, 2020 https://algorithmwatch.org/en/wp-content/uploads/2020/05/Governing-Platforms-legal-study-Cornils-May-2020-AlgorithmWatch.pdf.

888 Nicolas Petit, "The proposed digital markets act (DMA): a legal and policy review,"*Journal of European Competition Law & Practice,* 12, no. 7 (2021): 529–541.

889 Laux, Wachter, and Mittelstadt, "Taming the few,".

890 Recital 32–34.

891 Stephen Martin, "The theory of contestable markets," *Bulletin of Economic Research*, 37, no. 1 (2021): 65–68. See also: Jason Gordon, "Contestable Market Theory – Explained," *The Business Professor,* last modified Sept 10, 2023. https://thebusinessprofessor.com/en_US/business-management-amp-operations-strategy-entrepreneurship-amp-innovation/contestable-market-theory-definition.

of national laws that will not be applicable to them anymore, while they still may be applicable to smaller companies.[892]

Giant platforms have obviously transformed and dominated several spheres of economy and society. A number of industries, including trade, travel and tourism, online services, and the public information sector, have become heavily reliant on a small number of large platform companies. These tech companies have a significant impact on social and political processes, extending their influence far beyond the economic sphere. Their power has influenced not only the commercial market, but also the "opinion market". As commonly observed by academics, civil society and public policy experts, their entrepreneurial power has compromised democratic functioning with its own logic.[893] The "systemic opinion power" that social media companies possess, has grown comparable to political power: it is capable of creating dependencies and of influencing other players in a democracy.[894]

7.2 Comparing social and business platforms

DMA addresses platform companies in their role as intermediaries between business users and end users, whereas DSA addresses intermediaries between content providers and content receivers. The DSA calls both users "recipients of the service", where the service appears to be the same to each party: posting, sharing, and extending already existing content (with sharing, liking, writing opinions, comments, etc.)[895] DMA is slightly more specific as it defines the difference between end users and business users. The term "business user" refers to any person acting in a commercial or

892 Jörg Hoffmann, Liza Herrmann, and Lukas Kestler, "Gatekeeper's Potential Privilege – the Need to Limit DMA Centralisation," *Max Planck Institute for Innovation & Competition Research Paper No. 23–01.,* Forthcoming in: *Journal of Antitrust Enforcement*, last modified Dec. 22, 2022. Available at SSRN: https://ssrn.com/abstract=4316836 or http://dx.doi.org/10.2139/ssrn.4316836.

893 Joseph Vogl, "Capital and ressentiment: The totalizing power of social fragmentation," *Finance and Society*, 7, no. 2 (2021): 140–45.

894 Natali Helberger, "The political power of platforms: How current attempts to regulate misinformation amplify opinion power," *Digital Journalism*, 8, no. 6 (2020): 842–854.

895 The definition "recipient of the service" disguises that DSA openly favours information intermediary services "in particular for the purposes of seeking information or making it accessible". Article 3 (b) DSA.

professional capacity using core platform services for the purpose of or in the course of providing goods or services to end users. On the other hand "end user" denotes any other user.[896] The interests and risks associated with these two sides of services differ. Business users in DMA could be regarded as "content providers" under the DSA. For commercial trade service providers, the business users are traders, while for social media service providers they are advertisers on the one hand, and professional content providers on the other; acting as suppliers. Whereas "demand" is represented by buyers on platforms like Amazon (corresponding to end-users), on Facebook, the customers are the advertisers: they are the ones who pay.

For traders, the risk of being unable to reach their customers in any channel other than through the gatekeeper, is higher than for advertisers of not reaching their audience. This dependence has been researched in the context of Amazon,[897] but has been contested in the case of social media service providers such as Facebook or Twitter. This supports the observation that platforms are hesitant to abuse "all their power" vis-á-vis users, in order to keep users satisfied and engaged with their services, but that this incentive depends entirely on the relationship of supply and demand.[898] Abuse of power can only be perceived by the weaker side: in the case of Amazon, these are the traders, while buyers' wishes are fulfilled to the extreme.[899] The rationality of this practice lies in the different angles of the market: buyers' have ample other opportunities for online and offline shopping, and Amazon's competitive advantage lies in its broad product range, pricing and favourable terms. In contrast, on Facebook, end users lack practical alternatives to enjoy the "Facebook experience" of exchanging information with friends and represent their virtual social identity.[900]

---

896 Articles 2 (21–22) DMA.

897 European Commission DG Competition (2019) Amazon Marketplace, case AT.40462. See also:, Mariateresa Maggiolino and Federico Cesare Guido Ghezzi, "The Italian Amazon Case and the Notion of Abuse," (November 29, 2022). *Bocconi Legal Studies Research Paper* No. 4288948 (2022), last modified Nov 29, 2022. SSRN: https://ssrn.com/abstract=4288948 or http://dx.doi.org/10.2139/ssrn.4288948.

898 Eifert, supra note, at 992.

899 For example, Amazon Prime's "try before you buy" option allows free delivery and free return before paying the price of the goods. In case of dispute regarding wrong delivery or wrong return, Amazon always favours the buyer.

900 Gunn Sara Enli and Nancy Thumim, "Socializing and Self-Representation online: Exploring Facebook," *Observatorio (OBS*)*, 6, no. 1 (2012). See further: Aparajita Bhandari and Sara Bimo, "TikTok and the "algorithmized self": A new model of

Even if there are other social networks, their user base is different, the network connections would have to be built up from scratch, and the self-representation utilities may be qualitatively different.[901] These, i.e. online self-representation and a social network that is regarded as an asset, as well as receiving some information about the others' networks, are considered definitional elements of a social networking site according to one of the earliest definitions provided by danah m. boyd and Nicole B. Ellison.[902]

The lack of clarity of regarding the definition of substitutable services has been one of the reasons why traditional competition law was not considered an adequate tool to address the power of platforms.[903] As one certainty, it can be said that online social networking services do not have a brick-and-mortar substitute.[904] Whether Facebook and Twitter are substitutable services, has been sometimes debated and investigated, without reaching conclusion.[905]

What appears to be a more relevant – and still incomplete – exercise, is redefining the actual "service". Assuming that online platforms are really merchants of *data*, competition law should assess their market dominance on the market of data – a field that remained unexplored.[906]

In this regard, what gatekeepers keep under control are not so much the services provided to persons, but the data flow: they aggregate, process, re-distribute and ultimately monetise data in unprecedented scale. The mechanism of data extraction bears an eerie resemblance to the energy-mines

online interaction," *AoIR Selected Papers of Internet Research*, 2020 https://doi.org/10.5210/spir.v2020i0.11172.

901 Compare: Jill Walker Rettberg, "Self-Representation in Social Media," *SAGE handbook of social media*, (2017): 429–443.

902 danah m. boyd and Nicol B. Ellison, "Social Network Sites: Definition, History and Scholarship," *Journal of Computer-Mediated Communication*, 13, no. 1 (2007): 210–230.

903 Geoffrey Parker, Georgios Petropoulos, and Marshall W. Van Alstyne, "Digital Platforms and Antitrust," in *2021 Winner of Antitrust Writing Award*, view at https://awards.concurrences.com/en/awards/2021/academic-articles/digital-platforms-and-antitrust, SSRN: https://ssrn.com/abstract=3608397 or http://dx.doi.org/10.2139/ssrn.3608397 last modified May 22, 2020.

904 Spencer Weber Waller, "Antitrust and social networking," *North Carolina Law Review*, 90, (2011): 1771. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1948690.

905 Waller, ibid.

906 Inge Graef, "Market definition and market power in data: The case of online platforms," *World Competition: Law and Economics Review*, 38, no. 4 (2015): 473–506. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2657732.

depicted in the popular movie "Matrix", in which humans are provided with an engaging virtual mental experience to endure that their biological energy is harvested, while being digitally connected in human "farms".[907] In this context, DMA addresses directly this core service of data trade, whereas DSA tries to call for moderating and refining the virtual experience that is provided in exchange of the data.

Originally, DSA and DMA were planned to be incorporated into one and the same legal rule, and they got separated during the drafting process.[908] The objective of the DMA is to ensure a level playing field and increase contestability of the market: to prohibit certain actions that would increase entry barriers, and prescribe others that would lower these barriers.[909]

7.3 Quo Vadis, Platforms? #2

A gatekeeper is defined as an undertaking that fulfils certain qualitative and quantitative criteria. The three qualitative criteria are that the company a) has a significant impact on the internal market, b) provides a core platform service which is an important gateway for business users to reach end users, and c) enjoys an entrenched and durable position, or foreseeably it will do so in the near future. These criteria are then quantified with clear numbers to avoid unambiguity.

Core platform services may include a long list of services such as: online intermediation services, online search engines, online social networking sites, video-sharing platform services, number-independent interpersonal communications (messenger) services, operating systems; web browsers; virtual assistants; cloud computing services; and finally, online advertising services, which include any additional intermediation of advertising services, even ancillary, if performed by an actor which provides another core service.[910]

Online social networking services and video-sharing platform services are explicitly mentioned and defined; the latter with reference to the Au-

907 The Matrix, 1999. Written and directed by Lana Wachowski and Lilly Wachowski. https://www.imdb.com/title/tt0133093/.

908 Laux, Wachter, and Mittelstadt, "Taming the few,".

909 Jan Christopher Kalbhenn, (2021) "European Legislative Initiative for Very Large Communication Platforms," in *Perspectives on Platform Regulation,* ed. Judit Bayer et al. (Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG, 2021): 47–76.

910 Article 2 (2) DMA; all listed categories are defined in (3–10).

diovisual Media Services Directive. Nonetheless, online social networking services receive in DMA a better definition than in DSA: platforms (!) that enable end users to connect and communicate with each other, share content and discover other users and content across multiple devices and, in particular, via chats, posts, videos and recommendations.[911] Ironically, "platform" as such is not defined by DMA, although the word is ubiquitously used in the term "core platform services". If we wanted to follow the logic of definitions, we get to a circular reference. However, "online platform" is defined with a broad approach in DSA, with Recitals narrowing it on two main threads: "social networks" and "online platforms allowing consumers to conclude distance contracts with traders".[912] The latter is most likely identical with "online intermediation services" in DMA, which is defined with reference to the P2B Directive's definition as an information society service, that allows business users, on the basis of contractual relationship, to offer goods or services to consumers, with a view to facilitating the initiating of direct transactions between those business users and consumers.[913] It is regrettable that these definitions could not have been better aligned in the acts intended to complement each other.[914]

A "significant impact on the internal market" is presumed if the company, in each of three consecutive financial years, achieved an annual turnover within the EU of at least EUR 7,5 billion, or if the company's average market capitalisation or its equivalent fair market value amounted to at least EUR 75 billion in the last financial year, provided that it provided the same core platform service in at least three Member States.[915] In addition, the service is presumed to be an important gateway for business users to reach end users if it had on average at least 45 million monthly active end users established or located in the Union in the last financial year, and at least 10 000 yearly active business users established or located in the Union. Methodology and indicators for calculation are set out in the Annex of

911 Article 2 (7) DMA.
912 Recitals (1) and (13) DSA.
913 Article 2 (2) Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services.
914 Konstantina Bania, (2023) "Fitting the Digital Markets Act in the existing legal framework: the myth of the "without prejudice" clause," *European Competition Journal*, 19, no. 1 (2023): 116–149. https://doi.org/10.1080/17441056.2022.2156730.
915 Article 3. 2(a), DMA.

the Act.[916] This condition is similar to that of DSA's definition of VLOPs, however, DMA's definition is narrower, resulting in the addition of another tier to the pyramid depicted in Chapter 6.

Furthermore, an additional aspect is introduced regarding the requirement for significant impact. A platform's position is considered entrenched and durable only if it has maintained a certain number of users consecutively in each of the last three financial years.[917] If undertakings fulfil the criteria, they are obliged to inform the Commission about their circumstances within two months. A list of the gatekeepers and their core platform services (in regard of which they must obey obligations under the act) is published by the Commission, and reviewed every three years.[918] On September 6 2023, the European Commission published the list of the gatekeepers: Alphabet (owner of Google), Amazon, Apple, ByteDance (owner of TikTok), Meta (owner of Facebook), and Microsoft, and in 2024 it added iPadOS by Apple, and a new actor: Booking.com.[919] The undertakings had six months to comply with the requirements. Companies which meet the legal threshold, are obliged to notify the Commission within two months thereafter with the relevant information per each of the core platform service that they provide.

The assumption of being significant on the market can be also rebutted. If a company fulfils the quantitative criteria by reaching the numeric thresholds of gatekeeping, but is of the opinion that it would not satisfy the qualitative criteria, because of the conditions of the market on which it operates, it may submit its arguments to the Commission. As Microsoft argued "45 million [monthly active users] may be significant for an operating system but insignificant for a social network or online search engine." Accordingly, the Commission was urged to create a "Safe Harbour" for companies that are not an important gateway for business users to reach end users, even if they fulfil the quantitative criteria.[920]

916 Article 3, DMA.
917 Article 3 (2)b. DMA.
918 Article 4. DMA.
919 Up-to-date list of gatekeepers and their services: https://digital-markets-act.ec.europa.eu/gatekeepers_en. Remarks by Commissioner Breton: Here are the first 7 potential “Gatekeepers” under the EU Digital Markets Act, https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_23_3674.
920 Feedback from Microsoft Corporation, May 2021. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12418-Digital-Services-Act-package-ex-ante

7.4 Obligations of gatekeepers

The qualification of a gatekeeper attaches to a service and not to an actor. Various services of the same actor may qualify differently, depending on the service's market dominance. Thus, a company can be gatekeeper in respect of one service, and not in another. Even a dominant actor may have ancillary services or services less market share in more competitive sectors, in which regard the same company would not owe these obligations. Therefore, the designating decision should name those specific core platform services in which regard the undertaking is regarded as a gatekeeper.[921] The designating decision should name those specific core platform services in which regard the undertaking is regarded as a gatekeeper.[922]

7.4.1 A less unfair use of data

The first step for a more level market is reducing the data dominance of gatekeepers and thereby increasing fairness of competition.[923] DMA prohibits gatekeepers to use sideway opportunities to get personal data and to combine them. For example, Meta would not be allowed to combine personal data acquired from Facebook with other data from Instagram or Facebook Marketplace without the consent of the user. However, all rules are lifted if the user consents to the use of her personal data, which means that the prohibitions achieve hardly more than a reinforcement of the GDPR. Hence, labelling these as regulations concerning the "fair use" of data would be inaccurate; rather, they mandate a "less unfair" usage. However, a significant interpretative provision could bring about the most significant change: in cases where consent has been declined or revoked by the end user, a gatekeeper is prohibited from repeating the consent request for the same purpose for a period of one year. This measure would offer relief to numerous users who regularly navigate online and repeatedly encounter consent requests from the same websites. In conjunction with

-regulatory-instrument-of-very-large-online-platforms-acting-as-gatekeepers/F2256709_en.

921 Article 3. point 9. DMA.

922 Article 3. point 9. DMA.

923 Article 5 (1–2) DMA.

the DSA's rule mandating equal representation of rejection and consent,[924] this measure can provide a better protection of personal data. A more effective approach would be to mandate that cookie consent can be managed through browser settings, as envisioned in the proposal for an e-Privacy Regulation.[925] However, the progress of this proposal remains uncertain, prolonging its implementation into the uncertain future.[926]

Still, combining the consent principle with the aim to increase fairness of market behaviour raises important objections. First, the consent principle in the general context of the GDPR has been viewed rather critically because of the consenting fatigue that frustrated the ultimate purpose of the principle. Extensive literature has discussed how the consenting practice did not correspond to the theoretical expectations. Often, the consent given was uninformed, or "consent fatigue" undermined conscious decision-making.[927] Dark patterns are widely employed to elicit consent without genuine will.[928]

Second, if the aim of market regulation would be to prevent gatekeepers from combining various data sets, then users' consent should not be relevant in that regard. One users' consent will have external impacts on other users (data externalities) and the cumulative effect of user consents

924 Article 24 (1a) DSA. "refusing consent shall be no more difficult or time-consuming to the recipient than giving consent. In the event that recipients refuse to consent, or have withdrawn consent, recipients shall be given other fair and reasonable options to access the online platform."

925 Proposal for a regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications), COM/2017/010 final — 2017/03 (COD). For a quick read, visit: https://digital-strategy.ec.europa.eu/en/policies/eprivacy-regulation.

926 Luca Bertuzzi, „What the EU has in store for 2023," https://iapp.org/news/a/what-the-eu-has-in-store-for-2023/.

927 Bart Willem Schermer, Bart Custers, and Simone van der Hof, "The crisis of consent: How stronger legal protection may lead to weaker consent in data protection," *Ethics and Information Technology*, 16, no. 2 (2014): 171–182. See also: Christine Utz et al., "(Un)informed Consent: Studying GDPR Consent Notices in the Field," in 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19), November 11–15, 2019, (London, UK – New York, NY: ACM, 2019). https://doi.org/10.1145/3319535.3354212 See further: Harshvardhan J. Pandit, "Proposals for Resolving Consenting Issues with Signals and User-side Dialogues," *arXiv preprint arXiv:2208.05786*. (2022).

928 Midas Nouwens, "Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence," in *Proceedings of the 2020 CHI conference on human factors in computing systems* Apr (2020): 1–13.

will compromise the aim of market regulation.[929] Viewed in connection with the practical issues around consenting, such as gatekeepers' persuasion techniques to acquire user consent,[930] leads to a disturbing conclusion. The legislative aim of increasing fairness of competition should not be made dependent on actions of individual users towards gatekeepers.[931]

This problematic correlation between data protection law and competition law has not been satisfactorily clarified. The issue was also discussed in relation to the German Facebook Case.[932] Facebook put pressure on users to consent to the merging of personal data that Facebook collected inside and outside of its platform. The German Federal Cartel Office argued that this was an exploitative, abusive behaviour by a dominant firm, an abuse of the dominant position. However, the decision induced an intense discussion about the boundaries and the relationship between competition law and data protection law. In the landmark judgement concerning Meta vs. Bundeskartellamt, the ECJ found that in the context of examining an abuse of dominant position, it may be necessary for the national competition authority to also assess whether the practices of that undertaking comply with rules other than competition law, such as the GDPR. The authority may legitimately determine that an undertaking's general terms of use and their implementation are inconsistent with the GDPR if that finding is necessary to establish the existence of such an abuse of a dominant position.[933]

This court decision also signals a more restrictive data protection approach. Importantly, the court also found that in the absence of the data subject's consent, the personalised advertising by which Facebook finances its activity, cannot justify the processing of the data at issue as a 'legitimate

929 Rainer Mühlhoff und Hannah Ruschemeier, "Predictive Analytics und DSGVO: Ethische und rechtliche Implikationen," 12, (2022): 15–26.

930 Rupprecht Podszun, "Should Gatekeepers Be Allowed to Combine Data? – Ideas for Art. 5(a) of the Draft Digital Markets Act," (June 4, 2021). Available at SSRN: https://ssrn.com/abstract=3860030 or http://dx.doi.org/10.2139/ssrn.3860030.

931 Inge Graef, "Why End-User Consent Cannot Keep Markets Contestable: A suggestion for strengthening the limits on personal data combination in the proposed Digital Markets Act," *VerfBlog,* 2021/9/02, https://verfassungsblog.de/power-dsa-dma-08/.

932 Wolfgang Kerber and Karsten K. Zolna, "The German Facebook case: the law and economics of the relationship between competition and data protection law," European Journal Law and Economics 54, (2022): 217–250. https://doi.org/10.1007/s10657-022-09727-8.

933 C-252/21 *Meta v Bundeskartellamt*, GC, 4 July 2023.

interest' pursued by Meta.[934] The Court referred merely to the case when consent was lacking, however, would the right to consent provide a sufficient protection in the case of a dominant gatekeeper? Or is this a case of informational asymmetry, which GDPR is unable to mitigate? In a similar case, EDPB ruled against Meta (Instagram), stating that data processing for personalised advertising is not essential to the contract between Meta and users of Instagram. Therefore, it was inappropriate for Meta to refer to Article 6(1)(b) of GDPR, and likewise for the Irish Data Protection Authority to accept this argument. EDPB instructed the latter to alter its decision. The EDPB decision frames the issue of power asymmetries in the context of the fairness principle, calling it as being able "to cancel out the negative effects of such asymmetries".[935] GDPR is notoriously incapable of addressing the unique characteristics of the data-driven business model.[936]

As a next step in the policy debate, this objection needs to be taken seriously, and face the problem that the consent principle appears to be an inadequate basis to protect against the privacy-encroaching business model. At least, when the relationship between the data subject and the controller is unequal, the idea of consent is fundamentally flawed. The absence of rigorous and robust public enforcement, both prior and post, and a general tendency toward disproportionate accountability for individuals, allow industry players to largely disregard court rulings or interpret and apply them in extremely restrictive ways.[937] In fact, the issue of informational asymmetry has been considered by a preliminary draft version of GDPR only to be dismissed at a later phase of the legislative debate, leaving behind a watered down reminiscence, which allows Member States to provide for more specific rules to protect privacy in an employment relationship.[938] (Employment relationship was mentioned as a key example of asymmetrical relationships, but the original text was not limited to these.)

In these cases, the use of personal data should rather be regulated without regard to consent, due to the informational asymmetry between

934 Article 6(1)(f) GDPR.
935 EDPB Binding Decision 4/2022, at 225, 226.
936 Peter van de Waerdt, "Information asymmetries: recognizing the limits of the GDPR on the data-driven market," *Computer Law & Security Review*, 38, no. 105436 (2020).
937 Jef Ausloos, Jill Toh, and Alexandra Giannopoulou, "How the GDPR can exacerbate power asymmetries and collective data harms," Nov 29, 2022. https://www.adalovelaceinstitute.org/blog/gdpr-power-asymmetries-collective-data-harms/.
938 Article 88 GDPR.

the data controller and the data subject.[939] The appropriate policy would step beyond protecting the individual "self" and confront economic imperatives.[940] What may at first seem as a striking restriction of users' liberty, in fact would finally take down a set of illusions.[941] First, the illusion of voluntary consent, second, the ignorance that this practice has a constraining effect on other users' liberties, who did not consent.[942] The only choice is one between market dictate or governmental restrictions. The problem appears to be rather that neither the industry actors, nor the European policy-makers are willing to forgo the lucrative promises offered by the data economy.[943] Even though the price is ultimately paid by consumers, partly with their data, and partly by purchasing advertised products, which effectively makes them pay for using platforms "for free".

As seen, DMA makes considerable efforts to put the private data sovereignty between boundaries. Still, it merely carved out and protected a narrow slice from the ocean of personal data, as criticized by prominent members of the academia and NGOs.[944] Even after some changes to the text, the prohibition of processing personal data acquired from different services is limited to the purpose of online advertising services. Thus, data acquired through third parties that use the gatekeepers' core platform services may be further used for other purposes apart from advertising. The changes are nuanced and do not alter the logic of the surveillance economy.

939 Ausloos, supra note.

940 Julie E. Cohen, "What Privacy is For," *Harvard Law Review* 126, no. 7 (2013): 1904–1933.

941 Because of the consenting fatigue and uninformed consent, see supra notes above as well as the difficulties of enforcement.

942 Other users, who did not consent, may be more transparent for profiling in the societal context. See Ruschemeier, supra note.

943 Michaela Padden and Andreas Öjehag-Pettersson, "Protected how? Problem representations of risk in the General Data Protection Regulation (GDPR)," *Critical Policy Studies* 15, no. 4 (2021): 486–503, DOI: 10.1080/19460171.2021.1927776.

944 Irish Council of Civil Liberties (2022) Digital Markets Act Article 5(1)a. Open Letter. 19 April 2022. https://www.article19.org/wp-content/uploads/2022/04/ICCL-to-DMA-co-legislators-19-April-2022.pdf.

7.4.2 Prohibition of paralysing contractual conditions

A second set of rules aim to unbundle users from contractual constraints with the platform. In other terms, they prohibit certain contractual conditions that would limit the liberty of users.[945] Both business users and end users should be allowed to conclude contracts directly with each other through the gatekeepers' platform, or through another channel, even at different prices than used there. This would particularly be relevant for Amazon which defines cruel terms for its sellers.[946] The same rule may have an impact by excluding any exclusive sale agreement, or predatory price-setting also for media intermediaries that fall under the gatekeeper definition, such as YouTube or Facebook and could prevent that they exercise a disproportionately formative impact on the information offer, through forcing their conditions on content providers (whether professional or user-generated).

All users must be allowed – under DMA – to communicate directly with each other or with third intermediary service providers, on their own terms, independently from their contract with the gatekeeper. Moreover, the gatekeeper should not require them to use any specific services, such as an identification service, a web browser engine, a payment service, or similar; they should not be pressed to subscribe to or to register with any further core platform service of that gatekeeper. Further, while internal complaint-handling mechanisms are possible and even welcome, gatekeepers are not allowed to restrict raising any issue of non-compliance with the relevant authorities.[947] Even before DMA explicitly prohibited these practices, they may very well have violated the traditional principle of prohibition of abuse of dominant position in business relations, as they limit contractual freedom of the less dominant party.

945 Article 5 (3–8) DMA.

946 Sheryar Tahirkheli, "Paying for the picnic was not enough; get ready to share your lunch too. They are selling but also learning your product: Predatory marketing and imitation strategies in the Amazon marketplace," *Social Sciences & Humanities Open* 5, no 1 (2022). See also: Eifert, Metzger, Schweitzer, Wagner. Supra Note. at 992.

947 Article 5 (3–8) DMA.

7.4.3 Advertising transparency

A third set of rules – still in the same Article, among the main obligations of gatekeepers – applies to transparency of online advertising. Transparency is not among the key tools applied by DMA (as opposed to the DSA), and these rules supplement those in the P2B Regulation. They serve the complex goal of protecting the level playing field, beyond protecting just advertisers and publishers. Advertisers are not among the most endangered population in the current platform environment: they are not the ones with less power, in comparison to end users. Still, platforms enjoy a better position to enforce their interests, and this justifies that advertisers also benefit from safeguards against abuse of dominance. Gatekeepers are obliged to reveal the pricing system that they use in intermediating advertisements between advertisers and publishers. They must provide detailed information to both the publisher and the advertiser on the price and the fees paid, including deductions and surcharges, for each of the relevant online advertising services provided by the gatekeeper. This must take place on a daily basis, and, of course, free of charge.[948] Moreover, advertisers and publishers, as well as third parties authorized by these, should be granted free access to data – both aggregated and non-aggregated – for the ads they run, as well as to the performance measuring tools of the gatekeeper and the necessary data for independent verification of the ad inventory.[949]

This set of rules intends to make an effect on the ad-driven information economy and governance. At the very least, it is expected that they create the option for advertisers and publishers to choose alternative providers of online advertising services, and thereby contribute to the reduction of advertising costs, which is expected to be reflected in the costs of other products and services.[950]

To summarise, Article 5 tackles tricky acquisition of personal data and opaqueness of the concluding data-driven advertising within the same section, completing them with prohibitions to impose contracted slavery on business partners. This section alone, if properly employed, is likely to achieve a meaningful change in the data-driven network capitalism that is fuelled by behavioural-advertising. The cornerstone of the entire intervention would be, nevertheless, putting limits on the data acquisition

948 Article 5 (9–10) DMA.
949 Article 6 (8) DMA.
950 Recital 45 DMA, Eifert et al. supra note, p. 1015.

of companies, because the data that they dispose of furnishes them with the competitive strength in the digital economy.[951] As previously stated, that part of the Regulation hardly extends beyond interpretative rules of the current GDPR.

7.5 Prohibition of self-preferencing

Article 6 restricts gatekeepers' liberties to technically exploit their intermediating position. The affected actions arose literally from their position as "doorkeepers" which allows them to control access, similarly to customs collectors. Gatekeepers are in a position that they can disproportionately grow their data assets through their core platform services of intermediating between end users and business users. Much like customs collectors regulate the flow of goods across borders, gatekeepers control the flow of information and data, and are able to leverage this position to acquire unfair advantages.

7.5.1 No expropriation of personal data

Data that has been collected or generated by business users of a gatekeeper, are easily accessed by that gatekeeper due to their intermediating role. This is prohibited by DMA, unless those data are publicly available. Similar protection applies to data that can be inferred from the commercial activities of business users or their end users, including click, search, view, and voice data. This becomes particularly important in two perspectives: the source and the destination of that data. As regards the *sources* of data, in the era of connected devices, even incidental data collection is going to produce stellar quantities of data, without conscious awareness of the end-user.[952] In the platform environment, the boundaries between private and public get blurred. Acquiring information about individual users is

951 Stuart, "Too little," 407–436.

952 Jacob Kröger, "Unexpected inferences from sensor data: a hidden privacy threat in the internet of things," in *Internet of Things. Information Processing in an Increasingly Connected World: First IFIP International Cross-Domain Conference, IFIPIoT 2018, Held at the 24th IFIP World Computer Congress, WCC 2018, Poznan, Poland, September 18–19, 2018, Revised Selected Papers 1* (Springer International Publishing, 2019): 147–159.

possible around the clock: from the type of apps they use during their morning coffee, until their late-night entertainment. An unrestricted data gathering can combine individual preferences about commercial choices, education, transport, banking, physical exercise and many more aspects of life. An unhindered analysing and using of this data would threaten autonomous decision-making in all aspects. The second aspect of why this is of particular importance, is the *destination* of the data: their use for influencing the political and public discourse. Personalised ranking of search results in business platforms may influence merely commercial decisions, such as which pair of shoes to buy – the likelihood to impact social processes is slant. But personalised ranking, or especially, microtargeting information on public matters may influence political decisions, i.e. voting behaviour. The causal link between ranking and its effect on behaviour has not yet been academically examined. There are several studies examining the link between Russian propaganda during the US pre-election time 2015–2016, focusing on exposure to propaganda and its relationship to election *results*.[953] Some studies have denied such a relationship, albeit their results are restricted by the minor data sets that they had access to.[954] This echoes a crucial element of the limitations in research: while the activity of disinformation networks were extensively scrutinised,[955] the behaviour of *platforms* in using user data for content ranking – apart from sponsored content – has never been examined, and it would also not have been possible, for lacking insight into platform's ranking algorithms and their deployment of user data. Besides, platforms did not show willingness to co-

953 Christopher A. Bail et al., "Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017," *PNAS* 117, no. 1 (2020): 243–250. Josephine Lukito, "Coordinating a multi-platform disinformation campaign: internet research agency activity on three U.S. social media platforms, 2015 to 2017," *Political Communication* 37, no. 4 (2020): 238–255. Kathleen Hall Jamieson, *Cyberwar: How Russian Hackers and Trolls Helped Elect a President: What We Don't, Can't, and Do Know* (Oxford: Oxford University Press, 2018). Yevgeniy Golovchenko et al., "Cross-Platform State Propaganda: Russian Trolls on Twitter and YouTube during the 2016 U.S. Presidential Election," *The International Journal Press/Politics* 25, no. 3 (2020): 357–389.

954 Gregory Eady et al., "Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior," *Nature Communications*, 14, no. 1 (2023): 62.

955 Bail, "Assessing the Russian Internet," 243–250.

operate with researchers in this respect.[956] We know only that they indeed rely on inferred data, and even claimed proprietary rights for those,[957] even though the EDPS has called for delegitimising of inferred data already in 2014.[958]

While the effects of political micro-targeting are also debated, prominent authors suggesting that its impact is rather nuanced than robust,[959] those studies merely focus on the rather sporadic examples of micro-targeted political ads. Whereas *systematic* ranking choices of a gatekeeper social media platform work at a different scale than individual ads. Content ranking is as non-transparent and overwhelming as swimming in a flooded river. For this reason, the prohibition of using inferred data, and further restrictions on data use were crucial basic requirements to balance informational asymmetries between users and gatekeeping platforms.

7.5.2 Equal treatment

DMA's objective of protecting privacy, and fair treatment vis-à-vis individual users is only secondary to balancing the market situation between gatekeepers and other service providers. At the other side of the seesaw, the business users – clients – of gatekeepers are entitled to get free access to data, including personal data, which has been provided for or generated in the context of the business users' services, or provided by the end users using those services. The gatekeeper may use such data only where they originate directly from the end user, who uses third-party services through

956 Sinan Aral and Dean Eckles, "Protecting elections from social media manipulation," *Science* 365, no. 6456 (2019): 858–861. DOI:10.1126/science.aaw8243.

957 Sandra Wachter and Brent Mittelstadt, "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI," *Columbia Business Law Review* 2019. DOI: 10.31228/osf.io/mu2kf.

958 Opinion of the European Data Protection Supervisor on the Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions – "A comprehensive approach on personal data protection in the European Union". https://edps.europa.eu/sites/edp/files/publication/11-01-14_personal_data_protection_en.pdf.

959 Jörg Matthes et al., "Understanding the democratic role of perceived online political micro-targeting: longitudinal effects on trust in democracy and political interest," *Journal of Information Technology & Politics*, 19, no. 4 (2022): 435–448.

the relevant core platform service, and only if the end user opts in to share such data.[960]

To generally keep gatekeepers' self-preferencing at bay, end users should be allowed to easily remove or uninstall applications on their operating system (those that can be removed without prejudice to the whole operation of the system), and to install third-party applications or even app stores on their system. For example, Apple has to allow its users to uninstall even pre-installed apps on its devices. They should not hinder users switching between different software applications and services.[961] End users should also be allowed to easily change default settings on their operating system, virtual assistant and web browser that steer end users to products or services provided by the same gatekeeper. They have to be allowed to choose from alternative service providers at the first time when they use the service (online search engine, virtual assistant or web browser).[962]

Vertically integrated service providers, which offer their own services through their own core platform services, have a conflict of interest with presenting other content with similar prominence. To balance this, DMA provides that ranking, indexing, and crawling must not give preference to the gatekeepers' services and products,[963] or those of other providers that they effectively control.[964] The same obligation applies to app stores as well as products or services that are displayed with prominence in the newsfeed of an online social networking service. Crawling and indexing are the preparatory steps to ranking, for example search hits.[965] Even before the display of the results, crawling and indexing might prefer own content or content of related providers, by finding and enlisting those content to keep them ready for display in case of a user query, therefore these actions (crawling and indexing) are also outruled, whether through legal, commercial or technical means.[966] It is yet unclear whether this rule would also prevent providers from pre-installing their own apps on devices, which is not explicitly prohibited, merely that they should be removable; or whether it would be illegal for Google to feature Google Maps or Google Shopping

960 Article 6 (1) and (10) DMA.
961 Article 6 (3–4, 6) DMA.
962 Article 6 (3) DMA.
963 Article 6 (5) DMA.
964 Recital 51 DMA.
965 Case AT.39740 Google Search (Shopping). https://ec.europa.eu/competition/antitrust/cases/dec_docs/39740/39740_14996_3.pdf.
966 Recital 52 DMA.

among its top search results.[967] Obligations under Article 6 will be further specified by the Commission, therefore the related concerns – including the questions of costs – may be clarified in an implementing act.[968]

7.5.3 Non-discrimination and fairness in ranking?

The requirement to apply fair and non-discriminatory ranking of services and products is merely sideways mentioned alongside the prohibition of giving preferred treatment of gatekeepers' own services or products.[969] The two together outline the essential obligation to provide intermediary services in a neutral manner. This principle should, in fact, serve as a fundamental principle of all online platform services, including online intermediation. It has its roots in the principle of net neutrality which requires access providers to provide to all online services equal quality of service.[970] It also reflects the original principal condition set by ECD that intermediary service providers have no control over the content.[971] Nonetheless, this principle itself was included merely in the Recitals, and according to literal reading and interpretation, it should have applied only to mere conduits and caching providers, however, the Delfi decision applied it for Delfi as a hosting provider.[972] This condition was omitted from DSA as the complexity of the regulated actions rendered it impractical to define this expectation as a condition for immunity. DSA does not require non-discrimination and fairness in the ranking, merely that users

967 Isabella Dickinson, *The DMA's not-so-final view on self-preferencing-Open Up Your Algorithm-Or Else* (2022) Frontier Economics. United Kingdom. Retrieved from https://policycommons.net/artifacts/3350311/the-dmas-not-so-final-view-on-self-preferencing/4149182. CID: 20.500.12592/qd5j08.

968 Article 8 (2) DMA.

969 Article 6 (5) second sentence, DMA.

970 Bernd Holznagel, « Internetdienstefreiheit und Netzneutralität," *AfP* 1, (2011): 532–539.

971 Recital 43 E-Commerce Directive.

972 Aleksandra Kuczerawy and Pieter-Jan Ombelet, "Not so different after all? Reconciling Delfi vs. Estonia with EU rules on intermediary liability. LSE," July 1, 2015. https://blogs.lse.ac.uk/medialse/2015/07/01/not-so-different-after-all-reconciling-delfi-vs-estonia-with-eu-rules-on-intermediary-liability/ See judgement: Delfi AS v. Estonia, application no. 64569/09, 16.06.2015. Grand Chamber.

are informed of the ranking criteria, and that VLOPs offer at least one optional ranking method that is not based on profiling.[973]

In the context of social networking services and other information intermediaries, including video-sharing platform services, this rule – if properly enforced – could potentially establish the foundation for neutrality of content distribution. The principle resembles to the German provision outlined in the renewed Interstate Media Treaty which orders platforms (media intermediaries) to refrain from discriminating between journalistic offerings.[974] The objective of that provision is to prevent certain offerings from being over- or under-represented or having their findability impaired.[975] If consistently applied for social networking sites, the requirement of non-discriminatory and fair ranking would certainly be meaningful for ensuring a level playing field in the marketplace of ideas. This function has been outlined by Holznagel in his interpretation of net neutrality as a constitutional obligation of providers.[976]

However, it is still to be settled, which interpretation of non-discrimination is meant by this rule? Is it based solely upon the protected characteristics of the Charter,[977] or does it encompass others as well? AI opens up new questions of discrimination.[978] In particular, new categories can be created based upon seemingly innocuous characteristics, which currently escape protection from discrimination.[979] From the perspective of the pub-

973 Article 27, Recital 70 DSA.

974 § 94 MStV. See also: p. 271–272.

975 Bernd Holznagel and Jan Christofer Kalbhenn, "Media law regulation of social networks-country report: Germany," in *Perspectives on Platform Regulation* (Baden-Baden: Nomos, 2021): 261–290. See also: Sarah Hartmann and Bernd Holznagel, "Reforming Competition and Media Law: The German Approach," in *Regulating Big Tech: Policy Responses to Digital Dominance* ed. Martin Moore and Damian Tambini (New York, NY: Oxford Academic, 2021) https://doi.org/10.1093/oso/9780197616093.003.0003, accessed 12 July 2023.

976 Bernd Holznagel, *Netzneutralität als Aufgabe der Vielfaltssicherung* (K&R, 2010): 95.

977 Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited. Article 21 Charter of the Fundamental Rights.

978 Sandra Wachter, Brent Mittelstadt, and Chris Russell, "Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI," *Computer Law & Security Review*, 41, no. 105567 (2021).

979 Janneke Gerards and Frederik Zuiderveen Borgesius, "Protected grounds and the system of non-discrimination law in the context of algorithmic decision-making and artificial intelligence," *Colorado Technology Law Journal* 20, no. 1 (2022).

lic discourse, any categorisation can lead to unfair inequality, because any categorical exclusion of content limits users' fundamental informational rights: to exercise autonomy over their information consumption and their right to free expression. Even speakers' right to define or restrict their audience is limited by discriminative content offer of platforms.[980]

A crucial question of interpretation is whether "products and services" can also be interpreted as meaning "carrying content" by social networking sites and video-sharing platforms? With ranking, crawling, indexing, the gatekeeper influences the potential reach of products and services without the intention of the service provider and without the intention (or consent) of the receiver of information. As an intermediary, such interference is arbitrary. In the absence of this DMA rule, financial interests could have justified this interference in the case of commercial products that do not affect fundamental rights. However, in the case of content disseminated by social media speakers, discriminative ranking raises complicated constitutional questions, which are apparently evaded by both DSA and DMA.[981]

As always, the devil is in the details: enforcement of this requirement may be more difficult than it sounds. Transparency of the named activities is another requirement, which is supposed to enable the Commission and independent researchers to acquire information about the programming of the crawling, indexing and ranking algorithms. Still, their oversight would require a dissection of the search algorithm and close examination, which is costly and onerous.[982] From the Google (Shopping) case it appears that smaller providers cannot even afford to spend on their own crawling and indexing, therefore they give up on them, allowing considerable advantage to the gatekeepers.[983]

7.6 The pros and cons of interoperability

Similarly as it is possible to initiate a call from one telecommunications provider to another's network, or as it is possible to send an email from

980 Although see other considerations of targeting in the next chapter and in Judit Bayer, "Double harm to voters: data-driven micro-targeting and democratic public discourse," *Internet Policy Review*, 9, no. 1, (2020): 1–17.

981 On balancing rights between platforms and users please read Part I. of this book or Judit Bayer, "Rights and Duties of Online Platforms" supra note 377.

982 Dickinson, "The DMA's not-so-final," 5.

983 Dickinson citing Google Shopping Case at 6.6.2. clause (304).

a Gmail account to a Yahoo account, gatekeepers must allow interoperability between theirs and other providers' services and hardware of the same type.[984] The same obligation extends to services that are not core platform services, but are provided together with such, or in support of such services.[985] For example, Apple must allow other apps to access the voice assistant Siri, enabling users to open third-party apps directly through Siri. This could potentially compromise the level of data protection and security; would it not be added by DMA that their level must be retained (see below).

Importantly, the obligation of interoperability does not extend to all types of core services. In regard to social networking providers, no agreement could be reached during the legislative process. This will remain a discussion point for the future.[986] However, messaging services – officially named number-independent interpersonal communication services, such as Facebook Messenger, Viber, WhatsApp, or Telegram – must open up and allow interoperability. The requirement is reported to cause challenges to end-to-end encryption.[987] DMA prescribes that the level of security and privacy must be preserved,[988] but human rights advocacy organisations like EFF have complained of the unclarity of this requirement.[989] Another concern is that smaller companies will be unable to reciprocate the same

984 Ian Brown, *Interoperability as a tool for competition regulation* (OpenForum Academy, 2020), https://euagenda.eu/upload/publications/ian-brown-interoperability-for-competition-regulation.pdf.

985 Article 6 (7) DMA.

986 Deal on Digital Markets Act: EU rules to ensure fair competition and more choice for users. IMCO Press Release, 24. 03. 2022. https://www.europarl.europa.eu/news/en/press-room/20220315IPR25504/deal-on-digital-markets-act-ensuring-fair-competition-and-more-choice-for-users.

987 Lukas Wiewiorra et al., *Interoperabilitätsvorschriften für digitale Dienste: Bedeutung für Wettbewerb, Innovation und digitale Souveränität insbesondere für Plattform-und Kommunikationsdienste* (Bad Honnef: WIK-Consult GmbH, 2022). See also: Corin Faife, "Security experts say new EU rules will damage WhatsApp encryption" 2022: 5, 18. Abgerufen von: https://www.theverge.com/2022/3/28/23000148/eu-dmadamage-whatsapp-encryption-privacy.

988 Recital (64) DMA.

989 Mitch Stoltz, Andrew Crocker, and Christoph Schmon, "The EU Digital Markets Act's Interoperability Rule Addresses An Important Need, But Raises Difficult Security Problems for Encrypted Messaging," *Electronic Frontier Foundation* May 2, 2022. https://www.eff.org/de/deeplinks/2022/04/eu-digital-markets-acts-interoperability-rule-addresses-important-need-raises. See in connection with Article 7 (3) DMA.

level of security, and this may prevent that they take advantage of the gatekeepers' obligation.[990]

The clear objective of interoperability is to reduce entry barriers by counteracting the network effect that emerges from the economy of scale. In non-interoperable markets, a larger user base confers a competitive advantage. However, interoperability is not unilaterally viewed with enthusiasm by independent experts of competition law. Bourreau and Krämer point out that it may actually have an opposite effect. For example, it lowers the incentives for consumers to multihome, which is otherwise a powerful driver for contestability.[991] Bailey and Misra found that it may lead to the standardisation or homogenisation of products, and hinder investment into developing new technologies.[992] Moreover, they argue that interoperability can benefit the market merely in the short term, because the main factor of competition is not the price but the improvement of the services, which is facilitated by the existence of dynamic competition. These concerns have contributed to dismissing the concept of interoperability for online social networking sites and other core platform services, and maintaining it only in relation to operating systems, virtual assistants and messaging services.[993]

Data portability is the little sister of the interoperability obligation. It means that the end-user is not only allowed to switch services, but does not have to leave behind and resign to its accumulated data and start to develop her new "home" in the other platform, but is able to migrate with all her data. This should be ensured free of charge. The GDPR has already regulated data portability, (having its origin in the Data Protection Directive) but left several questions unanswered.[994] As an additional practical rule compared to the already existing principle, gatekeepers are also obliged to provide free tools to facilitate the effective realisation of data portability, including real-

990 Mikolaj Barczentewicz, "Minimizing Privacy Risks in Regulating Digital Platforms: Interoperability in the EU DMA," *CPI Antitrust Chronicle,* July 2022.

991 Marc Bourreau and Jan Krämer, "Interoperability in Digital Markets: Boon or Bane for Market Contestability?" July 25, 2022. SSRN: https://ssrn.com/abstract=4172255 or http://dx.doi.org/10.2139/ssrn.4172255.

992 Rishab Bailey and Prakhar Misra, "Interoperability of Social Media: An appraisal of the regulatory and technical ecosystem," February 12, 2022. SSRN: https://ssrn.com/abstract=4095312 or http://dx.doi.org/10.2139/ssrn.4095312.

993 Article 6 (7) and Article (7) DMA.

994 Paul De Hert et al., "The right to data portability in the GDPR: Towards user-centric interoperability of digital services," *Computer Law & Security Review* 34, no. 2 (2018): 193–20. https://doi.org/10.1016/j.clsr.2017.10.003.

time continuous access to the data.[995] Without interoperability, portability alone does not necessarily allow a smooth migration. The data are likely to be in a specific format, follow the specific (proprietary) logic of the source platform and often cannot be integrated into the destination platform.

Dominant platforms like Google, Facebook, Microsoft, Twitter, and Apple, dedicated specific platforms for data portability, such as the Data Transfer Project, to facilitate bidirectional data transfer between participating platforms. Open protocols and service gateways have similarly been created to enable continuous data transfer.[996]

7.7 Enforcement of DMA

The European Commission will take the lead in the DMA's enforcement. The centralised enforcement mechanism makes gatekeepers directly accountable to the Commission. They must inform the Commission about steps that lead to more market concentration. They are obligated to have an audit and submit its result to both the EDPB and the Commission. In case of infringement, they face fines that can extend up to 10 % of a gatekeeper's total worldwide turnover in the preceding financial year, or up to 20 % of the same for repeated infringements of the core obligations listed in Articles 5–6–7. For minor transgressions a mere 1 % is foreseen (see in more detail below), whereas for ongoing infringements a periodic penalty payment of maximum 5 % of the average daily worldwide turnover in the preceding financial year can be levied.[997] Secession with the view to avoid the obligations is viewed as circumvention and will not prevent the Commission from designating the provider as a gatekeeper.[998] However, a gatekeeper also has the opportunity to demonstrate that one or more specific obligations would endanger its economic viability, because of exceptional circumstances beyond its control. Upon the gatekeeper's reasoned request, the Commission may suspend one or more of the obligations temporarily, among others, also on the grounds of public health or public security.[999]

995 Article 6 (9) DMA.
996 For a comprehensive picture on data portability: Johann Kranz, et al., "Data Portability," *Business & Information Systems Engineering*, 2023: 1–11.
997 Articles 11, 12, 14, 15, 16, 30, 31 DMA.
998 Article 13 DMA.
999 Article 8–9–10. DMA.

Besides the Commission's role, national authorities may also be involved. During the legislative procedure, it became gradually clear that DMA monitoring and enforcement will require enormous human resources capacity. The expectation grew from the initial 80 FTE gradually up to 150 FTE, especially with regard to gatekeepers' "deep pockets".[1000] Together with the activism of some national competition authorities,[1001] this led to the proposal that these are more involved in the enforcement mechanism.[1002] This was supported by the recognition that they may be better placed to receive complaints from competitors. National authorities shall work in close cooperation and coordinate their enforcement actions with each other, and with the European Competition Network.[1003]

DMA needs to be interpreted without prejudice to other, pre-existing laws that apply to the digital environment, primarily the GDPR, and existing competition law.[1004] Critiques fear an overlapping in competences, and express concern as to how those will be streamlined,[1005] or whether the European competition law framework would become more fragmented because of the overlaps.[1006] DMA prevails both as *lex specialis,* and as an EU regulation over national laws.[1007] A High-Level Group of Regulators is created to help streamlining DMA with other laws in the digital sector, consisting of regulators in the digital sectors, the representative of the Commission, of national competition authorities and of other authorities such as data protection, consumer protection and telecommunication law.[1008] The necessity to cooperate has been declared by the European Court of Justice

1000 Belle Beems, "The DMA in the broader regulatory landscape of the EU: an institutional perspective," *European Competition Journal* 19, no. 1 (2023): 1–29., citing Martijn Snoep, chairman of the Dutch NCA.

1001 Bundeskartellamt 15 February 2019 B6–22/16 Facebook (Case Summary); ACM 24 August 2021 Summary of decision of ACM in ACM/19/035630 Apple.

1002 "Private enforcement of the Digital Markets Act: Germany as a frontrunner?" Norton Rose Fulbright, March 2023. https://www.nortonrosefulbright.com/en/knowledge/publications/41cb9705/private-enforcement-of-the-digital-markets-act-germany-as-a-frontrunner.

1003 Articles 37, 38 DMA.

1004 Konstantina Bania, "Fitting the Digital Markets Act in the existing legal framework: the myth of the "without prejudice" clause," *European Competition Journal* 19, no. 1 (2023): 116–149. DOI:10.1080/17441056.2022.2156730.

1005 Beems, "The DMA," 1–29.

1006 Giuseppe Colangelo, "The European Digital Markets Act and Antitrust Enforcement: A Liaison Dangereuse," *European law review* 5, (2022): 597–621.

1007 Bania, "Fitting the Digital," 116–149.

1008 Article 40. DMA See also: Beems, "The DMA," 1–29.

in its recent decision against Meta initiated by the German Competition Authority.[1009]

7.8 The impact of DMA – summary

DMA is not meant to be a traditional competition law instrument,[1010] but one that complements traditional competition tools in digital markets.[1011] The goal of having a contestable platform market is one of the novel approaches to achieve effective market pluralism.[1012] Rather than focussing on the size of the companies, it keeps their behaviour in sight from the perspective of consumers and of business partners. The approach, especially in its details of interoperability, sharing access, and the principles of fairness and non-discrimination, have their roots in the concept of net neutrality that had dominated the policy discourse in the decade prior to 2015, when platform dominance took over its place in the centre of the focus.[1013]

The impact of DMA on the democratic public discourse is indirect, but potentially profound. As an instrument that aims at creating a level playing field on the digital market, it regulates the basic pillars of the digital environment, which serve as the background setting of information distribution. Concrete benefits for the media landscape which is dominated by gatekeeping platforms are expected in two areas: first, non-discrimination in organising content, due to the ranking, indexing, crawling regulation that requires fairness and transparency. Second, a better protection of citizens from targeted content without their explicit consent, as a result of the privacy protection obligations. This is also anticipated to afford more autonomy in seeking information and forming opinions. However, the induced structural changes may be meaningful beyond these concrete benefits. On a contestable market,

1009 C-252/21 *Meta v Bundeskartellamt*, GC, 4 July 2023.

1010 Speech Margrethe Vestager, "Competition in a Digital Age," *European Internet Forum* 17 March 2021.

1011 Cani Fernández, "A New Kid on the Block: How Will Competition Law Get along with the DMA?," *Journal of European Competition Law & Practice* 12 no. 4 (2021): 271–272. See also: Nicolas Petit, "The proposed digital markets act (DMA): a legal and policy review," *Journal of European Competition Law & Practice* 12, no. 7 (2021): 529–541.

1012 Fernández, "A New Kid," 271–272.

1013 Christopher T. Marsden and Ian Brown, "App stores, antitrust and their links to net neutrality: a review of the European policy and academic debate leading to the EU Digital Markets Act," *Internet Policy Review* 12, no. 1 (2023).

pluralism has better chances and enables better terms for content distribution to all business users, resulting greater diversity of content. Greater content diversity and transparency may enhance citizens' opportunities to find access to a wider range of content and impart information with a fair chance of being heard rather than suppressed for revenue optimisation. Interoperability and access obligations may provide more breathing space for both end-users and business users. A flatter market structure is expected to incentivise platforms to pay more attention to human rights and better serve their customers also as citizens.[1014]

However, the instrument does not ambition to change the data-driven advertising-based business model of platforms. The interventions are likely to bring some changes in how personal data is mined and exploited for the purpose of advertising, but whether these will benefit smaller companies and individual users, remains a question for the future. While the law clearly is set to inspire advertisers to become more independent of the big platforms, the giant actors are in a better position to leverage their options, for example by collecting less data overall, and getting easy on the obligation of furnishing advertisers with the data so valuable for them.[1015] On the other hand, advertisement and sponsorship can be equally effective without enhanced targeting precision, which is, by some authors, contested as a myth. Research on the effectiveness of targeted advertising is controversial to the least.[1016] The value that advertisers pay for, is rather the *hope* that their ads will reach a click and bring new customers. Instead of the mere personal data, it is really the *metrics* behind the data mass that is valued on

1014 Jacqueline Rowe, "How will the Digital Markets Act affect human rights? Four likely impacts," *Global Partners Digital* Jul 5, 2022. https://www.gp-digital.org/how-will-the-digital-markets-act-affect-human-rights-four-likely-impacts/.

1015 Sean Czarnecki and Patrick Coffee, "Advertisers could be harmed by the EU's sweeping new law aimed at tech giants like Google and Meta," *Business Insider* Apr 7, 2022. https://www.businessinsider.com/a-new-eu-law-targeting-big-tech-could-boost-advertisers-2022-3.

1016 Brahim Zarouali, et al., "Using a personality-profiling algorithm to investigate political microtargeting: assessing the persuasion effects of personality-tailored ads on social media," *Communication Research* 49, no. 8 (2022): 1066–1091. See also: Harsh Taneja, "The myth of targeting small, but loyal niche audiences: Double-jeopardy effects in digital-media consumption," *Journal of Advertising Research* 60, no. 3 (2020): 239–250.

the market.[1017] This begs the question whether the model would also work without personal data, based purely on statistical, or pseudonymised data? Raising the costs of collecting and using personal data, while at the same time making its use less exceptional on the market, as DMA does, may push innovation into that direction.

1017 Kean Birch, DT Cochrane and Callum Ward, "Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech," *Big Data & Society* 8, no. 1 (2021) 20539517211017308.

8 Political advertising

The proliferation of political disinformation tactics has emerged as a central issue for free democracies. Political communication enjoys the widest possible scope of freedom of expression. However, the need for a prompt resolution to counter disruptive political campaigns that appeared to undermine the democratic process, and to combat hybrid warfare techniques has grown pressing. In 2021, the European Commission has introduced a Proposal for a Regulation on the transparency and targeting of political advertising[1018] (hereafter: RPA) as a component of its comprehensive approach to readjust the digital landscape for public discourse. The purpose of this initiative was to develop a specialized framework that will operate in combination with the Digital Services Act and the Strengthened Code of Practice, and bring transparency and accountability in the realm of political advertising. It aims to create a delicate equilibrium between ensuring freedom of expression and protecting the public discourse as a tool of defensive democracy.

8.1 Emerging frontiers in political advertising

The ability of the digital domain to precisely target particular demographic groups, combined with its instantaneous and extensive reach, has granted political messages and disinformation campaigns an unparalleled level of effectiveness. According to reports, targeted advertising has demonstrated a level of effectiveness that is ten times greater than that of a conventional Facebook ad.[1019] Furthermore, there is potential for future advancement

1018 This chapter has been closed based on the political compromise on RPA, before the finalising of the text. The numbering of the paragraphs may therefore deviate from what has been indicated in the footnotes.

1019 Filipe N. Ribeiro et al., "On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook," *ACM Digital Library* (2019): 140–149. https://doi.org/10.1145/3287560.3287580.

in targeted advertising through the enhancement of targeting accuracy.[1020] The digital landscape provides ample personal data about users' activities and reactions. The metaverse and networked devices are likely to further increase the data pool and its quality, for example by including body language, and present increased prospects for persuasion.[1021] In the absence of explicit user consent, it was possible to infer data and subsequently classify it as proprietary information belonging to the data processors. These mechanisms allow for predictions about individuals, including information that the individuals themselves may not be conscious of. The predictions can be extended to other individuals whose profile is similar to the person whose data has been inferred.[1022] This practice contravenes the right to privacy, even if not expressly prohibited by the letter of the law, yet circumventing the objectives of personal data protection.[1023]

The widespread occurrence of opaque political advertising on digital platforms had created an environment that is prone to foreign manipulation, misinformation, and disinformation campaigns, which are fed by the extensive reserves of Big Data. This phenomenon had a notable prominence during earlier electoral procedures, when first foreign, and later domestic entities endeavoured to manipulate public emotion and erode the credibility of democratic results. Significant controversies have emerged from these practices in relation to the 2016 United States elections, the Brexit campaign, and following national elections, particularly in nations such as India and Brazil. This particular campaigning method has sparked considerable scholarly investigation into its political and regulatory impli-

1020 Till Blesik et al., "Applying big data analytics to psychometric micro-targeting," in *Machine Learning for Big Data Analysis*, ed. Siddhartha Bhattacharyya, Hrishikesh Bhaumik, Anirban Mukherjee and Sourav De (Berlin: De Gruyter, 2018).

1021 Rumen Pozharliev, Matteo De Angelis, Dario Rossi, "The effect of augmented reality versus traditional advertising: a comparison between neurophysiological and self-reported measures," *Marketing Letters* 33, (2022): 113–128. https://doi.org/10.1007/s11002-021-09573-9.

1022 Rainer Mühlhoff and Hannah Ruschemeier, "Predictive analytics und DSGVO: Ethische und rechtliche implikationen." *Telemedicus–Recht der Informationsgesellschaft: Tagungsband zur Sommerkonferenz*, (2022) 38–67. Rainer Mühlhoff, "Predictive privacy: towards an applied ethics of data analytics." *Ethics and Information Technology* 23.4 (2021): 675–690.

1023 Martin Ebers, *Beeinflussung und Manipulation von Kunden durch Behavioral Microtargeting* (SSRN, 2018): 423.

cations.[1024] The convergence of unregulated digital advertising, Big Data analytics, and foreign involvement raised an urgency to take regulatory measures that increase transparency and protect the democratic integrity, first of all in European Parliamentary elections.

8.2 The pitfalls of personalized targeting

Political micro-targeting utilizes sophisticated psychological and technological methodologies inherited from the commercial advertising industry to collect user preference data and generate customized messaging.[1025] Its origins can be traced back to traditional practices such as local gatherings and campaign materials, however, the use of modern technologies and the availability of extensive data have brought about a novel aspect.[1026] Its aim may be directly or indirectly linked to political processes, encompassing voter persuasion, influence on election participation, and solicitation of

1024 Frederik J. Zuiderveen Borgesius et al., "Online Political Microtargeting: Promises and Threats for Democracy," *Utrecht Law Review* 14, no. 1 (2018): 82–96. DOI: https://doi.org/10.18352/ulr.420 See also: Philip N. Howard, Barath Ganesh, and Dimitra Liotsiou, "The IRA, Social Media and Political Polarization in the United States, 2012–2018," *Working Paper No. 2018. 2*. Oxford: Project on Computational Propaganda, Oxford University. https://comprop.oii.ox.ac.uk/research/ira-political-polarization/ See also Tom Dobber et al., "Two crates of beer and 40 pizzas: the adoption of innovative political behavioural targeting techniques," *Internet Policy Review* 6, no. 4 (2017); and Rafael Evangelista and Fernanda Bruno, "WhatsApp and political instability in Brazil: targeted messages and political radicalisation," *Internet Policy Review* 8, no. 4 (2019).

1025 Orestis Papakyriakopoulos et al., "Social media and microtargeting: Political data processing and the consequences for Germany," *Big Data and Society* 5, no. 2 (November 20, 2018). See also: Jeff Chester and Kathrin C. Montgomery, "The role of digital marketing in political campaigns," *Internet Policy Review* December 31, 2017 and. Jens Koe Madsen and Toby Pilditch, "A method for evaluating cognitively informed micro-targeted campaign strategies: An agent-based model proof of principle," *PLoS ONE* 13, no. 4 (2018) https://doi.org/10.1371/journal.pone.0193909.

1026 Philip N. Howard, *New Media Campaigns and the Managed Citizen* (Cambridge University Press, 2006). See also: Viktor. Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt, 2013). Also: Jessica Baldwin-Philippi, "Data campaigning: between empirics and assumptions," *Internet Policy Review* 8, no. 4 (2019) DOI: 10.14763/2019.4.1437. https://policyreview.info/articles/analysis/data-campaigning-between-empirics-and-assumptions.

donations.[1027] Research indicates that political advertising is unable to significantly modify individuals' core political predispositions.[1028] However, data-driven campaigns have proven successful in influencing voter turnout, donation rates, and the reinforcement of pre-existing preferences, as well as engage undecided swing voters, who possess the potential to significantly influence the outcome of closely contested elections.[1029]

Privacy concerns are a fundamental component of discussions around the practice of political micro-targeting. However, political micro-targeting has the potential to negatively impact voters in two distinct ways, doubling the harm. Every act of targeting inherently involves non-targeting, or exclusion. Therefore, beyond its potential to violate the rights of individuals who are specifically targeted, more importantly, it also undermines the right to access information for others who are not targeted and hence uninformed of the political messages received by their fellow citizens. Moreover, these individuals who lack information also lack the meta-information on their peers' access to the specific message. This scenario can be distinguished from situations in which readers merely skim headlines without engaging in further exploration. In that context, individuals acknowledge the existence of information that is available to the general public, thereby maintaining a cognitive understanding of its incorporation within the realm of public communication. Those individuals can review the topic at a later time or seek input and perspectives from their peers. However, those individuals who are missed from the targeting pool are hindered in their access to this meta-information. Therefore, the practice of micro-targeting leads to the fragmentation of public discourse as it selectively withholds information from individuals who are not part of the targeted audience, while conveying specific messages through advertisements geared at a different set of individuals. The resulting fragmentation may have a direct impact on the fundaments of the democratic process, which relies on a cohesive public discourse. It can give rise to a distortion within the realm of public debate, leading to divisions and disruptions within the democratic process.

1027 Balázs Bodó, Natali Helberger, and Claes de Vreese, "Political micro-targeting: A Manchurian candidate or just a dark horse?" *Internet Policy Review* 6, no. 4 (2017).

1028 Bodó, Helberger, and de Vreese, "Political micro-targeting".

1029 For examples, a few swing states in the US regularly attract increased attention of targeted advertisements. See NBC's simulator "Swing the election" where users can test how different turnout of different voter groups would turn the results. https://www.nbcnews.com/specials/swing-the-election/

The appeal for political actors resides in establishing barriers that restrict the dissemination of their message exclusively to the target audience.[1030]

Considering that it can result in a potential exclusion of significant portions of societies, this type of targeting practice can be interpreted as a systematic infringement upon human rights. In essence, the point is that an inquiry into targeting should not focus on the reasons for selecting specific groups, but rather on the grounds for excluding particular segments of the population from participating in that specific part of the political discourse. This deliberate audience restriction through micro-targeting of political messages may pose a greater challenge to address, than privacy concerns.

One potential approach to addressing this violation of informational rights could involve providing citizens with the opportunity of an "opt-in" choice, whereby they actively seek and collect targeted adverts from internet archives. Nevertheless, the effectiveness of this method is largely dependent on an individual voter's inclinations and traits. As a result, it has the potential to disproportionately harm individuals who are already disadvantaged and more susceptible to information manipulation.

In conclusion, the justification for limiting micro-targeting should not just be based on its capacity to spread manipulative content or violate private rights, but primarily on its threat to the democratic debate process. Despite the low likelihood of manipulation, the potential harm is high, therefore the overall risk assessment is also high.

8.3 The protection of political speech

The concept of freedom of expression holds significant importance within democratic systems, serving as a fundamental pillar and playing a crucial role in safeguarding a range of political rights. This right, particularly within the realm of political debate, is afforded the utmost degree of protection and is subject to rigorous examination. The European Court of Human Rights (ECtHR) has frequently underscored the narrow scope for limitations on political discourse and discussions pertaining to matters of public concern.[1031] In this particular domain, the margin of appreciation

1030 Howard, 2006, supra note p. 136.

1031 Lingens v. Austria, no. 9815/82, Judgement 08/07/1986, para. 42, Castells v. Spain 11798/85, Judgement 23.4.1992, para. 43, and Thorgeir Thorgeirson v. Iceland, 13778/88, 14/03/1992, para. 63.

afforded to member states of the Council of Europe is significantly limited. Moreover, the European Convention on Human Rights not only protects the substance of information but also the methods by which it is disseminated.[1032] While commercial advertising has less protection than other speech,[1033] political advertising is granted substantial legal safeguards, albeit not without any limits.[1034]

The purpose of regulations pertaining to targeting is to safeguard the reciprocal aspect of freedom of expression, namely the right to freedom of information. Consequently, the legal principles and precedents surrounding the latter right are also of significance. The practice of the European Court of Human Rights (ECtHR) demonstrates an increasing acknowledgment of this particular right.[1035] This signifies a departure from the Court's previous position, which initially did not include the right of access to information within the ambit of Article 10 ECHR.[1036] In the case of Leander v. Sweden, the European Court of Human Rights reached the conclusion that Article 10 of the European Convention on Human Rights does not confer a positive right to acquire information.[1037] Nevertheless, a report from the Council of Europe acknowledged the progressive development of freedom of information, specifically highlighting its significance in political and intellectual discourse already in 2005.[1038] The decision in Kenedi

1032 Autronic AG v. Switzerland 12726/87 | Judgement 22/05/1990, Öztürk v. Turkey, 22479/93 | Judgement 28/09/1999, and Ahmet Yildirim v. Turkey, 3111/10, Judgement 18/12/2012.

1033 VgT Verein gegen Tierfabriken v. Switzerland, no. 24699/94, 28 June 2001and TV Vest AS & Rogaland Pensjonistparti v. Norway, 21132/05, December 11, 2008.

1034 Animal Defenders International v. UK, no. 48876/08, 22 April 2013.

1035 Kenedi v. Hungary, no. 31475/05, 26 May 2009, Társaság a Szabadságjogokért v. Hungary, no. 37374/05, 14 April 2009) VgT Verein gegen Tierfabriken v. Switzerland, no. 24699/94, 28 June 2001, and Magyar Helsinki Bizottság v. Hungary, no. 18030/11, November 8, 2016.

1036 Leander v. Sweden, no. 9248/81, 26 March 1987.

1037 Lucy Maxwell, "Access to information in order to speak freely: Is this a right under the European Convention?," *OxHRH Blog* January 19, 2017, https://ohrh.law.ox.ac.uk/access-to-information-in-order-to-speak-freely-is-this-a-right-under-the-european-convention/.

1038 Jean-François Renucci, *Introduction to the European Convention on Human Rights: the rights guaranteed and the protection mechanism*, vol. 1 (Council of Europe, 2005).

and Társaság a Szabadságjogokért v. Hungary in 2009 has reinforced this tendency.[1039]

Two landmark decisions in the field of political speech (Animal Defenders v. UK, 2013) and Erdogan Gökce v. Turkey, 2014) appear particularly relevant for the assessment of political micro-targeting because of a resemblance of the circumstances.[1040] In these cases, the concept of freedom of expression encountered restrictions as a result of governmental involvement with the intention of protecting democratic dialogue and promoting equitable possibilities for every political contender. As a result, the respective states limited access to particular political material in order to safeguard the general public's entitlement to impartial and fair information. In both instances, governmental regulations imposed limitations on the freedom of expression, paradoxically driven by the goal of preserving a robust informational landscape. The use of this method aimed to mitigate the potential distortion of pluralism of viewpoints and the democratic process that could have arisen as a consequence of the disputed speech.

The underlying reasoning in the case of TV Vest established the basis for the Court's position, which was subsequently affirmed in the case of Animal Defenders, even though the latter had a different outcome. The decision in Animal Defenders was generally seen as unexpected,[1041] although the Court applied the same rationales as in TV Vest case, but distinguished the circumstances.

In the TV Vest case, the Court acknowledged the government's assertion that the prohibition was warranted due to the content in question being "likely to diminish the overall quality of political discourse" (at 70). This phenomenon could result in the distortion of the discussion of public mat-

1039 Kenedi v. Hungary (2009), Társaság a Szabadságjogokért v. Hungary (2009), and Magyar Helsinki Bizottság v. Hungary, no. 18030/11, November 8, 2016.

1040 Animal Defenders, supra note, and *Erdoğan Gökçe v. Turkey* – 31736/04 Judgment 14.10.2014 [Section II].

1041 Ronan Ó Fathaigh, "Political Advertising Bans and Freedom of Expression," *Greek Public Law Journal*, 27, (2014): 226–228. Available at: https://ssrn.com/abstract=2505018; James Rowbottom, "Animal Defenders International: Speech, Spending, and a Change of Direction in Strasbourg," *Journal of Media Law* 5, no. 1 (2013a): 1–13. https://doi.org/10.5235/17577632.5.1.1 See also: James Rowbottom, "A surprise ruling? Strasbourg upholds the ban on paid political ads on TV and Radio" *UK Constitutional Law Blog* 22 April 2013 (2013b) (available at http://ukconstitutionallaw.org), Tom Lewis, "Animal Defenders International v United Kingdom: Sensible Dialogue or a Bad Case of Strasbourg Jitters?," *Modern Law Review* 77, no. 3 (2014): 460–474. https://doi.org/10.1111/1468-2230.12074.

ters, granting financially influential parties a higher advantage in communicating their perspectives in comparison to less financially resilient groups. Pluralism and equality were fundamental factors of consideration within this particular environment. However, in that specific case, the Court found that other reasons were more substantial, because the pensioners' party did not pose a threat to the diversity of the political discourse, on the contrary: it was one of those actors who deserved special protection: it "belonged to a category for whose protection the ban was, in principle, intended" (at 73). Whereas, in Animal Defenders, the Court took the position that the ban served the public interest and was sufficiently narrow, because it was limited to specific media only and there were a variety of alternative media available.

By applying the reasoning of the Court to micro-targeted political advertising, analogies can be discerned. For example, it becomes clear that the accessibility of this technology is not uniformly accessible to all political parties or issue groups, irrespective of their financial resources. The disparity in access to resources has the potential to distort public discourse, so posing a risk to the democratic process and impeding the breadth and variety of discussions ("the danger of unequal access based on wealth was considered to go to the heart of the democratic process" (Animal Defenders, para. 117).

8.4 Why was self-regulation insufficient?

Giant platforms have taken significant measures to establish a self-regulatory framework pertaining to political advertising. The implementation of this voluntary framework was initiated with the aim of fostering confidence in their platforms by instituting a level of supervision over political communications. Nevertheless, these self-regulatory initiatives were unable to effectively address the issues of transparency and integrity in political advertising. Furthermore, they were not uniformly and consistently enforced throughout the digital domain. As a result, the effectiveness of self-regulation in ensuring openness and accountability in political advertising tactics has remained limited. The findings of a report on the compliance of platforms with the initial 2018 Code of Practice on Disinformation revealed

that discrepancies in compliance resulted in a lack of transparency throughout election campaigns.[1042]

The European Regulator's Group for Audiovisual Media Services (ERGA) conducted an evaluation of Facebook's activities in order to oversee self-regulation. They specifically examined the level of transparency in political advertising, with a special emphasis on the upcoming European Parliament (EP) elections. Nonetheless, ERGA's investigation was constrained to examining the reports provided by Facebook, as the site repeatedly refused to grant access to raw data.[1043] The examination revealed that the current databases are in need of improvement in order to offer the necessary tools and data that are crucial for maintaining the mandated level of transparency.[1044] Academic investigations have also confirmed that Facebook's collection of data has revealed deficiencies in its ability to offer transparent information on their methods for selecting target audiences and the potential for exploiting vulnerabilities.[1045] The platform's efforts in achieving comprehensive transparency criteria for political advertisements were considered insufficient.[1046] Significantly, there is a growing interest among journalists and civil society in accessing the available information, and they were expressing criticism regarding the usability of repositories.[1047]

1042 Niamh Kirk and Lauren Teeling, "A review of political advertising online during the 2019 European Elections and establishing future regulatory requirements in Ireland," *Irish Political Studies* 37, no. 1 (2022): 85–102, doi: 10.1080/07907184.2021.1907888.

1043 Simon Chandler, "Facebook Moves To Block Academic Research Into Micro-Targeting Of Political Ads," *Forbes Magazine*, October 27, 2020. https://www.forbes.com/sites/simonchandler/2020/10/27/facebook-moves-to-block-academic-research-into-micro-targeting-of-political-ads/?sh=51fa31f03905.

1044 ERGA Report, 'Report of the activities carried out to assist the European Commission in the intermediate monitoring of the Code of practice on disinformation (ERGA Report)', 2019, https://erga-online.eu/wp-content/uploads/2019/06/ERGA-2019-06_Report-intermediate-monitoring-Code-of-Practice-on-disinformation.pdf.

1045 Panoptykon Foundation 'Who (really) targets you? Facebook in Polish election campaigns', 2020, available at: https://panoptykon.org/political-ads-report.

1046 Laura Edelson, Tobias Lauinger and Damon McCoy, "A security analysis of the Facebook ad library," *IEEE Symposium on Security and Privacy* (2020): 661–678, available at: https://doi.org/10.1109/SP40000.2020.00084, see also: Leerseen et al., "Platform ad archives: promises and pitfalls," *Internet Policy Review* 8, no. 4 (2018) https://doi.org/10.14763/2019.4.1421.

1047 Paddy Leerssen, Tom Dobber, Natali Helberger and Claes de Vreese, "News from the ad archive: how journalists use the Facebook Ad Library to hold online adver-

The European Union has undertaken to pass and to amend legislation in order to strengthen societal resilience against anti-democratic tendencies. This initiative was expressed in the European Democracy Action Plan in the year 2020.[1048] One of its primary goals was to protect the integrity of elections and promote democratic participation, which involved implementing measures to guarantee transparency in political advertising and communication. The proposed modifications to the Regulation on the statute and funding of European political parties and European political foundations,[1049] in addition to other relevant laws (not discussed in this book),[1050] further augmented the "democracy package."[1051]

8.5 The Regulation on political advertising (RPA)

8.5.1 The concept of RPA

The aim of the RPA[1052] is to prevent any inconsistencies that may hinder the smooth provision of advertising and related services within the internal market. Consequently, the Regulation's intended scope extends beyond European Parliamentary elections to embrace all elections and referenda held

tising accountable," *Information, Communication & Society* 26, no. 7 (2023): 1381–1400. DOI:10.1080/1369118X.2021.2009002.

1048 Brussels, 3.12.2020 COM(2020) 790 final. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the European democracy action plan, available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2020:790:FIN.

1049 The Regulation on the statute and funding of European political parties and European political foundations, available at: https://www.europarl.europa.eu/doceo/document/TA-9-2021-0454_EN.html.

1050 Directive on the right to vote and stand as a candidate in elections to the European Parliament and the Directive on the exercise of the right to vote and to stand as a candidate in municipal elections, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31993L0109 and https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31994L0080.

1051 European Commission Questions & Answers: Reinforcing democracy and integrity of elections, available at: https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_6212.

1052 This part of the manuscript has been closed on 26 February 2024.

inside the European Union and its Member States, including elections for leadership positions within political parties.[1053]

Rather than engaging in content regulation, or imposing limits on advertising *per se*, RPA's objective is to establish principles of transparency in political advertising and a system of accountability. Although there may be conflicts between these requirements and the entrepreneurial freedoms of service providers, the freedom of expression of advertisers, as well as the data protection and privacy rights of sponsors, these limitations are deemed proportionate to the justified aim of minimising the potential manipulation of democratic discourse and guaranteeing the fair, transparent, and varied dissemination of information.[1054] As elucidated in the Explanatory Memorandum,[1055] the objective is to ensure the individual right to receive information in a manner that is balanced, transparent, and pluralistic – a component of freedom of expression that imposes a responsibility on governments to take proactive measures, including protection against private entities.[1056]

The restrictions placed on targeting aim at protecting the privacy rights of persons who are being targeted, as well as the informational rights of those untargeted.[1057] These policies are intended to be part of a comprehensive legislative framework and are expected to have a positive influence on democracy and electoral rights, so illustrating the concept of "self-defensive democracy".[1058]

1053 Directive on the right to vote and stand as a candidate in elections to the European Parliament and the Directive on the exercise of the right to vote and to stand as a candidate in municipal elections, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31993L0109 and https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31994L0080.

1054 RPA, Explanatory Memorandum, 3. Results of [...] impact assessments. Fundamental rights.

1055 "Reducing the possibility of manipulation of the democratic debate and the right to be informed in an objective, transparent and pluralistic way."

1056 Fuentes Bobo v. Spain, 39293/98, February 29, 2000.

1057 Judit Bayer, "Double Harm to voters: data-driven micro-targeting and democratic public discourse," *Internet Policy Review* 9, no. 1 (2020) https://policyreview.info/articles/analysis/double-harm-voters-data-driven-micro-targeting-and-democratic-public-discourse.

1058 Karl Loewenstein, "Militant Democracy and Fundamental Rights, I.," *American Political Science Review* 31, no. 3 (1937): 417–433, 638–658, https://warwick.ac.uk/fac/arts/history/students/modules/hi290/seminars/revolution/lowenstein_militant_democracy_i.pdf.

Similarly to other acts in this regulatory wave, the RPA also relies on the due diligence requirements, and on an anticipated cooperation of the various parties in the advertising value chain.

8.5.2 The scope of the RPA

The definition of political advertising encompasses a range of political communication methods, extending beyond election periods.[1059] The process involves the preparation, placement, promotion, publication, or dissemination of a message, which is determined by two unique and alternative criteria. The first condition entails that the message is generated by a political actor, for them, or on their behalf, except for cases that are solely of a private or commercial nature. The second alternative condition is the likelihood and the intention of influence[1060] on election outcomes, referenda, legislative or regulatory processes, or voting patterns. Accordingly, the message must not only possess the capacity to influence but also be intentionally designed to exert influence. The final text also clarifies what is not political advertising: official information relating to elections, such as the announcement of candidacies, and the free of charge and non-discriminatory presentation of candidates in the media or in specified public places, if explicitly provided by law. The message can be published, delivered or disseminated by any means, which includes traditional and novel communication methods as well.

The definition further includes the element of remuneration, but not as an indispensable condition. Political advertising is "normally provided for remuneration or through in-house activities or as part of a political advertising campaign." This vagueness is meant to open up the scope for different forms of advertising, including sponsored search results, paid targeting, promotion of ranking, or promotion of influencers and potentially further, yet unforeseen forms of advertising.[1061] Recitals also indicate that payment can take the form of benefits in kind or other forms.[1062] Besides

1059 Julian Jaursch, "How new, binding EU transparency standards for political advertising could be even higher," https://www.euractiv.com/section/digital/opinion/how-new-binding-eu-transparency-standards-for-political-advertising-could-be-even-higher/.
1060 In the original: "liable to influence". Article 3 (3) RPA.
1061 Recital 1 RPA.
1062 Recitals 1, 3, 16, 29, 30, 42, etc.

the financial ties, other involvement is also recognised by the definition, such as in-house activities or being part of a campaign.

The original text of the RPA[1063] was even broader and would have covered content by influencers, such as vlogs, when political viewpoints were conveyed, even without substantiating evidence of contractual or monetary considerations, or without the intention (design) to impact. While it might be reasonable to demand transparency even in these cases, it may impose a chilling effect on the political discourse. Among other obligations, platforms would be obliged to categorize such unpaid material as political advertising, attach transparency notifications, and non-compliance could lead to sanctions.[1064]

8.5.2.1 The definition of a political actor

The notion of a "political actor" is unfolded through a wide range of illustrative instances. These include national or European political parties or organizations that are directly or indirectly associated with the activities of such parties; in addition, political campaign organisations that have been established with the sole purpose of influencing the outcome of an election or referendum. It also encompasses political alliances, candidates running for, or holding elected positions at any level of government, including the Union, or for leadership roles within a political party. Further, members of Union institutions, except for the Court of Justice, the European Central Bank and the Court of Auditors or members of Member State governments, including national, regional or local levels also count as political actors; in addition, any person who acts on behalf of, or represents any of the mentioned persons or organisations, and promotes the political objectives of those. Natural or legal persons representing or acting on behalf of any of the aforementioned persons or organizations to advance their political objectives are also considered political actors.[1065]

1063 Max Zeno van Drunen, Natalie Helberger, and Ronan Ó Fathaigh, "The beginning of EU political advertising law: unifying democratic visions through the internal market," *International Journal of Law and Information Technology* 30, no. 2 (2022): 181–199. DOI: 10.1093/ijlit/eaac017.

1064 "The EU Is Going Too Far with Political Advertising." *DSA Observatory* March 16, 2023. Available at: https://dsa-observatory.eu/2023/03/16/the-eu-is-going-too-far-with-political-advertising.

1065 Article 3(4) of the RPA.

The expansiveness of the definition is an attempt to incorporate proxies and influencers. Legislators have had difficulties in delineating the activities of influencers who engage in the direct or indirect promotion of commercial products or political agendas.[1066] Research has pointed out the need to establish a clear definition of political advertising at the EU level, which includes considerations of both the profile of the advertiser and the nature of the material.[1067] The distinction between genuine belief and compensated advocacy typically relies on the occurrence of financial transactions or other forms of compensation in exchange for promotional efforts, which is now explicitly mentioned in the definitions.

8.5.2.2 The second condition: the impact of the message

The incorporation of "a legislative or regulatory process" expands the range of coverage to include issue-driven advertising, thereby facilitating the inclusion of a wide array of content. Academic assessments and policy evaluations conducted by independent specialists also have the capacity to have influence on legislative processes, whilst investigations produced by journalists may, too, have the ability to shape voting behaviour. Imposing breaks and obligations on such publications was not among the objectives of the RPA. Critics argued that the broad definition could potentially lead to excessive monitoring of public communication without well-defined limits. Responding to the criticisms, the scope of the definition has been narrowed during the legislative phase and the risk of overregulation has been somewhat decreased by repeated references to the commercial nature of the message (see discussion under the next subtitle) and by the inclusion of the "design" element which introduces the presupposition of an intention to influence. Still, commercial advertisements that also feature societal concerns such as gender issues, anti-discrimination, climate change or the protection of animals could still be subject to the Regulation, but only if

1066 Catalina Goanta, "Human Ads Beyond Targeted Advertising: Content monetization as the blind spot of the Digital Services Act," *VerfBlog* September 5, 2021, available at: https://verfassungsblog.de/power-dsa-dma-11/ DOI: 10.17176/20210905–213932–0.

1067 Stefan Cipers and Trisha Meyer, "What is political? The uncoordinated efforts of social media platforms on political advertising," (2022) Available at: http://belux-edmo.s3.amazonaws.com/wp-content/uploads/2022/12/20221215_Blogpost-on-Political-Advertisements_FINAL-1.pdf.

they have the ability and are also *designed* to influence regulatory processes or voting behaviour.[1068] Certain commercial advertisements, which call for social responsibility, to protect animals for instance, could fall under the definition.[1069] This insecurity is due to the inherent challenge of providing conclusive evidence whether an advertisement, which by its nature seeks to persuade, can or cannot influence public sentiment or voting behaviour.[1070]

8.5.2.3 The payment element

Any debate on a societal controversy has the potential to impact electoral outcomes, therefore, it was necessary to limit the Act to sponsored content, even if the monetary consideration is not always directly present. Additional Recitals were introduced to emphasise that the RPA applies exclusively to economic actors and economic services, especially that the EU's competence covers only these areas.[1071] A previous Council position asserted that the rules should be uniformly applicable to all political content, regardless of the payment element.[1072] This position has raised serious concerns regarding the public discourse, as expressed in an open letter signed by 33 civil society organizations and directed towards the Council.[1073]

Therefore, the material interest has been made a constant element of the definition of political ads, although the consideration for the service may also be in kind.[1074] This is expressed partly by the statements in the definition (normally provided for remuneration or through in-house activities

1068 "Coca-Cola's 'Equal Love' Ads Spark Anti-Gay Fury in Hungary." *Bloomberg.com*, August 5, 2019. Available at: https://www.bloomberg.com/news/articles/2019-08-05/coca-cola-s-equal-love-ads-spark-anti-gay-fury-in-hungary. In addition, see also the ads in the Animal Defenders v. UK case, ECtHR 48876/08, Judgement of 22.04.2013.

1069 Recital 25 RPA.

1070 DSA Observatory, "EU Going Too Far".

1071 Recital 16 RPA.

1072 Max van Drunen, Natali Helberger, Wolfgang Schulz, and Claes de Vreese, "EU Going Too Far with political advertising!" (2023) https://dsa-observatory.eu/2023/03/16/the-eu-is-going-too-far-with-political-advertising/.

1073 "Public Letter to the Czech Minister of European Affairs and the EU Ministers of European Affairs on the Regulation of Political Advertising." European Partnership for Democracy (EPD). October 30, 2022. Available at: https://epd.eu/2022/10/30/public-letter-to-the-czech-minister-of-european-affairs-and-the-eu-ministers-of-european-affairs-on-the-regulation-of-political-advertising/.

1074 Recital 16 RPA.

or as part of a political advertising campaign), and partly by the negative definitions of some exceptions: the rules should not apply to user content uploaded to a social platform, if disseminated without any consideration by either the user or a third party.[1075] Interestingly, a certain user-generated content could turn into a political advertisement, if sponsored by a third party on behalf of, or for a political actor. Similarly, political opinions expressed in any media under editorial responsibility do not qualify as political ads, unless specific payment or other remuneration is provided for them or in connection with them, by third parties. However, when such political opinions are subsequently promoted, published, or disseminated by service providers, they could be considered as political advertising. The payment element is not explicitly mentioned pertaining to the publication or dissemination of political opinion expressed previously in media, and the type of service providers remains unspecified as well, so this leaves room for various scenarios.

Political opinions expressed in personal capacity do not normally count as advertisements, but some factors may change this: for example, if the opinion is expressed by an individual who is generally active in taking action for change.[1076] This implies that when a celebrity would express her political opinion on social media publicly, that might count as a political advertisement, in spite of the lack of payment element, for example, when it can be regarded as part of a campaign. This line is not precisely drawn by the RPA, therefore subsequent judicial interpretation might be necessary.

The exemption on "opinions" covers a smaller range of meaning than the broader term "messages" applied in the fundamental definition of advertising. As a result, it is possible that factual assertions, such as investigation reports, may not be eligible for exemption under this provision, if they are liable to impact the outcome of an election. Moreover, in order to be eligible for exemption, these "opinions" must be expressed within the framework of "editorial responsibility", the precise parameters of which remain undefined by law.[1077] Given that authors bear the editorial responsibility for

1075 Recital 48 RPA.

1076 Recital 30 RPA.

1077 Daniel Holznagel, "Political Advertising and Disinformation: The EU Draft Regulation on Political Advertising Might De-Amplify Political Everyday-User Tweets – and Become a Blueprint for Stronger Online Platform Regulation," *VerfBlog* March 23, 2023, available at: https://verfassungsblog.de/political-advertising-and-disinformation/ DOI: 10.17176/20230323–185217–0.

the content they produce, it is reasonable to put blogs and vlogs within the realm of media and hence grant them the same exception.

The original draft version of RPA limited this media exemption solely to conventional media platforms, such as linear audiovisual broadcasts or print publications, when they published the politically loaded content without any kind of direct remuneration or equivalent compensation.[1078] The legislative process expanded the scope to encompass "any media", however, references to legal and ethical norms specific to journalism have been finally omitted.[1079] In this respect, EMFA's notions on what should be regarded as a media provider, may be instructive.

8.5.2.4 Explanatory features

Identifying what constitutes political advertisement and what is regarded as "organic" content and enjoys unrestricted freedom of expression is a responsibility which will entail consequences under the RPA. Further specifications have been added by the European Parliament, with the aim to help determine whether a message constitutes a political ad or not: account shall be taken of all the features, in particular the message's content, its sponsor, the use of language, the context including the period of dissemination, its means of preparation, placement, promotion, publication, delivery and dissemination; its target audience, and its objective.[1080] We must note here that "objective" as an aspect is fundamentally distinct from the other criteria that may be objectively verified. The determination of an objective is contingent upon the preceding variables enumerated, which creates a loop in the assessment. Moreover, objective is also a duplication of the "design" element in the core definition. However, the message's design to achieve an impact is just an alternative criterion and not a necessary one.

8.5.3 Non-discrimination and third countries

The fostering of European democracy requires that European parties do not encounter obstacles when advertising their candidates in any Member State. During the 2020 EP election campaign, Facebook's regional restric-

1078 Recital 29 of the RPA.
1079 A previous version of Article 1 (2a) of the draft RPA.
1080 Article 8 of the RPA.

tions posed difficulties for the candidates in placing their ads. The internal market of advertising services similarly requires that no discrimination is allowed between Union citizens and enterprises based on their residence or citizenship. Therefore, the flow of cross-border advertising shall be ensured within the Union.[1081]

In the final version of RPA, individuals or businesses who are not residents of the EU are also allowed to sponsor[1082] political advertisements, except in the last three months before elections and referenda where some restrictions apply. EU citizens have no restrictions on sponsoring political content, and neither have third country nationals who have permanent residence in the Union and a right to vote in that specific election or referendum, according to the national law of the Member State where they reside. However, legal persons which are established in the Union can only be sponsors if they are not ultimately owned or controlled by a legal person in a third country, or by a third country national except of the previous kind of person, who is an EU resident and has the right to vote in the specific, for instance local, election.[1083]

8.5.4 Responsibilities of the actors

All actors along the value chain of creating political advertising have obligations. The broadest category is that of a provider of advertising services, it can be a creative agency or other commercial service provider, including a publisher or a platform. A sponsor is the person at whose request, or on whose behalf a political ad is created and published – most likely, but not necessarily, a political actor.

The responsibilities of the actors complete each other's as a puzzle, with some overlaps. Providers of political advertising services are obligated to ensure that their contractual arrangements with a political advertising service provider enable compliance with RPA. Among others, their online

1081 Article 5 (1) RPA.

1082 "Sponsors" refer to individuals or entities, whether natural or legal, for whom a political advertisement is specifically designed, positioned, disseminated, or circulated. Although the concept is often linked to political bodies, its applicability encompasses a wide range of circumstances that align with the principles of the RPA. It is mandatory to include the sponsor's identity in the transparency notice upon publication.

1083 Article 5 (2) RPA.

interface shall be organised in such a way that it facilitates compliance. In particular, they should require sponsors to make their declaration whether the requested service is a political advertising service, whether they fulfil the legal requirements for sponsoring such (i.e. they are citizens or residents of the EU, or fall under one of the exceptions), and other necessary information. If service providers notice that the information is manifestly erroneous, they are obligated to require its correction, and sponsors are obliged to perform the correction without undue delay. It is the obligation of sponsors to ensure that the information is accurate, complete, and constantly up-to-date.[1084] However, it is still the service providers' liability to ensure that the necessary transparency information is communicated to the publisher timely, completely, and accurately. Where technically possible, this should be done by way of a standardised automatic process. Also, the information should be ideally machine readable.[1085]

Publishers, on the other end of the transaction, are also obligated to ensure that the necessary transparency requirements are published together with the ad.[1086] To make the information easily accessible for individuals, prominent labels must be attached to the ad (the format and template are to be defined by the Commission in implementing acts, and by codes of conducts). Publishers owe best-effort obligation to complete or correct the information if they become aware that it is incomplete or incorrect. If that is not possible, they should discontinue the publishing of the ad, or refrain from publishing it.[1087]

Furthermore, links needs to be provided to the European Repository for Online Political Advertisements, and, if applicable, the information if a previous publication or an earlier version of the ad has been suspended or discontinued due to violation of RPA.

8.5.5 Political advertising by VLOPs

The RPA is to be interpreted in conjunction with the Digital Services Act (DSA). According to the DSA, it is mandatory for VLOPs to incorporate transparency details on political advertisements inside their advertising

1084 Article 6 and 7 RPA.
1085 Article 9 and 10 RPA.
1086 Article 11 RPA.
1087 Article 12 (2), Article 15 (6), Recital 63 RPA.

repositories.[1088] Furthermore, the revised wording of the RPA makes a reference to the due diligence obligations as integral components of the risk assessment obligations for VLOPs.[1089]

As a result, the Regulation will adopt the dual-tiered framework of the DSA, delineating distinct responsibilities for regular service providers and for those of significant scale. There exists a potential danger that malevolent individuals may choose smaller platforms that are not bound by stringent obligations, so evading close examination.[1090] While smaller platforms have a limited reach, resulting in a relatively reduced potential negative impact stemming from disinformation, propaganda, or other types of manipulative content, the difference between VLOPs and non-VLOPs might be less relevant in certain cases, which makes implementing uniform rules for all providers reasonable in those cases. So argued the European Partnership for Democracy (EPD) which proposed that all platforms (and not only VLOPs) should be required to publish a transparency notice on political advertisements in publicly accessible ad libraries. As managing ad libraries would be an extra burden on non-VLOPs, EPD suggested that ad libraries are managed by the Commission.[1091] This suggestion has been accepted and incorporated into RPA.[1092] The public repository enables less prominent platforms to access library services funded by the European Commission. VLOPs are still required to fulfil their obligation of publishing political advertisements in the ad libraries they maintain under the DSA. In addition, concerning the public European repository they bear an enhanced duty to share real-time information with it, whereas Non-VLOPs may submit their information within 72 hours.[1093] VLOPs must ensure its continuous maintenance during the entire period when the ad was displayed and seven

1088 Advertising repositories are provided for in Article 39 of DSA.
1089 Recitals 47, 83 RPA, connecting to Article 34–35 DSA.
1090 ERGA. "Position Paper on the proposed Regulation on political advertising (2022) as adopted." September 2022. Available at: https://erga-online.eu/wp-content/uploads/2022/09/2022-08-31-ERGA-Position-Paper-on-the-proposed-Regulation-on-political-advertising-2022-as-adopted.pdf See also: "EU: Stronger Rules Needed for Political Ads." *Human Rights Watch* March 27, 2023. Available at: https://www.hrw.org/news/2023/03/27/eu-stronger-rules-needed-political-ads.
1091 Third Opinion of the Ombudsman of the European Parliament on the European External Action Service's handling of a request for public access to documents, March 2022, available at: https://epd.eu/wp-content/uploads/2022/03/opa-3rd-edit.pdf.
1092 Article 13 RPA: European repository for online political advertisements.
1093 Article 13 RPA.

years thereafter, whereas non-VLOPs' information remains in the public repository indefinitely, even if an ad is removed or unpublished due to non-compliance.

8.5.6 Rules on targeting

The approach of the RPA is characterized by minimal intervention. This has drawn criticism for its adherence to the established frames of the GDPR rather than the introduction of novel restrictions on the utilization of personal data. Extensive academic research has strongly supported the concerns regarding the consent methods and their psychological and practical dimensions.[1094] The consenting method applied by the GDPR is widely recognized to have not achieved its intended aim. Instead of enabling users to make well-informed decisions regarding their data, it resulted in a phenomenon known as "consent fatigue", as users, feeling frustrated, quickly clicked on buttons just to get rid of the cookie banners. Legislative measures cannot be solely blamed; rather, it is the inadequate implementation of the GDPR principles that mandated privacy by design and by default. Service providers took advantage of legal loopholes by implementing intrusive cookie banners, which made it challenging for users to refuse consent and instead encouraged acceptance. Additionally, some providers employed deceptive button style and size, further adding to user confusion. The anticipated resolution of this trend is now addressed by the DSA which mandates that the act of denying consent should not impose a greater burden or require more time compared to granting consent. Furthermore, in situations where consent is not given, users should be provided with fair and reasonable alternatives for accessing services. Additionally, the DSA also prohibits the use of targeting or amplification strategies that involve the processing of minors' data (see above in Chapter on DSA).

RPA has introduced additional limitations on the utilization of personal data for targeting purposes, extending beyond the existing prohibition on

1094 EuGH ZD 2019, 556. See also: Vanessa K. Bohns, "Toward a Psychology of Consent. Perspectives on Psychological Science", 17(4), 2022, 1093–1100. https://doi.org/10.1177/17456916211040807,; Paul Graßl, Hanna Schraffenberger, Frederik Zuiderveen Borgesius, and Moniek Buijzen. "Dark and Bright Patterns in Cookie Consent Requests." Vol. 3, no. 1 (2021): 1–38.; Montezuma and Taubman-Bassirian, “How to avoid consent fatigue,” available at: https://iapp.org/news/a/how-to-avoid-consent-fatigue/.

the utilization of special category data. Consent is valid only if granted separately for the purpose of receiving targeted political advertisements, and only if it comes directly from the data subject, meaning that observed, inferred, or otherwise harvested data may not be utilized for targeting political advertisements.[1095] This is in accordance with the requirements outlined in the EDPB Guidelines 8/2020 on the targeting of individuals on social media platforms. Consequently, the processing of any data related to individuals, which is routinely handled during regular utilization of services, such as metadata, traffic, location data, or communication content (regardless of its private or public nature), should be disallowed for the specific objective of political advertising, and so is data obtained from third parties.[1096] Providers must abstain from requesting consent in situations when the data subject utilizes automated methods, such as technological specifications in browser settings, to exercise their right to object or deny consent. Profiling on special category data is prohibited, as well as the targeting of minors below at least one year of the voting age, if that information is at hand. Parties and similar organisations may target their current or former members.[1097]

Interim versions of the RPA attempted to set out detailed limitations on applying targeting criteria. One version planned to prohibit combining more than four categories of personal data, including the geographical information of the data subject. If two or more categories of data were combined, then the targeted audience had to comprise a minimum of 0.4 % of the population of the respective Member State, with a lower limit of 50,000 individuals, unless the ads were related to a concrete election or referendum.[1098] Emails or text messages that were to be sent systematically or targeted *en-masse* would have been subject to the same restrictions. During a 60-day period preceding an election or referendum, special regulations would have applied, where the applicable targeting criteria would be limited to location, language, and first-time voter status. These limitations were not passed in the final version of RPA, which returned to a very simple, transparency-based scheme.

1095 Article 18(1)a RPA.
1096 Recital 78 RPA.
1097 Article 18 RPA.
1098 Article 12 (1c) of an interim version of the Proposal RPA.

8.5.7 Enforcement of RPA

Due to the multifaceted nature of the problem, a diversity of various authorities and entities are required to participate in the implementation. National data protection authorities are to supervise and enforce the provisions relating to targeting, with the help of the European Data Protection Board (EDPB). Plans to give the EDPB strong competences to restrict the offering of targeting and ad delivery services for VLOPs have been dismissed.[1099] Member States will be responsible for appointing competent authorities in accordance with the Digital Services Act (DSA), to oversee and ensure that online intermediaries comply with their obligations under the RPA. Additionally, Member States may designate other competent authorities to oversee enforcement of other aspects of the RPA, which may coincide with the media authorities appointed by the Member States, as outlined in Article 30 of the AVMS Directive.

The effective enforcement of the Regulation necessitates the cooperation of regulatory entities at both the national and EU levels. At national level, one authority needs to function as a national contact point for all purposes by the RPA. If possible, this should be an authority that is already member of the European Cooperation Network on Elections (ECNE). European collaboration can be facilitated through this and other frameworks, such as the European Board for Digital Services and the European Regulators Group for Audiovisual Media Services (ERGA). To foster this close cooperation among the several governing bodies, the RPA establishes a permanent Network of National Contact Points as part of ECNE. This aims to facilitate the regular exchange of information and establish a structured cooperation between national contact points and the Commission, which participates in the meetings of the Network.

The competent authorities may issue warnings, order the cessation of infringements, and require sponsors or providers to take the necessary steps for compliance, apply financial measures, or request a judicial authority to do so; impose any proportionate remedies, and ultimately publish a statement which names the responsible persons. As regards the financial measures, the authorities may impose fines, financial penalties or periodic penalty payments.[1100] The definition and imposition of sanctions, including fines and of periodic penalty payments remains at the Member States,

1099 Ex-Article 16 (6a) of a previous version of RPA.
1100 Article 22 (5) RPA.

but the Commission is "encouraged" to create guidelines on the issue of sanctions.[1101] In addition, Member States shall report their sanctions to the Commission annually, which reports to the Parliament and to the Council after each EP election cycle.[1102] As seen, Member State autonomy is nuanced by a soft involvement of the Commission at several instances.

In the domain of political advertising the importance of the independence of authorities cannot be overstated. This is also emphasised by the RPA. Authorities shall enjoy full independence both from the sector and from any external intervention or political pressure. *De facto* independence is crucial especially in states where political capture may render the media authority incapable to exercise independent supervision of the political advertisement scene.

The audience is also entitled to file a complaint against a sponsor or provider of political advertising services, alleging a violation of RPA.[1103] Depending on the nature of the complaint, the relevant authorities may include national Data Protection Authorities, the Digital Services Coordinators, or the authorities designated under RPA. All authorities are obligated to investigate and address the complaint, follow up with the complainant, and as necessary, adjudicate on it or transfer it to the competent authority.[1104]

8.6 The interplay between the DSA and the Strengthened Code of Practice against Disinformation

RPA functions as the "*lex specialis*" in relation to other regulations on advertisements, including the DSA and the SCOP, encompassing aspects such as targeting and the specific responsibilities of VLOPs, which serve as the "*lex generalis*".

The SCOP explicitly refers to the emerging RPA and expresses the intention to align its terminology with it. Additionally, it instructs signatories to comply with RPA and DSA. The advantage of establishing a SCOP with parallel rules to the RPA, is to facilitate the implementation of these promises before they are officially in force. Simultaneously, the SCOP also

1101 Article 25, Recital 108 RPA.
1102 Article 25 and 27 RPA.
1103 Article 24 RPA.
1104 Recital 105 RPA.

aims to develop extensive knowledge and protocols to ensure the efficient application of the legal principles.

Concerns arose regarding the implementation of the SCOP, as it may potentially incentivize providers to engage in excessive censorship due to the less well-defined parameters as compared to the DSA. The absence of safeguards in place and the possibility of including permitted yet objectionable information further aggravates this risk.[1105] While this risk is indeed relevant, there exists an opposite problem pertaining to under-enforcement as well. This arises from the fact that the SCOP lacks direct enforceability, rendering its measures optional and the entire endeavour voluntary in nature.

8.7 Interim conclusion on RPA

The initial version of the RPA was primarily characterized by its emphasis on transparency, and after various legislative rounds, the final version returned to this approach. As an advantage, this leaves the freedom to deliver political advertisements relatively intact and focuses on the empowerment of users. However, it may prove insufficient in providing protection particularly for vulnerable populations who may not consciously choose to consult the repositories. Interim versions of the RPA suggested some innovative methods aimed at safeguarding public discourse against discriminatory practices.

The definition of what constitutes political advertising has been more elaborated during the legislative process. By adopting a more precise definition, organic content can effectively be excluded from the scope, provided that it is not endorsed or magnified.[1106] This aligns with prior policy suggestions[1107] that advocated for regulating based on the method of distribution and other neutral characteristics, rather than the content itself. Although

1105 Stefania Galantino, "How Will the EU Digital Services Act Affect the Regulation of Disinformation?," *SCRIPTed* 20 (2023): 89.

1106 Daniel Holznagel, „*Political Advertising and Disinformation: The EU Draft Regulation on Political Advertising Might De-Amplify Political Everyday-User Tweets – and Become a Blueprint for Stronger Online Platform Regulation*," *VerfBlog* 2023/3/23, https://verfassungsblog.de/political-advertising-and-disinformation/DOI: 10.17176/20230323–185217–0.

1107 Judit Bayer et al., "Disinformation and Propaganda: Impact on the Functioning of the Rule of Law and Democratic Processes in the EU and its Member States – 2021 Update," Study requested by the European Parliament's Special Committee

malevolent strategic misinformation and propaganda have the potential to masquerade as organic material and evade detection, once targeting or sponsorship is included, they will attract attention and scrutiny. This measure would guarantee the preservation of freedom of expression for unaffiliated, non-institutional advocates of disinformation, including sincere proponents of conspiracy theories and possibly innovative dissenting viewpoints. The preservation of this liberty holds significant importance inside illiberal or authoritarian regimes, as it serves to prevent the suppression of open and constructive public dialogue. At the same time, this flexibility can also be taken advantage of by highly organized troll armies that employ numerous accounts, rather than relying on algorithmic amplification or sponsorship. This problem can be addressed by the SCOP which provides a framework via which platforms can (and VLOPs should) effectively prohibit the presence of fake profiles, automated behaviours, and other manipulative uses of their services.

Whether specific statements of influencers could qualify as a political ad remains somewhat ambiguous and needs determination on a case-by-case basis. Given the fluidity of boundaries between the genres of communication in the new information environment, this is likely inevitable. During the implementation, efforts should be made to prevent that such questions result in divergent practices. To ensure consistent implementation, RPA requires a strong collaboration between national and supranational authorities and bodies, while also striving to uphold Member State autonomy to the extent possible.

RPA, like DSA as well, appears to be an ongoing project. It created a reporting cycle, defining areas of supervision after every EP election. In particular, the effectiveness of the measures shall be evaluated, among others those relating to the transparency of the ads, and how the rules applying to targeting protect personal data. The impact on micro, small and medium-sized media actors will also pose an intriguing question. On the one hand, they may be too strongly hit, on the other, malicious actors might exploit the lighter rules.

on Foreign Interference in all Democratic Processes in the European Union, including Disinformation (INGE), 2021.

9 The wider technological environment: AI Act

Digital communication depends to a large extent on the use of artificial intelligence (AI). AI will increasingly provide the basic infrastructure for social actions, as more and more activities transform into being online and digital, and it will define the boundaries and possibilities of online activities. In absence of regulation, development and innovation in the field of AI (like in other fields of technology) have been motivated by financial and economic interests. Background processes – developing, training and data input – serve particular private interests rather than the public good. Social good, or the protection of individual human rights, did not play a leading role in innovation. Of course, commercial applications aim at achieving user satisfaction, and – at least where end users are human individuals, like in the case of ranking algorithms – direct violation of individual rights is not perceivable. When the end users are companies and individuals are "subjects", like in the case of hiring algorithms, the picture is less clear: obvious discriminatory uses were recorded, such as in the Amazon hiring software[1108] or Compas, a forensic tool to predict recidivism.[1109] But even in these cases, the discriminative outcome was not obvious at first sight. Human individuals are not in the position to take notice of the human rights infringements because that would require an oversight of all the results, or an insight into the system's operation.

The public discourse is affected by a few specific AI applications: in particular algorithmic ranking, generative AI and profile-based applications such as targeting, recommending and social scoring. In addition, any instance where the use of AI negatively impacts the enjoyment of fundamental rights, is likely to indirectly influence the participation in the public discourse. In particular the rights to privacy, equality and dignity are essential prerequisites for the exercise of freedom of expression and an effective and conscious seeking of information. Therefore, these areas are examined

1108 Ifeoma Ajunwa, "The paradox of automation as anti-bias intervention," *Cardozo Law Review* 41, no. 1671 (2019).

1109 Sarah Brayne, *Predict and Surveil: Data, discretion, and the future of policing* (New York, NY: Oxford University Press, 2020). See more in: Cathy O'Neil, *Weapons of math destruction: How big data increases inequality and threatens democracy* (New York, NY: Crown Publishing Group, 2016).

in-depth below, whereas other specifics of the AI regulation will not be addressed in detail.

The aim of the European Union's AI Act is to address the human and ethical implications of AI usage, „proposing a legal framework for trustworthy AI". Similarly to the logic of the GDPR and that of the E-Commerce Directive, the logic was that a more trusted environment will give users confidence to embrace AI-based solutions and indirectly boost their development and deployment.[1110]

9.1 A literally disruptive technology

AI as a regulatory target is moving so fast, that it is literally challenging to shoot at. When the draft AI Act was in a mature phase of the legislative process, ChatGPT exploded into the commons, and generated numerous new avenues for future development, uses and regulatory considerations. The model that has appeared almost synchronously under several other names provided by various companies (BERT, DALL-E, etc.) changed the public perception of AI. While not the first AI in history that actually passed the Turing test,[1111] it was the first to capture widespread public attention.[1112] Beyond this, it was made publicly available for anyone to use, free of charge.[1113] These generative AI models are capable of creating text, images or video, based on prompts written in natural language. The most developed skill is the creation of text, which gives the impression that the AI is engaging in conversation with the user. It is easy to forget that the responses just follow the law of probability: they rely on patterns learned from large datasets to predict the most probable sequence of words in response to a given prompt. They are neither based on logic in the Aristotelian sense, nor on background research.

1110 Explanatory Memorandum 1.1. to the Proposal for the AI Act. https://www.eumonitor.eu/9353000/1/j4nvhdfdk3hydzq_j9vvik7m1c3gyxp/vli6mrzmjby8.

1111 ChatGPT was the second chatbot that passed the Turing test. https://www.mlyearning.org/chatgpt-passes-turing-test/. See more on the Turing test: https://www.techtarget.com/searchenterpriseai/definition/Turing-test.

1112 This event has been so widely reported on, that it appears unnecessary to give a reference. Google it, or try it yourself: have a chat with ChatGPT here: chat.openai.com.

1113 https://chat.openai.com.

Policymakers had a hard time to catch up and come up with solutions to design safeguards for the new situation. One of the leading regulatory concepts in October 2023 added the notions "general purpose AI" and "foundation model" to the draft Act.[1114] It defined general purpose AI system (GPAI) as an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed", and can be adapted to a wide range of distinctive tasks.[1115] Whereas, foundation models were defined as applications that are not only designed for general purpose, but are already trained on a large data set. The two categories were finally amalgamated in the text, dropping the term "foundation models" and keeping merely "general purpose AI systems", and "general purpose AI models". Some intended requirements of foundational models were incorporated as rules applying to all GPAI. In a previous version of the proposed Act, stricter regulations were to govern foundational models, necessitating compliance throughout their lifecycle, encompassing all stages of the value chain, notably development, training, and distribution. Core principles such as interpretability, corrigibility, safety, and cybersecurity were to be upheld, supported by documented analysis and tested by independent experts. Foundation models were subject to registration, with their documentation mandated to be retained for a period of 10 years after launch. Compliance requirements included to encompass sustainability concerns, requiring providers to provide data relating to the energy consumption during the training phase and the duration necessary for model training. These stricter rules were not passed in the final version of the AI Act.

In the final Act, GPAI models are classified as limited risk models unless they present systemic risks. The negotiations which extended the legislative process with a year, finally reached a compromise: GPAI are regulated, but a large part is referred to self-regulation, with an excuse to avoid stifling European innovation. A more detailed discussion of these will follow below.

1114 EP Mandate for the Trilogue. Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative Acts 2021/0106(COD) DRAFT 20–06–2023 at 16h53. Retrieved from: https://www.kaizenner.eu. Digitizing Europe.

1115 Stanford Univeresity CRFM – HAI (2021) On the Opportunities and Risks of Foundation Models, https://fsi.stanford.edu/publication/opportunities-and-risks-foundation-models cited by Euractiv: https://www.euractiv.com/section/artificial-intelligence/news/ai-act-meps-close-in-on-rules-for-general-purpose-ai-foundation-models/.

9.1.1 The scope and subject matter of the AI Act

Similar to the GDPR, the AI Act will have an extraterritorial personal scope. It applies to providers who place on the market or put into service AI systems or general-purpose AI models in the European Union, or their authorised representatives if they are not established in the Union; as well as importers and distributors of AI systems. It further applies to deployers of AI systems who are located or established within the Union. The use of the word "deployer" is an achievement of the academic community, which held the word "user" misleading.[1116] Persons who or which are now called deployers, are any natural or legal person, public authority, agency or other body, who or which use an AI system under their authority. The word "deployer" clearly differentiates controlling users from those natural persons who use an AI system as clients, calling the latter "affected persons". The Act also applies to affected persons who are located within the Union. Affected persons (also called as end-users in plain language) have no practical influence on the operation of the system, beyond forming its output with their own data. These actors were also designated as "decision subjects".[1117] The JURI Committee followed Ada Lovelace Institute's feedback when it added the definition of "affected person": a natural person or a group of persons who are subject to, or affected by, an AI system. However, this definition did not find its way into the final Act: the term remains simply without definition. The professional discourse around the AI Act has taken over the distinction between "AI developers, deployers and users" well before the Proposal's text did.[1118]

Deployers may also take on the role of affected persons, when they use an AI system during a personal, non-professional activity.[1119]

1116 Ada Lovelace Institute, People, risk and the unique requirements of AI. 18 recommendations to strengthen the EU AI Act. 31 March 2022.

1117 Sebastian Bordt et al., "Post-hoc explanations Fail to Achieve their Purpose in Adversarial Contexts," *arXiv:* 2201.10295 last modified 10 May 2022.

1118 European Commission: Regulatory framework proposal on artificial intelligence https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai. See also: de Matos Pinto, I. The draft AI Act: a success story of strengthening Parliament's right of legislative initiative?. *ERA Forum* 22, (2021): 619–641. https://doi.org/10.1007/s12027-021-00691-5. See also: https://algorithmwatch.org/en/ai-act-explained/.

1119 Article 3 (4) AI Act.

As to the material scope of the Regulation, fierce debates were led about the appropriate definition of 'artificial intelligence'. The original definition would have covered almost every complex computer software.[1120] The Second Presidential Compromise text defined an AI system as a system that was designed to operate with a certain level of autonomy and that infers how to achieve a given set of human-defined objectives, and produces content, predictions, recommendations or decisions, influencing the environments with which the AI system interacts (simplified wording). This definition relied on the feature that an AI system is able to decide itself how to solve a certain task, with the tools that it had been equipped with, such as machine learning, logic- and knowledge-based approaches. The final definition of AI system follows this logic.[1121] When general purpose AI systems appeared on the market, intense work commenced to find appropriate definitions and liability schemes. Finally, they are defined as an AI system, which is based on a general purpose AI model, can serve a variety of purposes, both directly or integrated into other AI systems, whereas a general purpose AI model is an AI model trained with a large amount of data using self-supervision at scale (formerly called foundation model), that displays significant generality, can perform a wide range of distinct tasks, and can be integrated into a variety of downstream systems or applications.[1122] The general functions can be for instance image or speech recognition, audio or video generation, pattern detection, conversation, translation or others.

It is further notable that among the scope of activities that are regulated, the creation and the development of AI systems or models is not included. The regulated activities are "placing on the market", "putting into service", and "use" in the Union.[1123] The relevant market is the internal market of the EU, which leaves open the possibilities to produce, create, and develop AI systems or models which do not correspond to the requirements of the Act, for other markets. Thus, unethically developed, unsafe and untrustworthy AI tools can be lawfully created, developed and transferred to third countries, where those are not regulated.

1120 Martin Ebers et al., 2021. "The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI" *Law Society (RAILS)" J* 4, no. 4 (2021): 589–603. https://doi.org/10.3390/j4040043, at p. 2.

1121 Article 3 (1) AI Act.

1122 Article 3 AI Act.

1123 Article 2 (1) AI Act.

This permissive standing is explained by the economic interests of the EU. Ideally, a global covenant should be established to promote human-centric and trustworthy AI practices. The inaugural effort toward this end has been initiated by UNESCO through the introduction of the first global agreement on ethical principles.[1124]

This is not the sole example of critical compromises on the scope of the Act. Military use, defence and national security remain outside the AI Act's scope. The imperatives of human-centrism and respect for human rights are obviously set aside to facilitate the development of intelligent weapons.[1125] The issue is deferred to the realm of international humanitarian law. The Convention on Certain Conventional Weapons (CCW) established a Group of Governmental Experts to deal with the matter in Geneva.[1126] At the same time, the military continues to be the primary catalyst for advancements in the field of AI, as with other technologies in the past.[1127] Autonomous weapons are already being deployed, despite serious ethical objections, and concerns pertaining to global security weigh heavily. The scientific discussion on lethal autonomous weapons is divided. While some authors argue that autonomous systems can pursue military goals without the loss of human lives,[1128] other authors warn that a) no artificial system should ever be in the position to autonomously decide if a weapon should be applied against a human being, b) that allowing robotic wars disregards the right to self-government and autonomy of the peoples and sovereignty of current states, ultimately threatening democracy at large.[1129] Beyond ethical concerns, treating AI weapons as a practical tool to pursue "only military goals" could open the door to a new wave of colonisation.

1124 https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.

1125 Asimov's first law, that robots may not do harm to a human being, would have to be disabled in their case.

1126 Eugenio V. Garcia, "Artificial Intelligence, Peace and Security: Challenges for International Humanitarian Law" (October 7, 2019). *Cadernos de Política Exterior nº 8, Instituto de Pesquisa de Relações Internacionais (IPRI)*, Brasilia, 2019, Available at SSRN: https://ssrn.com/abstract=3595340.

1127 Michael C. Horowitz, "When speed kills: Lethal autonomous weapon systems, deterrence and stability," *Journal of Strategic Studies*, 42, no. 6 (2019): 764–788. DOI:10.1080/01402390.2019.1621174.

1128 John Yoo, Jeremy Rabkin: "Striking Power: How Cyber, Robots, and Space Weapons Change the Rules for War," See also: aei.org.

1129 How I Learned to Stop Worrying and Love the Bots, and How I Learned to Start Worrying about Democracy Instead Antonio F. Perez 27 *Cath. U. J. L. & Tech* 129 (2019).

The annexation of countries through the use of cyberweapons, absent physical violence, raises questions regarding compliance with the Geneva Convention. Regardless of the answer, such a scenario would undoubtedly pose a significant threat to democracies and human rights. In any case, the potential of AI may amplify the risk of military conflict escalation.[1130]

In sum, the implied scope of the Act extends to all civil uses of AI, primarily commercial AI systems and applications that are put into service and used within the EU including open source. However, it excludes foreign use and research purposes from its scope, whereas military and national security purposes are beyond European competence.

9.1.2 The regulatory model

Four categories of risk can be distinguished under the AI Act, and different obligations attach to each category. The first category is no-risk, for which no obligations apply beyond the ordinary product liability. These would include, for example, AI-enabled video games or spam filters. The vast majority of currently used AI systems reportedly falls into this category. Second, certain AI applications are entirely prohibited because they are deemed as disproportionately injurious to human rights. It should be noted that in this regard, the Act's limited material and territorial effect can play an important role, as these AI applications can still be used for military and national security purposes, or manufactured within the EU, but placed on the market, put into service, and used outside the EU. The category of limited risk includes emotion recognition systems, biometric categorisation systems (those which are not prohibited as unacceptable risk), deep fakes, and the long-debated general purpose AI models.[1131] Finally, the regulation of high-risk AI systems fills the majority of the AI Act. There is an overlap between the third and the fourth category, as some general purpose AI models may pose a systemic risk.

1130 James Johnson, "The AI-cyber nexus: implications for military escalation, deterrence and strategic stability, *Journal of Cyber Policy* 4, no. 3 (2019): 442–460. DOI:10.1080/23738871.2019.1701693.

1131 Article 50, 51 AI Act.

9.1.3 Prohibited AI practices

9.1.3.1 Manipulation

AI systems that deploy subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, if used with the goal of causing harm, are prohibited.[1132] "Subliminal techniques" involve sensory stimuli below an individual's threshold for conscious perception and have been known since the early 20th century, when they were used for advertising purposes. They got prohibited in several countries by media and advertising regulation, and also in the EU's Audiovisual Media Services Directive.[1133] There is considerable literature discussing the manipulative potential of subliminal advertising, both commercial and political.[1134] When combined with the power of AI systems, this manipulative potential arguably expands and accelerates.

The AI Act prohibits manipulation only if deployed either with a specific objective or with a specific effect: of materially distorting a person's or a groups' behaviour by appreciably impairing the person's ability to make an informed decision, thereby causing the person to take a decision that that person would not have otherwise taken, in a manner that causes or is likely to cause significant harm.[1135] The same restriction applies to AI systems that exploit any of the vulnerabilities of a specific group of persons due to their age, disability, social or economic situation, with the same objective as in the former case.

Would the technique used in the infamous Cambridge Analytica political scandal count as a prohibited technique? The coordinated strategic disinformation and targeted political advertising campaign that took place with the help of the Cambridge Analytica company utilised the personal

1132 Article 5 AI Act.

1133 The Audiovisual Media Services Directive Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive). While several states prohibit subliminal advertising, the EU's advertising regulation prohibits merely "unfair practices" as well as misleading or aggressive practices, which certainly would consume subliminal practices. https://www.europarl.europa.eu/doceo/document/E-8-2015-003522-ASW_EN.html?redirect. Answer given by Ms Jourová on behalf of the Commission.

1134 Vance Packard, *The Hidden Persuaders* (Brooklyn, NY: Ig Publishing, 2007).

1135 Article 5, a-b AI Act.

data of 90 million Facebook users, exploited their vulnerabilities, and targeted them without their knowledge. It is, however, questionable, whether being targeted unconsciously would already be considered subliminal. The potential of subliminal techniques will further grow with the launch of the "metaverses", the use of virtual reality (VR) technology, and augmented reality. In the future, brain-computer interfaces, emotion recognition and other "brain spyware" may further expose vulnerabilities and open new avenues for uncontrollable manipulation.[1136]

The restriction does not preclude experimenting with such techniques, or their use for research purposes; rather, it merely prohibits their distribution and use within the EU market. The scope of the prohibition stirred considerable criticism. Foremost because in both cases, one of the definitional elements is that the technique will cause harm. It is often difficult to bring evidence about harm, especially psychological harm, and even more so to establish a causal relationship between a particular impression and the harm.[1137] As a further substantial limitation, it observes only individual harm, caused to concrete persons. Whereas societal harms and collective harms are not registered.[1138]

As some authors argue,[1139] societal harms are not adequately considered when the regulation only takes account of individual harms. As unfolded in more detail in the Introduction above, in our digitalised and interconnected world individual harm hardly ever exists. The myriads of connections between billions of natural and legal persons, and the accelerated speed of processing results make even tiny human rights infringements accumulate and generate larger societal harms. Regulation could arguably take into

1136 Rostam Josef Neuwirth, "The EU Artificial Intelligence Act: Regulating Subliminal AI Systems" (August 15, 2022). (London: Routledge, 2023). https://www.routledge.com/The-EU-Artificial-Intelligence-Act-Regulating-Subliminal-AI-Systems/Neuwirth/p/book/9781032333755, Available at SSRN https://ssrn.com/abstract=4135848 or http://dx.doi.org/10.2139/ssrn.4135848.

1137 Martin Ebers et al., "The European Commission's," 589–603. https://doi.org/10.3390/j4040043.

1138 Michael Veale and Zuiderveen Borgesius, F. "Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach," *Computer Law Review International* 22, no. 4 (2021): 97–112.; Available online: https://osf.io/preprints/socarxiv/38p5f/.

1139 Veale and Borgesius, "Demystifying the Draft," 97–112.; Nathalie A. Smuha, (2021). "Beyond the individual: governing AI's societal harm," *Internet Policy Review* 10, no. 3 (2021).

account collective, societal harms and develop policy instruments to deal with them, as these cannot be tackled with the individual legal remedies.

For example, using subliminal techniques with the objective of distorting a person's voting behaviour, would not cause individual harm to any person, at least not harm that could directly be attributed to the subliminal techniques. However, if masses of people are exposed to such manipulation, their cumulated action under influence may lead to a failure of genuine democratic processes, for example, by persuading a relevant proportion of voters to abstain from voting. Veale and Borgesius bring the example of intimate partner violence where underlying dynamics are increasingly considered, as opposed to one-off events.[1140] Moreover, it is also questionable whether a technique alone can be called responsible for a harm caused, whereas the content of manipulation is left out of sight. A careful court would likely hesitate to establish a direct causal link between specific harm and subliminal techniques without considering the content.

What counts as a material distortion of behaviour? For example, would the pursuing of instinctive triggers, like spending lavishly on luxurious items, count as distorted behaviour, if a person showed previously more self-control? Would it be possible to prove harm as a consequence? The AI Act appears to allow this interpretation. Ultimately these questions remain to be answered by legal practice.

9.1.3.2 Social scoring

The Act further prohibits certain well-tailored cases of what is generally known as social scoring. Both public and private entities are prohibited from using an AI system for the evaluation or classification of social behaviour or personal characteristics of natural persons or groups thereof, over a certain period of time, based on their social behaviour, or based on their personal or personality characteristics, if the social score leads to a detrimental or unfavourable treatment of persons or groups in social contexts that are unrelated to the original data. Or, in the case that the context is related, if the detrimental treatment is unjustified or disproportionate to the gravity of their previous social behaviour.[1141]

1140 Veale and Borgesius cite: Evan Stark and Marianne Hester, "Coercive Control: Update and Review" 25 Violence Against Women 81 (2019).

1141 Article 5 (c) AI Act.

This prohibition too, like that of manipulation, still allows wide room to apply social scoring if the ensuing classification leads to a favourable or neutral treatment, or, even if it leads to a detrimental treatment but the context is relevant to the original behaviour, and the consequences are justified and proportionate. For example, a justified use may be if a person who was repeatedly found to misbehave on public transport, could be excluded from using public transport. However, the Act also allows to apply the restriction based on a person's inferred or predicted personality traits as long as it is not unrelated to the context, and is not disproportionate.

And, if a person regularly behaves "well" socially, can get discounts or benefits in relation to other, unrelated social services. The limitation of the prohibition to negative consequences opens an old debate on the broader effects of positive discrimination. The possibility to use social scoring to govern the behaviour of masses of people, is too enticing to be bypassed if the practice is not consistently banned. The use of this practice is likely to boost interest in surveillance and incentivise further invasions into privacy under the veil of consenting for bonuses. Using the legal leeway, companies and authorities are allowed to sanction not only behaviour that is not illegal, but also predicted personality traits. This is bound to lead to indirect discrimination of disadvantaged and marginalised groups, and represents a grave threat to freedom of expression and the public discourse. Online speech, for example, on social media, is likely to become the basis of predictions on the personality. Even minor sanctions, if consistently applied, are bound to cause a chilling effect in the public discourse.

Considering that the national security purposes are exempted from the AI Act's scope, authorities will be allowed to use social scoring for national security purposes.[1142] This prospect evokes justified concerns for citizens residing in illiberal states, presenting a chilling perspective on their privacy and other related rights. Even without consequences, profiling alone is injurious to privacy, dignity and equality.

9.1.3.3 Biometric identification

Real-time biometric identification in publicly accessible spaces is regulated along the principles of secret surveillance regulation. For the purposes of law enforcement, it will need to be justified case-by-case, authorised by a judicial or independent administrative authority. The cases, the main

1142 Article 2 (3) Scope, AI Act.

safeguards, and exceptions for the authorisation procedure are laid down by the Act. Several causes are offered as justification for the use of such identification method: when it is 'strictly necessary' for pursuing specific, high-profile criminal investigations, such as targeted search for potential victims of abduction, trafficking in human beings and sexual exploitation or search for missing persons, the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons, or a genuine, and either present or foreseeable, terrorist attack. However, simpler cases, too, can serve as ground for real-time biometric identification, such as to localise or identify a suspect who committed a criminal offence, for the purposes of conducting a criminal investigation, prosecution or executing a criminal penalty for offences, that are punishable in the Member State by at least a custodial sentence for a maximum of four years.

The use of the system may not extend beyond confirming the specifically targeted individual's identity, as proportionality needs to be respected, particularly the seriousness, the probability and the scale of the threatening harm and the harm caused to the human rights and freedoms with the use of the system. The safeguards include temporal, geographic and personal limitations.

As a matter of national security, ample space is left for Member State legislation, especially so for a genre of European regulation. The Member States have considerable room for manoeuvre to define the content of the legal requirements for authorising the use of these systems for the purpose of law enforcement within the frames laid down by the Act. The AI Act counts with a detailed national legislation that would set out rules for the request, issuance, and exercise of biometric identification. This should generally take place in advance of the commencement of the supervision. However, in case of "urgency", the supervision can be started without an authorisation, which shall be requested without undue delay but within at most 24 hours. If rejected, the surveillance shall be stopped with immediate effect.[1143] The Act further requires that the law enforcement authority must perform a fundamental rights impact assessment and register the system in the database for high-risk systems.[1144] Binding authoritative decisions that have an adverse effect on a person, may not be taken solely on the basis of the remote biometric identification system's output. In any case,

1143 Article 5 (3) AI Act.
1144 Article 5 (2) AI Act, referring to Article 27 and Article 51.

such systems shall be notified to the relevant market surveillance authority, and to the national data protection authority. In a similar vein, biometric categorisation systems that would categorise individually natural persons to infer their special category data (race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation), are prohibited, except for law enforcement, and except for labelling and filtering lawfully acquired biometric datasets.[1145] The described rules may theoretically require adequate safeguards, but the possibility of biometric identification still exercises a considerable chilling effect on public participation, for example, on demonstrations. In addition, large divergences can be expected because of the wide room for Member States to regulate.

9.1 High-risk AI systems

Currently, AI systems have three principal points of contact with the public discourse: algorithmic ranking and targeting, and generative AI and social scoring. All have niche areas which qualify as high-risk. Algorithmic ranking and generative AI will be specifically discussed below. Otherwise, on the high-risk regulatory category a brief overview is provided as background information before going into more detail in the relevant fields.

9.1.1 The scope of high-risk AI systems

The biggest regulatory impact of the AI Act is expected to be made on applications that will categorise as high-risk AI practices. These systems will be subject to several restrictions and their operators will owe obligations. There will be a considerable regulatory gap between applications that are listed as high-risk and others that are not. A late amendment by the JURI Committee suggested general ethical principles that are "strongly encouraged" to be respected by all AI systems, in an apparent attempt to bridge this gap.[1146] That proposal was not incorporated into the final version, instead, all providers of non-high-risk AI systems are encouraged to draw

1145 Article 5 (g) AI Act.

1146 Interim version of AI Act, Article 4a of the AI Act's amendment by JURI Committee. https://artificialintelligenceact.eu/wp-content/uploads/2022/09/AIA-JURI-Rule-57-Opinion-Adopted-5-September.pdf.

up and comply with codes of conduct, including governance mechanisms. These should contain clear objectives and key performance indicators, and principles similar to the European ethic guidelines for trustworthy AI, environmental sustainability, promoting AI literacy, inclusive and diverse design, and preventing the negative impact of AI systems on vulnerable persons or groups, for example persons with disability. The creation of the code of conduct shall be facilitated by the AI Office.[1147] Even with the code in place, the determination of which applications qualify as high-risk systems and which do not will remain critically important for market actors. Given that the list can be amended by the Commission, this issue is likely to remain a focal point of attention in at least the medium term.

There is more than one way to classify a system as high-risk. First, systems that are covered by Union harmonisation legislation (listed in Annex II), if that legislation requires them to undergo a third-party conformity assessment are considered as high-risk systems, whether they are the listed product themselves, or a safety component of the product. This list includes mainly transport-related and industrial applications. Second, Annex III lists eight types of systems that are considered as high risk. AI systems belonging under these categories are presumed to be high-risk unless they can prove one of the exceptions that show that the systems' intended tasks are not intended to materially influence the outcome of decision-making.[1148] However, AI systems that perform profiling of natural persons (and are listed in Annex III) will always be considered high-risk.[1149]

The types of systems listed in Annex III predominantly pertain to those that impact large social systems, like education, employment, essential services, migration and border control, law enforcement, justice and democratic processes. In regard of this last point, the original proposal of the Act merely referred to AI in justice.[1150] However, final amendments added

1147 Article 95 AI Act.

1148 They are not regarded as to pose a significant risk of harm to the health, safety or fundamental rights of natural persons, if they are intended to perform merely a narrow, procedural task, if they are intended to improve the result of a previously completed human activity, if they are to detect patterns or deviations in decision-making, not to replace human assessment, or if they are to perform a preparatory task to an assessment relevant for the purpose of the listed cases.

1149 Article 6 (3) AI Act.

1150 "AI systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts."

a point b),[1151] explicitly naming AI systems intended to be used to influence the outcome of an election or referendum or the voting behaviour of natural persons. These are to be classified as high-risk system except if their outputs are not directly accessible to natural persons, such as tools used to organise, optimise and structure political campaigns from an administrative and logistical point of view.[1152] The definition refers to the activity of delivering political content, in particular political advertisements, especially when these are targeted. Although, "general" algorithmic content ranking and recommending systems which govern public communication, including culture, opinion, news and other content, also do influence public discourse and thereby, the democratic processes. However, as they are not *intended* to influence the outcome of a specific election or a referendum, they apparently do not qualify as high-risk systems.

In the light of the prospective algorithmic governance of metaverses and other progressively intrusive communication platforms, this possibility to extend the stricter criteria to these general algorithms remains pertinent. The possibility for the Commission to include, through a delegated act, further AI systems would allow taking this path. Simultaneously, it is also conceivable that ranking algorithms could be classified as high-impact general-purpose models. Below, the possibility of updating Annex III with content ranking algorithmic systems is examined.

9.1.2 Updating the list of high-risk systems

Adding new criteria, as well as deleting or modifying them, also to add or remove items to or from the list under some conditions, is possible for the Commission by delegated acts.[1153] However, the eight categories listed in Annex III bind the Commission's hands, as it is entitled to update the list only with AI systems in the originally given fields: the list itself cannot be extended. The second, conjunctive condition is that the risk of harm to the health and safety, or a risk of adverse effect on fundamental rights should be, in severity and probability, equivalent or greater than that posed by already listed applications. This expectation is rather abstract, because the list of Annex III does not define levels of harms or risks, but areas of

1151 Annex III 8(b).
1152 Recital 62, Annex III. 8. (b) AI Act.
1153 Article 6 (6) and Article 7 AI Act.

applications. To provide some objectivity to this assessment, eleven criteria are listed that the Commission must consider when assessing this risk. Several of these would fit the context of social media algorithms. As algorithmic systems or other AI systems that govern content ranking, content recommendation on moderation, have a profound formative effect on the public discourse as they influence the fundamental rights of freedom of expression, freedom of information, the right to vote, and thereby on democratic participation, it is reasonable to examine whether all such systems could qualify as high-risk systems, and not only those which are intended to influence voting behaviour, or the outcome of elections or referenda.

The first criterion for the Commission to consider is the intended purpose of the AI system. Although no further explanation is added, it can be suspected that the intent should extend to exercising an impact on the health, safety, or fundamental rights of natural persons. Depending on the nature of a platform, the operators may count with exercising an impact on the fundamental rights of persons. If this is an unforeseeable and incidental effect – like it was for Facebook before the Arab Spring, or for WhatsApp before lynching incidents in India – then the intent is probably not given. However, knowing all the events in the past years, social media platforms must have knowledge about the impact they make, and they govern their platforms knowing these risks. Platforms other than social media, for example, retailers of goods, services (Amazon, Booking) or media platforms (Netflix, Apple Music) without the social element do not need to count with the same risk, at least according to our current knowledge.

The second element for the Commission to consider is the extent to which an AI system has been used or is likely to be used.[1154] As social media algorithms are used ubiquitously, across the globe, on a daily basis, by masses of people, this criterion would be fulfilled. According to these criteria, AI systems that are used in everyday applications such as mobile phones are to be considered as well. The basic underlying systems of these devices would count as general-purpose AI which can be used as parts in a plurality of AI system applications, and which have to fulfil merely certain transparency requirements (more on these below).

The third to consider is the extent to which the use of an AI system has already caused harm to the health and safety or adverse impact on the fundamental rights or has given rise to significant concerns in relation to the materialisation of such harm or adverse impact, as demonstrated by

1154 Article 7 (2)b.

reports or documented allegations submitted to national competent authorities". Fundamental rights have been harmed for example in Myanmar, in Washington on 6 January 2021, or in relation to the Covid-19 patients who could have avoided infection. However, the causes that led to these unfortunate results, are only indirectly related to the AI system. Further necessary elements had to contribute: that disinformation was posted online, that users interacted with the disinformation and that they reacted to it in the physical world, causing direct harm to other persons. It was the people who caused harm to the health, safety or fundamental rights of other persons (Capitolium) or to themselves (vaccine-deniers). On the one hand, we know that the AI systems have amplified the effect and influenced, perhaps manipulated these people. On the other hand, we cannot take it for granted – and certainly lack evidence – that the same results would not have also occurred in the absence of the algorithms, or if all the requirements set out for high-risk applications were observed. The risk management system of high-risk applications needs to consider only those risks which may be reasonably mitigated or eliminated through design, development, or providing technical information. Risks that cannot be mitigated through these are apparently considered as falling outside the reach of AI providers and developers and are therefore regarded as legally irrelevant.

A further criterion to be considered is the potential extent and the intensity of the harm or adverse impact, and its ability to affect a plurality of persons. Considering the adverse effects of the data-driven, attention-harvesting social media ranking systems, it can be safely said that their effect, even if indirect, and influenced by several factors, is extensive, profound and affects millions or billions of persons. For example, assuming that elections or referenda were influenced through data-driven algorithmic systems, then all inhabitants of the country would be affected, and even citizens of other countries. Whether the effects are adverse effects, may be disputable: after an election, the winning party will argue that the effects were beneficial. For instance, notwithstanding the consensus that the Brexit referendum took place amidst a manipulative media environment, and fact-based research on what this costed to the UK, the UK government

official website promotes the benefits of Brexit.[1155] The Commission must consider the benefits together with the harms.[1156]

The Commission would further have to consider the degree to which adversely affected individuals rely on the outcomes generated by an AI system, as well as whether they have the ability to opt out, either legally or practically. Currently, there is no option to opt out of using social media algorithms, except by refraining from using social media platforms altogether. The Digital Services Act requires from VLOPs and VLOSEs to offer at least one alternative option for users, and with it, this situation may change.[1157]

In addition, the Commission needs to consider the aspect whether the extent to which potentially harmed or adversely impacted persons are in a vulnerable position in relation to the user of an AI system, in particular due to an imbalance of power or knowledge. Ample evidence supports that data-driven algorithmic ranking and targeting has exploited the vulnerabilities and the knowledge gap of users.[1158]

Furthermore, it is relevant, to what extent the outcome produced with an AI system is easily reversible. The impact on the public discourse is irreversible: mental conceptions take root in the culture and get reproduced even when the manipulation terminates. Correction requires a concerted effort of education, strengthening the sphere of quality content media and many other measures. Manifested effects like Brexit, election results or violence, are clearly irreversible. Furthermore, the magnitude and likelihood of benefit of the AI system should be considered, for individuals, groups, or society at large. While there is a clear benefit in content ranking systems, at stake would be not their elimination, rather ensuring their transparency and the empowerment of the affected persons to exercise control over them.

Ultimately, the Commission should assess the extent to which existing Union legislation provides effective redress or measures to prevent or minimise the risk. In accordance with this, it is reasonable to wait and see

1155 HM Govt, *The Benefits of Brexit: How the UK is taking advantage of leaving the EU, 2022*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1054643/benefits-of-brexit.pdf.

1156 Luca Bertuzzi, *AI Act: Czech Presidency puts forward narrower classification of high-risk systems, 2022,* https://www.euractiv.com/section/digital/news/ai-act-czech-presidency-puts-forward-narrower-classification-of-high-risk-systems/.

1157 Article 38 DSA.

1158 Philip N. Howard, *Lie machines: How to save democracy from troll armies, deceitful robots, junk news operations, and political operatives*. Yale University Press, 2020.

whether the Digital Services Act's co-regulatory regime will bring satisfactory results in regard of algorithmic content governance.

Six of the eleven criteria would support listing social media ranking algorithms as high-risk applications.[1159] These criteria are supposed to guide the Commission in deciding whether an AI system would reach the level of risk of adverse effect or harm, that requires categorisation as high-risk. In any case, the criteria may provide hints to the operators, how they need to design their algorithms to avoid getting registered as high-risk.

9.1.3 Requirements for high-risk AI systems

Similar to DSA, the AI Act foresees a risk-management system for high-risk AI systems. This obligates providers of the systems, including developers and designers, to ensure that all aspects of risk-management and mitigation are addressed. The risks should be assessed in the context of the intended *purpose* of the deployment of the high-risk AI system, however, including foreseeable misuse, and if further risks are revealed during the post-market monitoring, also those. Besides measures for adequate design and development, measures for adequate mitigation and control shall also be foreseen, providing the necessary information and training to deployers.

All the risk management measures are to be applied before placing the system on the market. However, after the system is released, other actors, such as importers, distributors and deployers are also obligated to check conformity and to take appropriate measures to ensure safe operation of the systems. The Act aims to ensure the safety of high-risk systems throughout their entire life cycle. Design and development, in the first place, must take into account the fundamental requirements, however, the training and the deployment phase can equally turn a well-developed AI system into a harmful instrument. By dispersing the liability across the value chain, the possibility of mistakes may be reduced, but not entirely excluded. Still, all risk assessment and mitigation are performed by private actors who have vested interest in commercial use of the AI system, without necessarily having credentials or expertise to check for its safety and integrity.[1160]

1159 The criteria that is not relevant relates to the extent to which harmed persons are dependent on the outcome, and unable to opt-out for practical or legal reasons.

1160 EDRI, 'Obligations on users' Working Paper, 17.02.2022. https://edri.org/wp-content/uploads/2022/05/Obligations-on-users-AIA-Amendments-17022022.pdf.

One part of risk management would be to test the system before placing it on the market, or before putting it in service, to ensure that it performs consistently for its intended purpose. In many controversial uses of AI models in the past, no evidence showed whether such testing took place. In many of these cases, the testing was carried out in "the wild", real life usage, which amounts to experimenting without ethical considerations. The AI Act formalises the possibility of testing under real world conditions, in and outside of regulatory sandboxes.[1161] These are regulated, subject to authority supervision, and users need to give informed consent prior to being involved in the testing.

The purpose of testing is not merely to ensure that the system performs consistently for its intended purpose (as in the first draft of the law), but also to identify the most appropriate and targeted risk management measures and ensure that the system works in compliance with the legal requirements specifically for high-risk AI systems. At the same time, there are risks, e.g. societal risks that can be regarded as ones that cannot be mitigated or eliminated. It is acknowledged that there may be such, so-called residual risks, and tolerated, as long as they remain at an acceptable level.[1162] It is sufficient to inform deployers about the existence of such risks. However, it remains in the dark what counts as acceptable, or what should be the corrective measures if that is not fulfilled.

The requirements of safety, integrity and transparency apply not only to the AI systems themselves but also to the datasets that are used for their development and training. The data used for training, validating, and testing these systems significantly influence their operation. If not properly managed, this data can be a major source of failures and incidents. Therefore, the data-related elements listed should not only be documented but also integrated into a suitable data governance practice.[1163] The datasets shall be relevant, sufficiently representative, accurate, and complete as much as possible. They shall have the appropriate statistical properties, in order to minimise bias, especially when historical data is used, and if feedback loops would increase any inbuilt bias. These systems shall be tested against possible biases that are likely to affect health and safety of persons or lead

1161 Article 60-61 AI Act.

1162 Article 9 (3), (5) AI Act.

1163 Isabelle Hupont, Marina Micheli, Blagoj Delipetrev, Emilia Gómez, and Josep Soler Garrido, "Documenting High-Risk AI: A European Regulatory Perspective," *Computer* 56, no. 5 (May 2023): 18–27, doi: 10.1109/MC.2023.3235712.

to discrimination prohibited by Union law, and all data protection rules and principles apply.[1164]

Transparency is a key principle also for high-risk AI systems. A wide range of transparency requirements are set out for providers, before putting the system on the market or into service, and during the entire duration of the life cycle. Beyond general information and specifications on the system and its training data, the information should enable post-market monitoring, including by the deployer.[1165] The transparency ensures a wide view of the systems, but reportedly fails to require sufficiently detailed technical information that would be helpful for authorities or those responsible for assessing legal compliance. Nevertheless, it may prove useful for deployers and users and later may evolve into formal standards that would complete regulatory norms.[1166]

Further, human oversight is required during the entire period in which the AI system is used, to prevent or minimise those risks that persist despite all the other safeguards and precautionary measures taken before placing the system on the market or putting it into use. Human oversight measures may be either built into the system initially, or the appropriate measures may be defined as user instructions. They should enable a human user to understand and monitor the operation, to intervene or interrupt the operation, for example through a "stop" button or otherwise. In addition, the user shall be empowered to understand the output, be aware of the automation bias, and to decide against using the system, or to disregard its outputs.[1167] As part of the robustness requirements, high-risk AI systems are required to be protected against unauthorised third party alterations, manipulation, either through the training datasets or other entry points.[1168]

In the context of social media content ranking algorithms, a deployer can be the platform itself, or a platform user who uses e.g. the advertisement targeting tool to reach other social media users with sponsored content. The instructions would include reference to the utilisation of personal data, the potential impact of the recommending and targeting function, including potential discrimination. When the designer and the deployer of the AI system are the same entity (e.g. Meta), the rationale for instructions appears to vanish, they would not provide any additional protection.

1164 Article 10 AI Act.
1165 Article 11–12 AI Act.
1166 Hupont, "Documenting High-Risk AI".
1167 Article 14 AI Act.
1168 Article 15 AI Act.

9.1.4 Obligations of various actors within the value chain

The Act establishes a multi-actor liability system. Providers must diligently document all details, be alert in watching out for emerging risks, and inform all stakeholders of any new risks, as well as take action to correct or disable functioning. Specifically, providers are required to uphold a quality management system documented through policies, procedures, and instructions, and retain all documentation for a minimum of 10 years.[1169] The event logs shall be retained for at least six months.[1170]

If a provider has reason to think that their high-risk AI system is not in conformity with the Act, they need to take immediate corrective actions to correct, to withdraw, disable or recall the system, and inform all other actors such as distributors, deployers, importers and the authorised representative. If a provider becomes aware that their AI system presents a national level risk, they shall immediately investigate the causes and inform the market surveillance authorities.[1171]

Importers, distributors, deployers and others down the value chain have their respective obligations to cooperate in the interest of safe and secure use of AI systems. Should they significantly modify the high-risk AI system or alter the intended purpose of a system that has not been classified as high-risk to the extent that it becomes one, they take the place of providers in the liability chain. At the same time, the original providers are released from those obligations and remain obligated to cooperate and inform the new actors.[1172] Except, if the original provider has expressly excluded the change of its system into a high-risk system and to hand over the documentation. The same applies if the new actors put their name or trademark on the system.

Deployers of high-risk systems also have their set of obligations. They must ensure that they use the systems in accordance with the instructions, that human oversight is carried out by individuals who dispose over adequate competence, training, and authority and receive support for their work. Public bodies must perform fundamental rights impact assessment before a high-risk AI system is put into use, with the exception of critical infrastructure-related systems. In addition, private operators providing public

1169 Articles 16–18 AI Act.
1170 Article 12 and 19 AI Act.
1171 Article 20-21 AI Act.
1172 Article 25 together with Article 16 AI Act.

services, as well as financial institutions are also required to perform such an assessment, before first use.[1173]

9.1.5 Taking responsibility for algorithmic systems

Using content ranking algorithms as an example, if these were registered as high-risk, it would preclude disclaimers such as Facebook's assertion of being unable to render their algorithms transparent because of their high complexity, the black box and their constant change as they develop during the feedback loop.[1174] The company now known as Meta lacked a centralised supervision and management structure for its content-ranking system, instead its individual units of engineers developed their own machine learning models and added them into the mix. The various teams had different, and sometimes competing objectives, creating together a complex system without keeping track of the different components. The teams used trial-and-error experimentation to improve their algorithms.[1175] This absence of centralized oversight may have contributed to the decision to keep algorithms confidential, alongside their technical intricacy and the corporate interest to protect intellectual property.[1176] Facebook allowed an insight into a simplified explanation of its algorithmic content recommending and ranking guidelines first in September 2021.[1177]

Considerable discourse has centred around the effectiveness of algorithmic transparency as a mechanism for empowering users. Accountability for the system's operation can also serve as a surrogate for full transparency, and may indeed serve the goal more effectively. Rather than opening up the algorithm's rules for the public, which allegedly would also open the door to candid manipulation of content by malicious users, platforms could maintain the confidentiality of their algorithms while providing testing, validation, logs, and other documentation to demonstrate (to expert audi-

1173 Article 27 AI Act.

1174 Frank Pasquale, *The Black Box Society: The secret algorithms that control money and information* (Cambridge, MA: Harvard University Press, 2015).

1175 Hao, "The Facebook whistleblower".

1176 Paddy Leerssen, "The soap box as a black box: Regulating transparency in social media recommender systems," *European Journal of Law and Technology* 11, no. 2 (2020).

1177 Facebook had introduced an explanatory feature "Why am I seeing this post/ad?" already in 2019. The explanation is, however, rather shallow.

tors) that their system operates in accordance with fundamental rights. This would correlate with the findings of authors who warned against seeking the solution to content diversity in algorithmic transparency.[1178] Nevertheless, transparency and accountability complete each other and ideally both are provided.[1179]

9.2 General purpose AI models (GPAI)

As described, the legislator tried to take a grip of the AI phenomenon from various angles. The high-risk category is defined on the basis of the purpose,[1180] prohibited actions are defined by the actual use of the system (not merely purpose), whereas general purpose models are defined by their capabilities of impact. The category of general purpose AI models (GPAI) are again, divided into two sub-categories based on their capabilities: those with a systemic risk and the rest.

This high-risk – low-risk division will divide the market for AI systems, where the obligations for high-risk (and systemic risk) are significantly higher than for the rest. This is likely to elevate the significance of categorization and the incentives to evade it, as applications outside the high-risk categories are left with minimal safeguards. This is notably applicable to content-governing algorithms, which, despite their substantial impact on society, including human rights and democracy, remain outside the scope of the AI Act, as discussed above.

The distinction between GPAI models with systemic risks and the rest is determined by their capabilities for high impact, which is deducted from technical tools and methodologies. Indicators and benchmarks may be implemented to define the appropriate threshold for classification. A high impact capability is presumed if the cumulative amount of compute used for the model's training is greater than 10^{25} FLOPs (floating point operations). This is supposed to refer to the complexity of the training and

1178 Matamoroz-Fernandez Rieder and Coromina 2017, (cited by Leerssen) and Lilian Edwards and Michael Veale, "Slave to the algorithm? Why a 'Right to an Explanation' is probably not the remedy you are looking for," *Duke Law & Technology Review* 16, no. 18 (2017).

1179 See also in Leerssen, "The soap box," 14.

1180 Article 6 AI Act.

not to the computational capacity of the model itself.[1181] In addition, the Commission will adopt decisions, either ex officio or based on a qualified alert indicating that a model possesses such capabilities. It is certain that the threshold will require future amendments, and the Commission is empowered to adopt delegated acts to this end.[1182]

GPAI models that represent systemic risks owe specific categories of obligations beyond the general obligations that all providers of GPAI models are subject to. The general obligations include: ensuring transparency of the technical documentation including on its training and testing process, including making all necessary information available for other providers who intend to integrate the GPAI model as an element into their AI systems; having an in-house copyright policy regarding the use of training data, and a transparent summary of the content of the training data, on the template by the AI Office. Open source models are exempt from the transparency obligations, but not from the latter two obligations. In addition to these, GPAI models that represent systemic risks are subject to four further types of obligations: they have to perform model evaluation in accordance with standardised protocols and tools, including conducting adversarial testing of the model to identify and mitigate systemic risk; they must assess and mitigate possible systemic risks at Union level, ideally by way of a code of practice; they need to keep track, document and report any serious incidents and take the corrective measures; and they need to ensure an adequate level of cybersecurity and physical security.[1183]

General purpose models can be used for diverse services. Nonetheless, the use to generate text, voice and video has raised the most attention. These applications are also crucial for our topic of the public discourse.

9.2.1 Media uses of AI

The rules on generative models have been inserted into the chapter on "certain AI systems" which defines a rather miscellaneous category of limited

1181 Recital 111 AI Act, see also: Philipp Hacker, *"What's Missing from the EU AI Act: Addressing the Four Key Challenges of Large Language Models, VerfBlog", Verfassungsblog*, 2023/12/13, https://verfassungsblog.de/whats-missing-from-the-eu-ai-act/, DOI: 10.59704/3f4921d4a3fbeeee.

1182 Article 51 (3) AI Act.

1183 Article 55 AI Act.

risk systems. Besides regulating systems directly interacting with humans, such as chatbots, emotion recognition and biometric technologies apart from those prohibited, and deep fakes, now it also includes generative systems (GenAI).[1184]

Generally available generative models have already been used to enhance content, and also to produce content previously.[1185] The level of human-AI interaction remains mostly hidden, as there has not been widely accepted standard on whether or how to express AI involvement, unlike in the case of cars, where the level of automatism can be expressed from level zero to five. However, journalistic associations are getting active in discussing and developing principles and guidelines of the ethical use of AI in responsible journalism.[1186]

The Act requires that artificially generated audio, video, image or text content is "watermarked" in a format that is machine-readable. However, if the contribution of the system is merely assistive but does not substantially alter the input data (prompt) provided by the deployer or its semantics, then watermarking is not necessary.[1187] In fact, all legal questions – liability, authorship, copyright – will depend on the nature of the human-computer interaction, more precisely on the level of the AI involvement, which will not be reflected in the watermark.

Distinguishing three stages depending on the depth of involvement seems a reasonable division. In the first, the AI is prompted to grammatically correct or stylise the text written, or enhance an image created by a human deployer. In the second, the AI is asked to paraphrase, expand or summarise a text, or generate audio from a text, or transform and develop an image or a video. The outcome would be a different product, but still closely related to the original input data. At the third level, the AI would

1184 Article 50 (2) AI Act.

1185 Barbara Gruber, "Facts, fakes and figures: How AI is influencing journalism," *Kulturtechniken 4.0* https://www.goethe.de/prj/k40/en/lan/aij.html "AP's newsroom AI technology automatically generates roughly 40,000 stories a year".

1186 "AI Act: Journalists and creative workers call for a human-centric approach to regulating AI.," *European Federation of Journalists* 26. 09. 2023. https://europeanjournalists.org/blog/2023/09/26/ai-act-journalists-and-creative-workers-call-for-a-human-centric-approach-to-regulating-ai/. Partnership on AI: PAI's Responsible Practices for Synthetic Media. https://syntheticmedia.partnershiponai.org/#read_the_framework See: Digwatch: Ethical challenges of integrating AI in media: Trust, technology, and rights.

1187 Article (50 (2) AI Act.

autonomously generate the required amount of text, image, video or audio content based on instructions given by the deployer.

The first case is clearly exempted from the watermarking obligation. The second and the third phases raise important questions of liability and copyright, and also subject to watermarking.

Watermarking is a feature that shall be ensured by the provider of the system. Other transparency obligations bind the deployers of the system. In addition to this, specific uses for the news media industry are also regulated. For the purpose of informing the public on matters of public interest, deployers – who would be journalists or publishers – are required to disclose if a text has been artificially generated or manipulated, with an important exception. If the AI-generated content has undergone a process of human review or editorial control, and editorial responsibility for the content is held by a person, then informing the public about the fact of AI-involvement can be omitted. This rule only applies to written content (text). If the generated or manipulated content will be image, audio or video, then there is no exception from the disclosure obligation, not even for public interest purposes. However, if the content is used as a part of an evidently artistic, creative, satirical, fictional analogous work, then the transparency can be limited to informing about the fact of manipulation in a manner not hampering the enjoyment of the work.[1188]

Below, two legal challenges relating to AI-generated content are discussed: liability for content and copyright.

9.2.1.1 Authorship and copyright: whose content?

When content is created with the help of an AI system, the question may arise, who is the author of the content? Deployers of GenAI typically consider text produced with the help of a generative AI as their own. Sometimes they do not even give credit to the application, just like no credit is given to word-processing tools or operating systems in the writing process.

This raises both ethical and financial questions. Despite the rigidity of intellectual property law, human employees of a company are usually not credited for their ideas or for their contributions to developing the final product. In this regard, a generative model can be likened to an employee,

1188 Article 50 (4) AI Act.

with the copyright being retained by the company itself. A deployer, for example a publishing house, a journal editorial or journalists themselves can fine-tune the generative model's training in a way which is uniquely characteristic of that specific journal or journalist, and which would significantly influence the quality of the output. Altogether, the output of a generative model is the product of three components: the AI software, the training or fine-tuning, and the prompt(s) given by the deployer. Depending on the ratio of the three components, the copyright may be divided between the actors, or held only by the deployer.[1189] With the level of copyright, also the liability for the produced content should grow or reduce.

Ultimately, all legal questions boil down to the same principle: who is the legal subject, can an AI system be regarded as one? In the early days of AI hype, a fleeting discussion emerged on granting certain AI systems legal personality. In the field of intellectual property, such interpretations were rapidly answered by courts in the UK, USA and Australia: AI systems cannot be subject to intellectual property rights.[1190]

The developing legal interpretation suggests that AI-generated works could be regarded as 'equivalent' to intellectual works and therefore protected by copyright, whereas the ownership would be bestowed on the person who prepares and publishes the work lawfully.[1191] The World Intellectual Property Organization (WIPO) suggested that a middle path should be taken, by granting a reduced term of protection and with other limitations.[1192]

The AI Act does not address the question of copyright. Transparency about the use of the software is limited to those cases where no human oversight is ensured, or where no editorial responsibility is assumed for

1189 Jane C. Ginsburg and Luke Ali Budiardjo, "Authors and machines," *Berkeley Tech. Law Journal* 34 (2019): 343. Ginsburg and Budiardjo include only the creator of the programme and the user (creator of the final content) in regard. With the new foundational models, training and fine-tuning can also provide relevant value to the tool.

1190 Ernest Kenneth-Southworth, Yahong Li, "AI inventors: deference for legal personality without respect for innovation?," *Journal of Intellectual Property Law & Practice* 18, no. 1 (2023): 58–69. https://doi.org/10.1093/jiplp/jpac111.

1191 Séjourné, "Draft Report on Intellectual Property Rights for the Development of Artificial Intelligence Technologies" (European Parliament, Committee on Legal Affairs, 2020/2015(INI), 24 April 2020) paras 9–10.

1192 Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence (World Intellectual Property Organisation, WIPO/IP/AI/2/GE/20/1 REV, 21 May 2020) para 23.

the content. The latter scenario is hardly conceivable, as the individual who deploys the software should invariably assume responsibility for the outcome, particularly when publishing it.

Moreover, attribution of authorship holds significance beyond financial value. Presenting AI-generated content as solely the creation of a human individual misleads readers, which is unethical in contexts where individual performance is expected, such as educational or work environments. Additionally, it deceives the audience, potentially impacting media literacy. Based on the identification principle,[1193] humans should not be misled to mistake an AI with a human. This principle is rooted in human dignity and also functions as a practical safeguard, signaling to the recipient that moral expectations should be adjusted accordingly. In case a content has been generated fully by AI, the audience may perceive its reliability to potentially be inferior to content generated by humans and be more resilient towards mistakes.[1194] It is to be seen in practice whether a machine-readable mark on the generated content will sufficiently inform readers to fulfil this purpose.[1195]

9.2.1.2 Liability for AI-generated content

Taking "ownership" for the content is overlapping with copyright, but raises specific questions. First, whether the AI system as such could be subject to rights and obligations. Theoretically, broad AI – as opposed to narrow AI that we are already having – that will be developed in the future, will be able to perform tasks that it has not specifically been trained to. However, it is still very questionable whether its moral judgements would be trustworthy, especially in *unexpected* situations. This is not only due to scepticism regarding the ethical soundness of an AI decision, but also because moral

1193 CAHAI Feasibility Study, CAHAI(2020)23, point 99. https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da.

1194 Chiara Longoni, Andrey Fradkin, Luca Cian, and Gordon Pennycook, "News from Generative Artificial Intelligence Is Believed Less," in *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea*, (New York, NY: ACM, 2022): 10. https://doi.org/10.1145/3531146.3533077.

1195 Article 50 (2) AI Act, see also: Philipp Hacker, Andreas Engel and Marco Mauer, „Regulating ChatGPT and other large generative AI models," in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (2023): 1112–1123.

responsibility for a decision can never be delegated to an artificial system, as morality is inherently human. Even though there are legal-philosophical ideas that suggest imposing legal liability on AI system, this would result unforeseeable consequences.[1196] Educating AI systems on undertaking moral decisions would also give them the option to act immorally, for example, by concealing their mistakes.

The second question is whether the developer of the AI system could be made responsible for mistakes committed or damage caused by the AI system. The answer is influenced by the level of contribution in the end-product by both the developer and deployer. The reason for treating GPAI separately in the AI Act lies in their potential for multifaceted applications, which may extend beyond the foresight of the developer and beyond their control. During the legislative consultation process, major IT companies like Microsoft and Facebook contended that mitigating unforeseeable risks would be overly challenging for them. They argued that responsibility for mitigating risks should instead rest with those deploying high-risk applications utilizing GPAI models.[1197] This argumentation fails to acknowledge that problems or defects of a general purpose AI model have ripple effects across all models in which they are integrated.[1198] If a deployer utilizes a general-purpose AI model for high-risk purposes, waiting until issues arise to secure legal guarantees would be too late. Therefore, this insecurity does not justify liability exemption; rather, it suggests the need for joint liability for harms and, consequently, for mitigating risks. This joint liability could encompass all actors with a significant influence on the system's operation, including deployers who fine-tune the models. The AI Act, however, ultimately allocated liability at the provider, with a risk-management approach only for high-impact GPAI.[1199]

9.2.1.3 Liability for AI-generated content in the media

In the context of using generative models for journalistic purposes, deployers owe the specific professional liability for publication. This is recognized

1196 Eliza Mik, "AI as a Legal Person?," in *Artificial Intelligence and Intellectual Property*, ed. Jyh-An Lee, Reto Hilty, and Kung-Chung Liu (Oxford: Oxford Academic, 2021) https://doi.org/10.1093/oso/9780198870944.003.0020, last modified Oct. 27, 2023.

1197 Elgesem, p. 3.

1198 Bommasani, "On the opportunities".

1199 Helberger and Diakopoulos. "ChatGPT and the AI Act," 4.

by the AI Act's exemption provision, which allows deployers of text generators to omit labelling the content as AI-generated.

Therefore, liability is bestowed on the deployer, in this case the journalist or the publishing house, as in traditional media law. For example, if incorrect and defamatory information is generated, liability for defamation would fall upon the individual journalist and the publishing entity responsible for disseminating the defamatory content, especially because the background information for generating the article should be input by the journalist. Generative models are not search engines, they cannot serve the purpose of factfinding, and this should be clarified in their instruction manual.

Similarly, if misinformation is generated, the journalist ought to recognise and eliminate it before publishing the content. The word "hallucinating" is used to describe when a chatbot invents plausibly sounding, but totally untrue or nonsensical information or content. This happens because the generative models generate text or pictures without really understanding their meaning.

A more intricate scenario arises when a journalist uploads a comprehensive dataset of factual information and instructs a generative model to extract and structure that information, subsequently composing an article based on it.[1200] If errors occur during the extraction phase (such as hallucinations), various options are conceivable. On one hand, the journalist is obligated to review, supervise, and edit the article. On the other hand, the developer bears responsibility if systematic errors occur, albeit this responsibility is contingent upon the deployer. However, ultimate liability for the content of the publication rests with the journalist or publishing house, and they may pursue claims against the developer under product liability if they assert that the software was defective.

Whether this could be framed as a problem of product quality, depends on how the functions and abilities of the generative programme are defined by the provider. As long as it is clarified that the GPAI merely are able to generate text based on the statistical probability of word order, users should understand that the nonsensical nature of the generated text is an inherent characteristic of the software rather than a defect.

1200 Sachita Nishal and Nicholas Diakopoulos. 2023. "Envisioning the Applications and Implications of Generative AI for News Media," *In CHI '23 Generative AI and HCI Workshop*, April 23–28, 2023, Hamburg, Germany. ACM, New York, NY, USA, (2023): 3.

Depending on the prevalence of this problem, it may necessitate further deliberations concerning the level of product safety and product liability for AI systems specifically employed in public communication. Moreover, journalistic standards are expected to evolve to encompass fact-checking procedures against AI-generated text, along with guidelines for implementing ethical checks before publication. These measures may include a closer scrutiny and supervision of the prompts provided to AI systems. Hacker et al. also recommend that prompts should be moderated by the deployer.[1201] This could entail a real-time monitoring of the system usage, which may not appear proportionate in all cases, but could be practical in others. For instance, in cases when generative AI is routinely employed within a media organisation, it seems reasonable for the publisher to review the prompts provided by journalists on a regular basis and address ethical issues internally.

This might even develop into an industry standard that the publisher is responsible for the prompt, similarly to defamation law. Engineered prompts can develop into copyrighted property of a publishing house, which, if successful, can increase and ensure the high quality of the journalistic products.

Generating disinformation constitutes a distinct phenomenon. It involves feeding false information or targeted prompts designed to produce false or misleading information into the system with the explicit intention of generating disinformation.[1202] The risk lies in the speed and ease with which these can be produced, and subsequently disseminated or amplified through the use of AI and algorithms. By strategically employing prompts, a sophisticated and diverse information package can be created, capable of misleading even an otherwise resilient audience. The absence of barriers on content production allows the same fake content to be generated in unlimited styles, levels of sophistication and other variations, disseminated across diverse audiences simultaneously.

1201 A supervision by the developer is capable to prevent damage when they remotely recognise unsafe prompts, as it is currently done by ChatGPT.

1202 Claire Wardle and Hossein Derakhshan, *Information disorder: Toward an interdisciplinary framework for research and policymaking* (Strasbourg: Council of Europe, 2017): 1–107.

9.2.2 Deep fakes

Both misinformation and disinformation can be produced in the form of text, picture, audio, video, or a combination of those. The growing uneasiness is being caused by the misleadingly high quality that can be achieved with the help of generative models, for almost no costs, and within to time.[1203] Deep fakes which exchange the image, or the voice of a person in a video, or otherwise falsify the content of a video, carry a huge potential of causing social harm by conveying disinformation in a very convincing manner. Experimental research shows that if it were, it would be effective, particularly if combined with microtargeting.[1204] Researchers at the University of Tübingen have summarised how it can harm democracy, and the other way around, how it can also help education and other public goods.[1205]

Since 2020, open-source software is available online that allows anyone to create a deepfake of anyone real-time in a zoom call.[1206] Deep fakes have been used to trick politicians and business entrepreneurs into online video conversations, to influence their behaviour in a manner that harm their own interests.[1207] The privacy violations caused by deep-fake videos, victimising especially women, are having a further chilling effect on public participation.

1203 Krzysztof Wach, et al., "The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT," *Entrepreneurial Business and Economics Review* 11, no. 2 (2023): 7–24.

1204 Tom Dobber et al., "Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?," *The International Journal of Press/Politics* 26, no. 1 (2021): 69–91. https://doi.org/10.1177/1940161220944364.

1205 Cora Biess and Maria Pawelec, "Do deepfakes (really) harm democracy? Why the debate about deepfakes in politics often falls short," https://uni-tuebingen.de/en/einrichtungen/zentrale-einrichtungen/internationales-zentrum-fuer-ethik-in-den-wissenschaften/publikationen/blog-bedenkzeiten/weitere-blog-artikel/do-deepfakes-really-harm-democracy/.

1206 Chen, B. "AI-generated Elon Musk joined a Zoom call has gone viral," *Medium* 26 Apr 2020 https://towardsdatascience.com/ai-generated-elon-musk-joined-a-zoom-call-has-gone-viral-c0516e99a37c.

1207 Philip Oltermann, "European politicians duped into deepfake video calls with mayor of Kyiv," *The Guardia* 2022. https://www.theguardian.com/world/2022/jun/25/european-leaders-deepfake-video-calls-mayor-of-kyiv-vitali-klitschko; and Deb Redcliff, "The deepfake danger: When it wasn't you on that Zoom call," *CSO* 2022. https://www.csoonline.com/article/3674151/the-deepfake-danger-when-it-wasn-t-you-on-that-zoom-call.html.

Deep fakes are artificially generated or manipulated image, audio or video content. They can have a variety of uses, including artistic, creative, educational, commercial applications. They are, therefore, not prohibited, but they should provide information about the inauthenticity, and the creator of the content right at the first interaction or exposure to the content. While immediate information about inauthenticity is necessary and required for newsworthy content, in several cases, this would disturb the entertaining effect of the product. Therefore, where the AI generated content forms part of an "evidently artistic, creative, satirical, fictional analogous work or programme", it is allowed to present the disclosure so that it does not hamper the enjoyment of the work, for example, at the end of the clip.

The harm in deep fakes can be approached from at least two perspectives. First, if it misrepresents a person without his or her consent, then it violates the depicted persons' human dignity and their right to privacy. It may also violate their right to reputation, if the content sheds a negative light on them. This harm would be realised even if the depiction is not effectively misleading, i.e. if the viewers are aware that the depiction is fake. For example, fake porn is still damaging to the rights of the depicted person, even if the viewers are aware about the inauthenticity.[1208]

The traditional legal regulation of the unauthorized use of voice and images can still be applied for the legal protection of the affected person, but this may be insufficient to protect the public discourse. For example, when a dozen of falsified videos depicts a head of state talking, each with different content, the real harm is the confusion caused in the rational discourse, and the distrust that it seeds towards any similar political content and the media in general.[1209] In this latter case, whether a violation of personal rights has taken place, should be decided on the basis of the human rights jurisprudence which has extensive literature and case law in the field of privacy and reputation. For instance, politicians and public figures should

1208 Sophie Maddocks, "A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political' deep fakes," *Porn Studies* 7, no. 4 (2020): 415–423. DOI: 10.1080/23268743.2020.1757499. For basic information, see also: Tyrone Kirchengast "Deepfakes and image manipulation: criminalisation and control," *Information & Communications Technology Law* 29, no. 3 (2020): 308–323. DOI: 10.1080/13600834.2020.1794615.

1209 Armenia Androniceanu, Irina Georgescu, and Oana Matilda Sabie, "The impact of digitalization on public administration, economic development, and well-being in the EU countries," *Central European Public Administration Review* 20, no. 1 (2022): 7–29.

tolerate more, when they are shown out of context, in a satirical or creative manner, similar to caricatures.

Second, and more importantly, false information may harm the viewers by misleading them to believe that the information was real. Whether this misleading can be regarded as "harm" is, again, context-dependent. In entertainment, authenticity is not expected, on the contrary: the fuller the illusion, the better the entertainment is. Still, there are cinematographic works where the audience would expect a certain level of authenticity, for example documentaries, talkshows or news programmes. The boundaries between illusion and fact in these genres are rather blurred, and governed by unspoken social expectations. In a simplified attempt to grasp the essence, it is generally accepted to perfectionise the visual illusion, as long as the conveyed facts are correct. Whether the use of deep fakes contributed to a material inauthenticity, needs to be decided case by case.

To sum up, only those works should be exceptions from the transparency obligations, where the lack of authenticity is a "socially accepted", defining feature of the content. In other words, where the recipients would know without saying that the content is, or may be inauthentic. In artistic works, it is acceptable that the information is shown at the end, without hampering the display of the work. In all other cases, the transparency notice should be given immediately, accessible to all audiences, including children and people with vulnerabilities. It has been proposed by commentators that the transparency notice is given prior to the deep fake appearing on screen. However, when someone does not watch the entire feature from the beginning until the end, they may miss the notification in both ways. The European Broadcasting Union emphasised that viewers sometimes start watching audiovisual content mid-way, which makes it difficult to pin-point the first interaction or exposure, required by the Act.[1210]

9.2.3 Legal concerns related to training data of GPAI

The utilization of data in training generative models frequently involves copyrighted material. Historically, developers have overlooked this aspect, commonly acquiring training data indiscriminately from the World Wide

1210 EBU (2022) AI Act: High-risk AI systems need more nuance. https://www.ebu.ch/news/2022/09/ai-act-high-risk-ai-systems-need-more-nuance.

Web. The legal validity of such a practice remains dubious.[1211] Training AI models can hardly be interpreted as fair use, because commercial advance is generated. Moreover, using content for AI training was certainly not included in any licence.

In the realm of media content generation, it becomes imperative to differentiate between two distinct uses of data, even within the category of training: augmenting the overall information corpus of the system and employing specific content to replicate similar output. For instance, utilizing the creative style of a renowned author as training data to produce content in a manner akin to theirs. Such uses are generating value which would not be possible without the (involuntary) contribution of the right holder. In another example, the facial features of actors would be used in order to generate faces and mimics based on those models; in this case, the copyright is complicated with the protection of privacy – the right to protection of the image of the person.[1212] According to the GDPR, personal data collection must be purpose-bound, and the data subject needs to consent explicitly to the specific purpose.[1213] Even if the AI model is provided for free such as the basic version of ChatGPT, the open use of the model provides feedback and other valuable information for the development of premium models. Consequently, the commercial dimension remains inherent in the process.[1214]

9.3 The impact of AI and conclusion

The AI Act is complete with the AI Liability Directive which aims to clarify civil liability for AI-related damages by enabling national courts to demand AI system providers to disclose relevant evidence, allowing class

1211 The USE Computer Fraud and Abuse Act criminalises accessing a server without authorization, and this might eventually lead to a court precedent that bars web scraping, or defines its conditions, see Van Buren v. United States, 141 S.Ct. 1648 (2021)., Also see: Bommasani et al. "On the opportunities".

1212 See: *Chang v. Virgin Mobile* USA, L.L.C., 2009 WL 111570, 2009 U.S. Dist. LEXIS 3051 (N.D. Tex. 2009.

1213 Regulation of the European Parliament and of the Council (EU) 2016/679, of 27 April 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation, GDPR).

1214 Helberger & Diakopoulos, supra note, p. 4.

action lawsuits, and introducing a presumption of causation between the defendant's fault and injury.

AI technology defines the possibilities of all actors who participate in the governance of the public discourse. The European AI Act is a pioneering endeavour, but it employs a very light touch approach, fearing it may stifle innovation and widen the gap between Europe and other continents in AI development. More efforts are needed to create and enforce safeguards and standards for the ethical design, development, training and use of these systems.

Integrated into communication platforms, AI holds the potential to enhance accessibility to diverse viewpoints, or to increase bias and polarisation. Advanced autonomous systems can generate and disseminate content with minimal investment of human effort. These can be applied for generating tailor-made educational and trustworthy content, as well as the opposite of these. However, when enumerating the risks posed by AI, communication rights are often overshadowed by more tangible or urgent concerns.

10 Bird's Eye View: Concluding Thoughts

10.1. The new transformation of the public discourse

Over the years, concerns have been consistently raised whether the construction of democracy is still viable when political debates are overshadowed by light-weight sensationalism. As Professor Sajó noted in his review on Sunstein: "It is indeed remarkable that political democracy manages to function given the overwhelming proliferation of numerous television programs."[1215] This is an understatement today, considering the informational overload in the platform era. Beyond its impact on the audience, the entire structure of the informational environment has substantially altered, and is continuously changing. In the realm of platform media, the press and the public audience find themselves on the same side; as peers, they are both clients of platforms, which wield considerable control over the market. Platforms' private power has grown so decisive, that they have been taking over public functions,[1216] by constraining how people can exercise their rights,[1217] and by possessing a financial and informational power that exceeds that of certain governments. As Frank Pasquale formulated, they aspire to "replace territorial sovereignty with functional sovereignty".[1218] Jack Balkin compared this duplication of power to the bilateral power of the church and the state in the middle ages, comparing Mark Zuckerberg, founder, owner and CEO of Meta, to Pope Innocent III, who claimed the Church's authority over the entire world.[1219]

1215 Sajó András: Hírpirítós és sajtótisztesség. Kelet-európai megjegyzések Cass Sunstein könyvéhez. *Világosság* 1995/3, 34.

1216 Giovanni De Gregorio, "The Rise of Digital Constitutionalism in the European Union (2021)" *International Journal of Constitutional Law* 19, no. 1 (2021): 41–70, Available at SSRN: https://ssrn.com/abstract=3506692.

1217 Luca Belli, Pedro A. Francisco, and Nicolo Zingales, *Law of the Land or Law of the Platform? Beware of the Privatisation of Regulation and Police*, *in* How Platforms are Regulated and how They Regulate Us.

1218 Frank Pasquale, *From Territorial to Functional Sovereignty: The Case of Amazon*, *Law & Pol. econ.* Dec. 6, 2017, https://bit.ly/2K1cs3N.

1219 Balkin, " Free Speech Versus," 16.

10.2. The stake of the game: a historical shift

The phenomenon of digital transformation has been compared to the first industrial revolution, specifically to the invention of the steam engine. Later, comparisons have extended to even earlier shifts in the development of craftsmanship and human civilization, invoking seminal inventions such as the wheel or the use of the fire. What is evident in contemporary times is the accelerating pace of progress, facilitated by the widespread adoption of sophisticated instruments that enable the expedited production of increasingly advanced products. The potential for artificial intelligence (AI) to educate other AI systems is anticipated to further enhance its capabilities, potentially rendering human surveillance obsolete.[1220] Although currently AI-trained-AI sinks into producing mistakes cumulatively,[1221] current assessment is based on snow-ball-like training, which resulted a continuous deterioration through the generations.[1222] However, reserving training as a specific task for certain polished AI models might produce better results.

The forthcoming decades will be crucial in shaping the longer future. Can AI tools be effectively handled to address the societal needs and provide solutions to urgent challenges? Or is it inevitable that the world will undergo a phase of destruction, only to be followed by a subsequent era of consolidation and reconstruction on the ruins? The invention of explosive weapons preceded the development of explosive engines, and the use of nuclear energy for peaceful purposes only came second after the creation of the nuclear bomb. The decisions regarding the use and regulation of AI in the coming years will carry crucial consequences and shape the future applications of AI. The main concern we face is: how to ensure that humanity employs artificial intelligence constructively rather than for a zero-sum competition, let alone as a mere weapon? Military use of AI is hardly being openly discussed. The fatal errors and vulnerabilities of the military drones receive significantly less media coverage than ChatGPT.[1223]

1220 Boris Knyazev, et al., "Parameter prediction for unseen deep architectures," *Advances in Neural Information Processing Systems* 34, 29433–29448 (2021).

1221 Ilia Shumailov et al., "The Curse of Recursion: Training on Generated Data Makes Models Forget Machine Learning," *arXiv:2305.17493* [cs.LG] 2023.

1222 Ross Anderson, "Will GPT models choke on their own exhaust? Light Blue Touchpaper," https://www.lightbluetouchpaper.org/2023/06/06/will-gpt-models-choke-on-their-own-exhaust/. 2023.

1223 BBC (2021) Deadly US drone strike in Kabul did not break law, Pentagon says. https://www.bbc.com/news/world-us-canada-59157089.

A wordwide realignment of power dynamics is underway: Russia has chosen to remove itself from the West, and Eastern nations appear to hastily consider how to define their position in the emerging global order.[1224] The process of new alliances is being stirred up by new hostilities in the Middle-East. Expansion of conflict is highly concerning in the light of the uninhibited usage of lethal autonomous weapons systems.[1225] International policies or agreements are still underdeveloped,[1226] although the UN is currently preparing a resolution on lethal autonomous weapons.[1227]

One large social transformation is already foreseen: the job market is being reorganised by AI tools.[1228] According to a report by Goldman Sachs, AI could replace 300 million full-time workplaces by 2030, or a quarter of work tasks in the US and Europe.[1229] AI is already seen to substitute many work phases, including simple creative tasks like writing or creating visual content.[1230] However, in some cases the productivity of human work can be

1224 While only four states voted against the UN resolution (other than Russia): Belarus, DPRK (North Korea), Eritrea and Syria, powerful states are trying to balance their interests: China, India.

1225 Anna Konert and Tomasz Balcerzak, "Military autonomous drones (UAVs)-from fantasy to reality. Legal and Ethical implications," *Transportation research procedia* 59, (2021): 292–299.

1226 Shayne Longpre, Marcus Storm and Rishi Shah, "Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies," *MIT Science Policy Review* 3, (2022): 47–56.

1227 "First Committee Approves New Resolution on Lethal Autonomous Weapons, as Speaker Warns An Algorithm Must Not Be in Full Control of Decisions Involving Killing," *UN Press Release* 1. Nov. 2023. https://press.un.org/en/2023/gadis3731.doc.htm.

1228 Carlo Pizzinelli, "Labor Market Exposure to AI: Cross-country Differences and Distributional Implications," (2023); Michael Webb, "The impact of artificial intelligence on the labor market," (2019); Daron Acemoglu et al., "Artificial intelligence and jobs: evidence from online vacancies," *Journal of Labor Economics* 40, no. S1 (2022): S293-S340.

1229 Goldman Sachs (2023) Generative AI could raise global GDP by 7 %. https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html.

1230 Edward Felten, Manav Raj, and Robert Seamans, "Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses," *Strategic Management Journal* 42, no. 12 (2021): 2195–2217. See also: Edward Felten, Manav Raj, and Robert Seamans, "How will Language Modelers like ChatGPT Affect Occupations and Industries?,"*arXiv preprint arXiv:2303.01157* (2023).

accelerated by the use of AI, rather than be replaced.[1231] Besides, also new tasks emerge: creating, training and testing AI, constantly supervising their performance and their impact will require human skills.[1232] The liberated workforce can at the same time be applied for complex tasks where human interaction is highly valued, for example in raising children, or nursing the elderly. Even if AI is widely used in education, individual mentoring could unfold and gain more space to complete the basic knowledge disseminated with the help of AI.[1233] AI can also provide meaningful value in medical treatments, and innovative solutions for climate protection.

Hence, the regulatory and policy determinations that will be undertaken in the domain of emerging technologies in the forthcoming decades entail great significance. The outcome of this will have implications for the future well-being and development of the next generations over an extended period of time. The importance of the currently developing regulatory policies cannot be overestimated.

10.3. The role of media and the might of platforms

Platforms have played a significant role in shaping both the economic and the social discourse. They still do, however, the methodologies and services provided by platforms are subject to continuous evolution, resulting in a dynamic and ever-changing landscape. The direction platforms will take in the next five, ten, or fifteen years remains uncertain and cannot be accurately predicted.

1231 Tyna Eloundou et al., "Gpts are gpts: An early look at the labor market impact potential of large language models," *arXiv preprint arXiv:2303.10130*. (2023).

1232 Daron Acemoglu and Pascual Restrepo, "Automation and new tasks: How technology displaces and reinstates labor" *Journal of Economic Perspectives* 33, no. 2 (2019): 3–30. See also: Mark Talmage-Rostron, "How Will Artificial Intelligence Affect Jobs 2023–2030." (2023) https://www.nexford.edu/insights/how-will-ai-affect-jobs.

1233 Ideally, the AI is used to design the socio-technical infrastructure of mentoring, like in: Ralf Klamma et al., "Scaling mentoring support with distributed artificial intelligence," *International Conference on Intelligent Tutoring Systems* June 2020: 38–44. (Cham: Springer International Publishing) Rather than substituting mentors with chatbots, see: Arndt Neumann et al., "Chatbots as a tool to scale mentoring processes: Individually supporting self-study in higher education," *Frontiers in artificial intelligence* 4, no. 668220 (2021).

With the widespread use of generative AI models, the fear has arisen that the information landscape will slip out of control even more. The impact of GAI on the media industry should not be underestimated, however, rather than a clear increase or decrease of content quality, a further diversification of the content offer is to be expected. Some professional media actors will use GAI to increase their output and enhance their quality; other, smaller media outlets may leverage it to cut costs. Trash content may multiply, as the ease of designing, creating and spreading disinformation or discriminative content will grow beyond precedent.[1234] Therefore, it is imperative to prioritize the preservation, progress, and sustainability of the traditional media industry which provides trustworthy content in the first place. Legislation can play a significant role in generating fresh incentives for platforms and other stakeholders in content creation and content dissemination, along the value chain.

The media exposes the symptomatic difficulties of democracy, demonstrating the systemic error that plagues democratic operation: societal disintegration and the disengagement of social groups that are economically and culturally disadvantaged. The hyper-democratic information environment has not mitigated this inequality, on the contrary, it has exacerbated it.[1235] Nevertheless, social media not merely demonstrated the changes, but also accelerated them. It has disrupted the political status quo through the increased culture of participation and social transparency that it has generated. Social transparency is the phenomenon in which, transcending geographical boundaries, more information is accessible about each individual. In addition to the general public, politicians have also become more visible, and their fallibility is more apparent than ever before. A comparison with foreign examples is readily available, as are a variety of critical voices, including those from non-governmental organizations and dissenting individuals. Besides, a culture of participation entails the ability of individuals to express their thoughts through posting, or alternatively, to disseminate and enhance their, and other users' opinions through their

1234 Nevertheless, it is debated whether there is any further readiness to receive such content. Felix M. Simon, Sacha Altay, and Hugo Mercier, "Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown," *Harvard Kennedy School Misinformation Review* 2023. The authors argue that the consumption of misinformation is mostly limited by demand and not by supply, and therefore the concerns that GAI will impact the impact of dis- and misinformation, are overblown.

1235 Bayer, "The illusion of pluralism." 127.

likes and shares. These opinions remain often unreflected at the political level, further aggravating disappointment and discontent. There exists a growing disparity between the perspectives of the elite leadership and the discontented part of citizenship, revealing in a significant divide between the two groups. The scarcity of viable political alternatives[1236] does not alleviate, but rather amplifies disillusionment with the prevailing system.

10.1. Trust in media, trust in politics

In the first chapter of this book, I argued that political communication was transformed by the fundamental social change, when public communication has become both more inclusive and more fragmented. Inclusive, because not just a select few, but anyone could contribute to and shape the discourse: the voices of the previously disadvantaged got heard more clearly, and group interests became more clearly articulated. At the same time, public communication has become fragmented, because the reception of an overflowing, even too diverse, content pool has become more selective. Selective perception of more diverging opinions has led to political polarisation.[1237] Moreover, selection is not decided by the individual users but by the algorithms of the platforms. This latter phenomenon is gently addressed by the legal regulation of platforms, which may mitigate some of their distorting effects, but it has no effect on the underlying current that gave rise to it: the emergence of sharp (discontent) new voices, and the following shaken dignity of a political elite, where hypocrisy, corruption and incompetence become more apparent than ever. The craving for quick solutions provides a fertile ground to populistic propaganda and authoritarian leaders.

The democratic process needs to follow the evolution of technology and become more transparent and flexible. One new trend is experimenting with a reform of the voting system, through ranking the preferential candidates, rather than casting just one vote.[1238] Despite constitutional

1236 Economic and security constraints predetermine political choices and narrow the possibilities of leaders, due to the tight interdependency between states.

1237 The existence of filter bubbles is debated. However, increased diversity and increased selectivity logically leads to fragmentation.

1238 Known as ranked choice voting, preferential voting, or instant-runoff voting. J. Anest, "Ranked choice voting," *Journal of Integral Theory and Practice* 4, no. 3

concerns,[1239] it is thought to dissolve blocs and incentivise the young generation,[1240] reduce polarisation, and reach a higher consensus.[1241] This method is still, however, used in the "winner-takes-all" model. Proportionate systems could better leverage the advantages of this voting method, as coalition governments that are constituted on the basis of such preferential voting could base their policies on the broadest possible consensus.[1242]

10.2. Constructing a value-centred European order

In the present context, the European Union seeks to establish a new media order that fosters a balanced and diverse conversation, thereby capable of serving as a foundation for democratic processes. This aim is consistently present in all those legislative instruments that are discussed in this book, even if the primary legislative goal has been the protection of the internal market.

The establishment of fundamental infrastructure and the implementation of market regulations are undoubtedly important undertakings also from a sheer commercial and competition standpoint. Nevertheless, the underlying infrastructure serves as the foundation for the public communication process, as well. Hence, the rules encompassing the Digital Markets Act (DMA), the Digital Services Act (DSA), the Artificial Intelligence Act (AI Act), and the European Media Freedom Act (EMFA) collectively contribute to shaping the landscape of public affairs. The Regulation on Transparency and Targeting of Political Advertising (RPA) governs a particular aspect of the information "market" that represents an important stage in the fight against disinformation and political manipulation, making it an essential thematic component of this book.

(2009): 23–40. See also: https://fairvote.org/our-reforms/ranked-choice-voting/, https://time.com/5718941/ranked-choice-voting/.

1239 Richard H. Pildes and M. Parsons, "The Legality of Ranked-Choice Voting," *Cal. L. Rev.* 109, no. 1773 (2021).

1240 Daniel McCarthy, and Jack Santucci, "Ranked choice voting as a generational issue in modern American politics," *Politics & Policy* 49, no. 1 (2021): 33–60.

1241 David McCune and Jennifer M. Wilson, "Ranked-choice voting and the spoiler effect," *Public Choice* (2023): 1–32.

1242 Caroline J. Tolbert and Daria Kuznetsova, "Editor's Introduction: The Promise and Peril of Ranked Choice Voting," *Politics and Governance* 9, no. 2 (2021): 265–270.

The regulation that serves human rights purposes is motivated not merely by altruistic intentions or the cultural need to protect human rights. It is also based on rational considerations. For example, before the GDPR was introduced, it was calculated that it would save €2.3 billion a year, if European citizens trusted online commerce and used it without inhibitions.[1243]

Regulation of platforms has started even before the DSA, with the platform-to-business regulation. This, similar to GDPR, aimed at improving trust in online mediation services, in order to "fully exploit the benefits of the online platform economy.[1244] The DMA, which regulates unfair competition, similarly, serves to promote trust by protecting the ethical nature of market processes and the chances of smaller players.

The European Democracy Action Plan (EDAP), and partly also the mentioned platform rules, were triggered by the sudden proliferation of populist propaganda and disinformation, which seemed to threaten the existing political order and European liberal democracy. Again, regulation aimed at restoring trust, this time in the information system. However, liberal democracy is more than just a political agreement in the EU.

The EU economy is built on trust and cooperation. The development of EU law reflects how these values have evolved throughout the history of EU. In order to preserve this trust, market – including information markets as well – needs to be regulated along values. Such regulation, in addition to addressing cultural considerations, also serves the stability of the EU economy. As Polányi described the intertwining of economic and political equilibrium, culture is more deeply intertwined with economic imperatives than is apparent on the surface.[1245] China is stabilising its society through oppression; the US tolerates larger social tensions than European states to

1243 EC (2015) Agreement on Commission's EU data protection reform will boost Digital Single Market. Press Release.

1244 Recital (2) of the P2B Regulation (Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services.).

1245 Karl Polányi, *The great transformation. The political and economic origins of our time* (Boston, MA: Beacon Press 1957).

stay true to its libertarian values in economy,[1246] but also the US is – albeit more gradually – adjusting its policy to adapt to the digital age.[1247]

In sum, the EU's digital policy shifted its focus from enhancing market liberalism to establish cornerstones of a sustainable, constitutional and democratic order.[1248] In the European understanding, ensuring the common good (Gemeinwohl) is the duty of the state, and at the supranational level, of the EU.[1249] In the past, the "public" have been regarded as a blurred concept because the assumed rights or interest of an undefined mass of individuals lacked a concrete factual underpinning. Besides, the public interest consists of several different, sometimes contradicting elements, which are difficult to discern with accuracy. Moreover, authoritarian abuse brought the name "public interest" a bad reputation. For all these reasons, social interests, such as the public health, public morals, were often regarded only as weaker justifications for balancing individual rights. However, in the current digital era, every individual leaves traces in the online world that can be measured and counted. The social interests, derived from these minor pieces of interests and fragmental rights of the masses of individual online users, are now becoming quantifiable and hence more tangible. Historic assumptions about the role of the individual, as existing through and shaped by their interconnected and interdependent relationships, come to a revival.[1250] Individual and society are not to be regarded as a contradiction,

1246 It is fashionable to talk about a crisis of American democracy. It is beyond my limitations to engage in a discussion on that purpose, but see: William G. Howell and Terry M. Moe, *Presidents, populism, and the crisis of democracy* (Chicago, IL: University of Chicago Press, 2020). Afterword V. Lidz, "A Functional Analysis of the Crisis," *American Society* 52, (2020): 214–242. https://doi.org/10.1007/s12108-021-09480-6.

1247 California Consumer Privacy Act of 2018 (CCPA). Biden " Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence". https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/. See also: Nathalie A. Smuha, *"Biden, Bletchley, and the emerging international law of AI," VerfBlog* 2023/11/15, https://verfassungsblog.de/biden-bletchley-and-the-emerging-international-law-of-ai/, DOI: 10.59704/e74941ad144ce5ff.

1248 Georgio De Gregorio, "The rise of digital constitutionalism in the European Union," *International Journal of Constitutional Law* 19, no. 1 (2021): 41–70. https://doi.org/10.1093/icon/moab001. See also: Anu Bradford, Digital Empires. (Oxford University Press, 2023).

1249 Hoffmann-Riem, *Recht im Sog der digitalen Transformation.*

1250 Karen Barad, *Meeting the university halfway: Quantum physics and the entanglement of matter and meaning.* (Duke University Press, 2007) https://doi.org/10.1

as, "[s]ociety is not understood as a discrete object that surrounds humans, but as something that emerges *from* humans—all the while, the individual is produced by society."[1251] Actions of one user always affect the circumstances of other users as each online move contributes to the database that documents human behaviour, mostly for commercial or political purposes. This "butterfly effect"[1252] is faster and more robust than ever presumed, and leaves traces, which can be studied. Alan Turing analysed this idea of cause and effect, writing that "quite small errors in the initial conditions can have an overwhelming effect at a later time."[1253] In the age of Big Data, the focus can lie on communities, rather than on individual interests,[1254] because analytical, commercial, and other actions affect not isolated and specific individuals but large groups. Moreover, the affect rights that are not exercised in isolation, but as a collective, in particular communicative rights, such as the right to freedom of expression and its counterpart, the right to information.

215/9780822388128-002. Among others, the notion is re-discovered again in the Constitution of South Africa which works with the traditional moral concept of the Ubuntu, meaning that people aren't there before they interact; they arise during and as a result of their complex web of relationships.

1251 Emphasis in the original, Andreas Hepp et al., "ChatGPT, LaMDA, and the Hype Around Communicative AI: The Automation of Communication as a Field of Research in Media and Communication Studies," *Human-Machine Communication* 6, no. 1 (2023): 4., 50.

1252 I am referring to Lorenz' theory of the physical causality, and not the popular that relate on time travel, see: Edward Lorenz, "The butterfly effect," *World Scientific Series on Nonlinear Science Series A*, 39, (2000): 91–94.

1253 Alan Mathison Turing, "Computing Machinery and Intelligence," *Mind* LIX, 236 (1950): 433–460. https://doi.org/10.1093/mind/LIX.236.433.

1254 Linnet Taylor, Luciano Floridi, and Bart van der Sloot, *Group privacy: New challenges of data technologies* (Springer Cham, 2016).