STUDY

Requested by the IMCO committee

European Parliament

New aspects and challenges in consumer protection

Digital services and artificial intelligence

Policy Department for Economic, Scientific and Quality of Life Policies
Directorate-General for Internal Policies
Author: Prof. Giovanni SARTOR
PE 648.790 - April 2020

EN

New aspects and challenges in consumer protection

Digital services and artificial intelligence

Abstract

The study addresses the new challenges and opportunities for digital services that are provided by artificial intelligence, in particular which regard to consumer protection, data protection, and providers' liability.

The discussion addresses the way in which digital services rely on AI for processing consumer data and for targeting consumers with ads and other messages, with a focus on risks to consumer privacy and autonomy, as well as on the possibility of developing consumer-friendly AI applications.

Also addressed is the relevance of AI for the liability of service providers in connection with the use of AI systems for detecting and responding to unlawful and harmful content.

This document was provided by the Policy Department for Economic, Scientific and Quality of Life Policies at the request of the committee on the Internal Market and Consumer Protection (IMCO).

This document was requested by the European Parliament's committee on the Internal Market and Consumer Protection.

AUTHOR
Prof. Giovanni SARTOR, European University Institute, Florence

ADMINISTRATORS RESPONSIBLE
Mariusz MACIEJEWSKI
Christina RATCLIFF

EDITORIAL ASSISTANT
Roberto BIANCHINI

LINGUISTIC VERSIONS
Original: EN

ABOUT THE EDITOR
Policy departments provide in-house and external expertise to support EP committees and other parliamentary bodies in shaping legislation and exercising democratic scrutiny over EU internal policies.

To contact the Policy Department or to subscribe for updates, please write to:
Policy Department for Economic, Scientific and Quality of Life Policies
European Parliament
L-2929 - Luxembourg
Email: Poldep-Economy-Science@ep.europa.eu

Manuscript completed: April 2020
Date of publication: April 2020

This document is available on the internet at:
http://www.europarl.europa.eu/supporting-analyses

DISCLAIMER AND COPYRIGHT

For citation purposes, the study should be referenced as: Sartor, G., *New aspects and challenges in consumer protection*, Study for the committee on the Internal Market and Consumer Protection, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2020.

CONTENTS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
EDPS	European Data Protection Supervisor
EU	European Union
GDPR	General Data Protection Regulation
PDA-CDA	Privacy Digital Assistants/Consumer Digital Assistants
US	United States

LIST OF FIGURES

EXECUTIVE SUMMARY

Background

Online consumers find themselves in an unbalanced relation to service providers and traders. A range of powerful intermediaries has emerged, which deliver key services, such as access to the Internet infrastructure, online search, content sharing, cloud computing, and online payments. Some of these services are offered for free to final users, being supported by advertising revenue. Ads are automatically targeted to individual consumers, the targeting being based on information collected by tracking them.

AI has provided technologies with which to exploit the wealth of consumers' information so as to better target individuals. In particular, machine learning has enabled traders to grasp correlations between consumer data (purchases, sites visited, likes on social networks) and possible responses to ads. The ability to predict consumers' reactions provides traders with the ability to trigger such reactions through appropriate ads and other messages. This ability could morph into manipulation, as consumers' responses could be based on irrational aspects of their psychology, on a lack of information, or on a situation of need.

A widespread mechanism for changing behaviours has emerged whose final purpose is to modify people's purchasing behaviour through targeted ads. Thanks to big data and AI, traders may come to know what may influence particular consumers or groups of them one way or the other. More generally, a personal data economy is emerging where all kinds of personal data are collected and exchanged, their value consisting in their possible uses to anticipate and modify the behaviour of people.

The business model based on providing "free" services paid through advertising has an impact that goes beyond e-commerce. In order to expose consumers to ads, platforms have to attract and keep consumers on their websites. AI can discover what kinds of messages and information are more likely to achieve this goal. These tend to include messages — among which rumours or fakes — that please or excite users, confirm their prejudices, trigger negative feelings (such as rage or disgusts), and provide for additive symbolic rewards and punishments. Moreover, individuals tend to be served with kinds of content and messaging that have attracted or pleased similar people in the past. This may lead to separation and polarisation in the public sphere.

AI technologies are also increasingly used by online service providers, to detect and react to unlawful and inappropriate online behaviour. While AI technologies can contribute to effective moderation, enabling providers to cope with the huge growth and accelerated dynamic of user-generated content, they may also deliver inaccurate, biased or discriminatory responses, to the detriment of freedom of speech and users' rights.

Aim

The aim of this report is to identify key issues concerning the situation of consumers relative to service providers and traders as well as the ways in which AI is impacting relations and interactions, and to propose possible solutions.

Presently, consumers' personal data are most often extracted from online services at no cost, and then used and exchanged to the benefit of providers and traders. One way out of this predicament consists

in accepting that personal data are a tradable commodity, but ensuring that data subjects draw some benefit from the use made of their data, while also enabling them to exercise some control over these data. The other way out consists in ruling out the possibility that personal data can be a tradable commodity, i.e., in barring vendors from offering services or benefits in exchange for personal data. On the latter approach, personal data should be used only when necessary to deliver a service requested by consumers, not as something given in exchange for a different service. EU law has not yet chosen between these two models, nor has it found a way to reconcile them. This report argues that guidance should be provided in this regard. Consumer choice can play an important role, whichever approach is adopted, but effective protection of consumer privacy requires that consumers should not be deceive by "design tricks" or "dark patterns" that stealthily induce them to consent to the processing of their data.

The AI-based processing of consumer data is relevant to the main goals of consumer protection law, such as protection of the weaker party, regulated autonomy, and non-discrimination. First, as noted above, the use of AI by vendors/retailers and service providers may introduce further imbalances between these parties on the supply side and consumers on the demand side. Second, the manipulative use of big data and AI may limit the independence of consumers. Third, automated decisions may work to the disadvantage of certain individuals and groups, and without any acceptable rationale. In this regard, some clarification is also needed, possibly through soft law instruments.

Even though the risks that AI poses to consumers are significant, no less important are the opportunities opened up by AI. AI can support citizens and their organizations so that they may not only make better use of the opportunities available in the market, but may also resist and respond to unfair and unlawful behaviour by AI-powered companies. Consumer-empowering AI technologies can support consumers in protecting themselves from unwanted ads and spam; they can enable consumers to identify cases where unnecessary or excessive data is being collected or where fake and untrustworthy information is provided; they can enable support consumers and their organisations in detecting violations of the law, assessing compliance, and obtaining redress. The public could support and incentivise the creation and distribution of AI tools for the benefit of consumers, as data subjects and citizens.

In the domain of Internet moderation, AI may enhance the capacity of providers to detect and react to unlawful and inappropriate online content and behaviour. AI systems can filter out some unlawful content or flag content for human moderation. AI can also assist human moderators by increasing their productivity. It may also be used to reduce the potentially harmful effects of content moderation of individual moderators. However, there is the risk that AI-based moderation may lead to outcomes that are inaccurate, unfair or discriminatory, to the detriment of freedom of expression and information.

Recommendations

The report includes some policy recommendations as follows:

a) Consumers should have the option not to be tracked and (micro)-targeted and should have an easy way to express their preferences;

b) The grounds on which service providers and traders cannot price-discriminate should be specified;

c) It should be considered how discrimination in ad targeting is to be addressed;

d) Guidance should be given concerning what algorithmic practices count as instances of aggressive advertising;

e) Guidance should be given concerning cases in which consumers have a right to contest a decision that undermines their interests;

f) Consumers should be provided with information on whether and for what purposes they are tracked and on whether they are receiving information for advertising purposes;

g) Protection of consumer privacy requires preventive risk-mitigation measures in combination with collective redress;

h) The development of consumer-friendly AI-technologies should be encouraged and supported. Service providers should be prevented from blocking legitimate tools for the exercise of consumer rights;

i) Liability limitations for online providers should also apply to "active" providers, such as search engines, online repositories, and social networks, regardless of whether user-generated content is organised, presented and moderated by humans, by algorithms or both;

j) Limitations on providers' secondary liability should not apply when providers have failed to adopt reasonable precautionary measures that could have prevented that behaviour or mitigated its effects. This failure may also depend on not having adopted the most effective AI technologies;

k) The availability of AI technologies for detecting unlawful online content and behaviour should be encouraged, in combination with human judgment;

l) Third-party filtering/moderation should be encouraged so as to broaden access, and so should the sharing of datasets (to train AI classifiers) and software, so that both are accessible to small companies as well.

1. INTRODUCTION

This study aims to provide an account of the impact of Artificial Intelligence (AI) on consumer protection in the context of e-commerce, and of the Digital Services Act.

AI systems are populating the human and social world in multiple ways: new or expanded online digital services industrial robots in factories, service robots in houses and healthcare facilities, autonomous vehicles and unmanned aircraft in transportation, autonomous electronic agents in e-commerce and finance, autonomous weapons in the military, intelligent communicating devices embedded in every environment.

AI can provide great new opportunities for individuals and society: enhancing human abilities, increasing productivity, improving security and efficiency in public and private services, enabling the universal provision of knowledge and skills. On the other hand, it may increase individual and social risks: providing for pervasive manipulation, and discrimination; disrupting social interactions; and exposing humans to harm resulting from technological failures or disregard for individual rights and collective values.

This study will focus on the impact of AI on new digital services and consumers. This focus is significant not only since consumption is at the basis of the European economy, but also since consumer protection is a key component of EU law. Moreover, a focus on consumers is highly significant because AI applications in the consumer domain are paradigmatic of the role of AI in the context of today information society and information capitalism. The use of AI in advertising and consumer management has indeed paved the way for AI applications in other domains, such as access to information and political campaigning.

In the context of the current debate on a future Digital Services Act, the study will also consider the extent to which AI may affect the delivery of services to consumer.

Though the focus will be on consumer protection, some references will be made to other sources — prominent among these the GDPR[1] — that contribute to addressing the opportunities and risks that come with AI.

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation).

2. DIGITAL SERVICES AND CONSUMERS: THE RISE OF INFLUENCE MACHINES

KEY FINDINGS

Online consumers often find themselves in an unbalanced relation to service providers and traders. A range of powerful intermediaries has emerged, which deliver key services, such as access to the Internet infrastructure, online search, content sharing, cloud computing, and online payments. Some of these services are offered for free to final users, being supported by advertising revenue. Ads are automatically targeted to individual consumers, the targeting being based on information collected by tracking them.

A significant shift in the sentiment towards information technologies in the consumer domain can be observed to have taken place in the last decades. This shift reflects the evolution of the information economy, in particular the emergence of monopolies over data and services as well as the massive collection of personal data, and their use to influence consumer behaviour.

Two or three decades ago, when the Internet was still in its beginning, the mainstream assumption was that information and communication technologies would deliver a new economic environment, making for new exiting opportunities for both producers and consumers: disintermediation, unlimited access to information, larger and open markets, the opportunity for global interactions. It was even assumed that the distinction between producers and consumer could also be overcome, thanks to production models where individuals self-select for participation in common projects, inspired by a culture of sharing and openness[2]. Creative outcomes could also be made public, as information technologies would make sharing costless, enabling unlimited access.

This perspective reflected the emergence of a new model for the production and delivery of information, the so-called web 2.0, in which the public actively participates in the creation of content, through only blogs, wikies, repositories, and social networks,. This new model would supplement and sometimes replace traditional forms of cultural production, once more taking down the distinction between traders and producers of information. The main focus of the debate was on the conflict between, on the one hand, the legal and socio-economical restrictions to the free-flow of information (such as censorship and copyright) and, on the other hand, the need to enable the free growth of the new online ecology. In this context, the newcomers, the netizens enjoying the free flow of information across the new infrastructure, and the newly emerging web entrepreneurs, were opposed to the incumbents, copyright holders and states, whose interests lay in limiting and controlling that flow[3].

[2] Benkler (2006).

[3] For an analysis see Lessig (2001).

Figure 1: Consumer power

Source: Economist April 2003

In the consumer domain, it was assumed that the internet would strengthen the market power of consumer relatively to traders: any consumer would be able to access a global marketplace, where he or she would select the most convenient opportunities. The market would discipline the behaviour of traders; consumers would obtain information on products and prices through search tools, and this information would be expanded and validated through collaborative tools, such as consumers' ratings on their purchasing experience. As the cover of the April 2003 issue of the Economist claimed, the Internet would bring power to consumers: "the internet means that the consumer really is king (and queen)" (see Figure 1).

Todays a more sober attitude can be observed, which reflects the emergence of a new sociotechnical infrastructure. On the one hand, a set of powerful intermediaries have emerged, in multiple domains, from access to the internet infrastructure, to search engines, to platforms for sharing online content, to e-commerce, to cloud services, to online payments[4]. These new intermediaries tend to enjoy a monopoly or oligopoly position, as in information technology services size is usually an advantage, due to well-known aspects of the information economy, such as network effects, small marginal costs, the possibility of packaging and integrating multiple services, the advantage of possessing vast amounts of information. Much information is collected in the context of the provision of services: in online services to consumers a two way transmission of information takes place: from the provider to the consumer, but also from the consumer to the provider[5]. Computer systems run by providers/traders can observe, verify and analyse any aspects of the transaction, recording every character typed on a

[4] As it was already recognised in OCDE (2010), see also Sartor (2017).

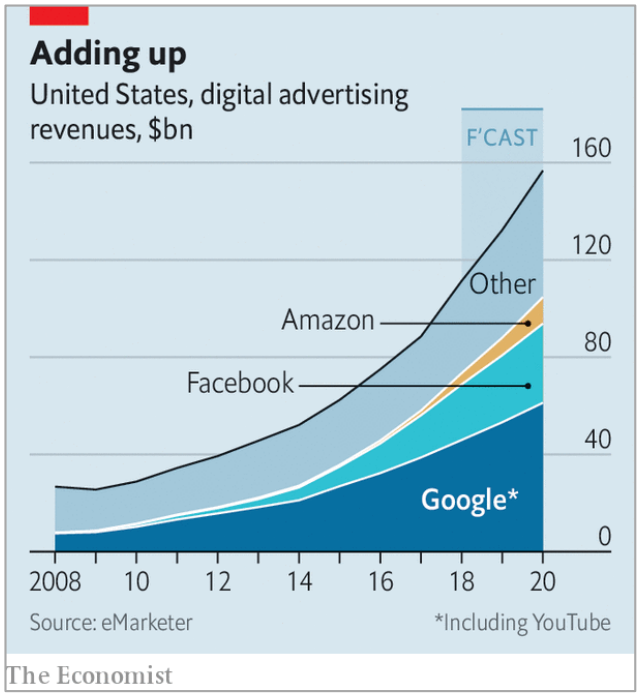
[5] Varian (2010, 2014).

keyboard and every link clicked. Thus, monopolies over the online provision of services tend to become monopolies over the collected data.

With regard to the online provision of information services — search engines, online repositories, social networks — the business model has emerged according to which services are offered for free to final users, but they are backed by advertising revenues. Thus, such key services for the information society are offered on two sided markets[6]: providers have two different classes of clients — advertisers, and users — and have to take both into account. There is an interdependence between advertisers and users: to satisfy advertisers, intermediaries must attract and retain users. We may also say that consumers' attention and information about consumers are the key commodity that providers sell to advertisers. As a popular meme says, users of free services are not consumers, they "are the product."

It has been observed[7] that the emergence of this model was due to the convergence of two ideologies, or value-frames, both playing a powerful role in the Internet culture[8]: on the one hand the libertarian-egalitarian strand, according to which information ought to circulate freely and online services ought to be freely accessible to everybody (where freely means both without proprietary constraints and at no cost), and on the other hand an entrepreneurial strand, focused on successful business and money-making. Unfortunately, while this business model was undoubtedly successful on the entrepreneurial side, leading to the emergence of some of the richest and most innovative companies of today's economy, such as Facebook and Google, its record from a liberal-egalitarian perspective is questionable, as this business model contributes to pervasive surveillance and influence over citizens, and in particular over consumers.

Figure 2: Concentration in online ads



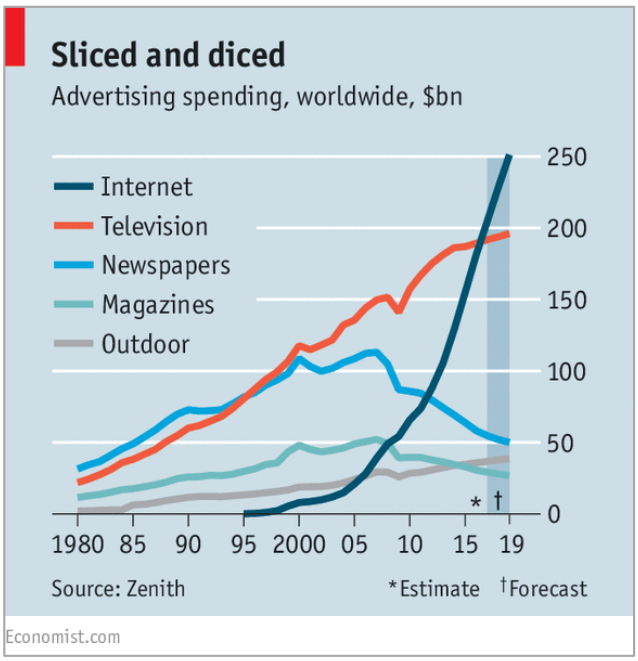
Source: Economist March 2017

6 Rocher and Tirole (2003), Hagiu 2009.

7 According to Lanier (2018).

8 See Castells (2001), on the different cultures of the Internet.

Figure 3: The share of digital advertising



Source: Economist October 2018

Relatively to other form of advertising, web-based advertising has a decisive advantage: messages can be automatically targeted to the individual consumers, the targeting being based on information about such consumers. In the beginning, the targeting would take place in innocuous ways, similar to the way in which in the pre-internet context, different ads would be placed in different periodicals or in different features or departments of them: in the so-called context-based advertising, users are presented with ads that fit the page they are browsing, or the search they have just made. However, the internet service providers learned very quickly that further information could be obtained about users, the more so as the internet expanded, becoming a universal medium for the delivery of any kind of services. Thus, the possibility was open to use this information for the purpose of targeting consumers. Online advertising boomed, in few years overtaking traditional forms of advertising (such as, particular newspaper advertising), which lost considerable part of their advertising revenue (see Figure 2 and Figure 3).

3. AI AND TARGETED ADVERTISING

KEY FINDINGS

AI has provided technologies with which to exploit the wealth of consumers' information so as to better target individuals. Machine learning has enabled traders to grasp correlations between consumer data (purchases, sites visited, likes on social networks) and possible responses to ads. The ability to predict consumers' reactions provides traders with the ability to trigger such reactions through appropriate ads and other messages. This ability could morph into manipulation, as consumers' responses could be based on irrational aspects of their psychology, or a lack of information, or on a situation of need.

At this point AI entered in the scene, providing technologies through which to exploit the wealth of consumers' information so as to better target individuals: big consumer data and AI have converged, providing a new infrastructure for addressing and managing consumers. In fact, in the last couple of decades the statistical machine-learning, approach has become dominant in AI. This approach uses big data sets to automatically build models that track correlations, and then uses such models to make predictions for new cases. In the consumer domain this has meant that records of past consumer behaviour could be used to grasp correlation between consumer data (purchases, sites visited, likes on social networks) and possible responses to ads and other consumer related messages).

On this basis, consumers could be targeted with offers similar to those that had a response with similar consumer in the past. Moreover, such AI systems would be able to learn from their own successes and failures (according to the model of the so-called reinforcement learning), namely, they would learn how to accurately address consumers sharing certain features exactly with the messages that had been successful with the same kind on consumers in the past. Thus, the ability to predict consumers' reactions provides traders with the ability to trigger such reactions through appropriate ads and other messages. This ability could become manipulation, as consumers' responses could be based on irrational aspects of their psychology, rather than on reasoned choice, and consumers may be unaware of the way in which they are being influenced.

In fact the correlations discovered by AI systems may correspond to multiple causal mechanisms in consumers' psychology: maybe there is a genuine fit between reasoned preferences and the purchases suggested by targeted ads, and this explains why the consumers follows such suggestions; or maybe that the ad-sending machine is just profiting from a weakness in the targeted consumers: it is exploiting their anxieties, insecurities, credulousness, addition in order to lead them into choices they will later regret. AI-based learning systems are not behaving badly on purpose, they are not immoral, but rather amoral, namely, they discover and applies the most effective solutions in view of their purpose of prompting consumers to make purchases, regardless of the ethical implications of such solution. The success of ads may be owned to their fit between what is presented and the reasoned preferences of consumers or, on the contrary. people may also be primed to purchase goods they do not need, to overspend, to engage in risky financial transactions, to indulge in their weaknesses (e.g. gambling or drug addiction). It has been said that such systems engage people based on a "radical behaviourist approach:" they rely on cognitive and behavioural patterns that often operate on automatic, near-instinctual levels and that may be manipulated instrumentally[9].

[9] Cohen (2019), see also Zuboff (2018).

Algorithms can indeed learn to exploit all kinds of consumer biases: availability cascades (as when messages are repeated to convince consumers); bandwagon effects (as when the message convinces consumers that they should do what many others are enthusiastically doing); confirmation biases (as when the offer is linked to what the consumers already likes or knows); fears of missing out (as when a short time frame is presented for choices), herd behaviour (as when offering rewards under referral programs), anchoring/framing (as when artificially high prices are shown with large discount offers).

The goal of sending more and more effective targeted ads to consumers provides a key incentive for mass surveillance, leading to the massive collection of consumer data: all online activity, every click or message, can be recorded in order to use it subsequent discover possible correlation that may be useful in influencing consumers through the most effective adds. Psychographic techniques can be deployed to extract the personality types and psychological attitudes of consumers. This discloses new opportunities for manipulation, as consumers can be targeted with the ads that are more effective, relative to their personality[10]. Emotion detections techniques are also increasingly available to traders: these enable the monitoring of facial expression and record voices to infer emotional states and reactions of consumer and use this knowledge in transactions.

In the contexts of computer-powered smart homes, cars, and cities, AI is embedded into physical objects (e.g. house appliances, cars, roads, etc.). Interconnected sensing devices increase the amount and the specificity of data collected and enable ubiquitous commercial interactions with consumers[11].

[10] Burr and Cristianini (2019).

[11] Helberger (2016).

4. FROM ADVERTISING TO FILTER BUBBLES

KEY FINDINGS

The business model based on providing "free" services paid through advertising has an impact that goes beyond e-commerce. In order to expose consumers to ads, platforms have to attract and keep consumers on their websites. Machine learning can discover what kinds of communications are more likely to achieve this goal. These may include messages — among which rumours or fakes — that please or excite users, confirm their prejudices, trigger negative feelings (such rage or disgusts), and provide for additive symbolic rewards and punishments. Moreover, individuals can be served with kinds of content and messaging that have attracted or pleased similar people in the past. This may lead to separation and polarisation in the public sphere.

The business model based on providing "free" serviced paid through advertising has an impact that goes beyond e-commerce strictly understood. In order to expose consumers to ads, platforms have to attract and keep consumers on their webpages. Machine learning methods can be used for this purpose, too, namely, to discover what kinds of messages and information are more likely to achieve this goal. The system's working is amoral here, too: what matters is that the attention of users is obtained, it does not matter whether this is obtained by sending them relevant and useful information, or rather by exposing them to messages — including rumours or fakes — that please or excite them, confirm their prejudices, trigger negative feelings, such rage or disgusts, provide for additive symbolic rewards and punishments.

It has been observed that the purpose of targeting each user with the information that is mostly likely to "engage" or rather addict him or her may have the effect of carving fault lines into the public sphere: no longer can citizens rely on a common shared set of facts, rather tendencies toward separation and polarisation are emphasised. It had been anticipated decades ago that the free choice of individuals — concerning what digital content to access or with whom to interact online — could lead to the splitting of a national audience in separate non-communicating and polarised factions[12]. This is happening to some extent today, though in a way that is different was from how it was anticipated: the fragmentation of the public sphere, no longer results only from individuals' choices, but also from algorithms aimed at attracting our persistent attention. The outcomes are what we know as filter bubbles or echo chambers: the information that people receive is selected — by search engines and news feeds — is based on the degree to which similar people have been attracted or pleased by such information. As people are usually affected by what is called "confirmation bias" (they prefer to see what coheres with their mindset), they are provided with information that validates and strengthens their current convictions[13]. They are similarly directed by social networks to interact with those who share their same background, since this would lead to more frequent and pleasant information exchanges. This dynamic may lead to political communities being split into non interacting factions, each one aimed at pursuing its goals according to its different beliefs, without engaging in dialogue, mediation, and the search for truth. In fact, polarization on the one hand is a side effect of the search for people's attention, but on the other hand it is a booster of attention. It has been observed that the appeal to tribal instincts,

[12] Sunstein (2001).

[13] Pariser (2011).

putting us vs. them, often is the cheapest way to create the appearance of a relationship, to create a fake community unified by a common enemy[14].

[14] As argued by Lawrence Lessig in his 2018 Ted talk "How Digital Destroyed Democracy."

5. A NEW LANDSCAPE

KEY FINDINGS

A widespread mechanism for changing behaviours has emerged whose final purpose is to modify people's purchasing behaviour through targeted ads. Thanks to big data and AI, traders may come to know what may influence particular consumers or groups of them one way or the other. More generally, a personal data economy is emerging, where all kinds of personal data are collected and exchanged, their value consisting in their possible uses to anticipate and modify the behaviour of people (and in particular of consumers).

The combination of the trends just described leads to the emergence of a widespread mechanism for behavioural modification: the final purpose is to modify people's purchasing behaviour through targeted ads. However, as just noted, this final goal also determines the instrumental goal of modifying the behaviour of users of online services, by sending them engaging/addictive items of information, in particular to the users of social networks and online ads repositories.

Personalised advertising also leads to personalised consumer management, as offers or rejections of requests as well as further interactions can be based on the knowledge obtained about consumers. In this context, a new imbalance emerges between traders (supported by the AI-driven influence technology) and consumers. Not only do traders know their products and services better than the consumer does, but they may know about consumers much more than the consumer knows about them. Traders, or rather the big data and AI-based systems supporting them (possibly thanks to the data and technologies provided by online intermediaries) may know what may influence the consumer one way or the other.

By analysing offers to similar consumer and possibly experimenting with them, AI system may come to know the reserve price of particular consumers, i.e., the highest price they may accept. As a consequence, new market arrangements may emerge: traders may be able to push onto the consumer the highest offer that the latter may accept, eroding consumer surplus. This may not only affect consumers' wellbeing, but also the efficiency of markets, which is based indeed on the assumption of the absence of price discrimination [15].

More generally, it has been argued that systematic surveillance for the purpose of influencing people has led to a new economic model, the "surveillance capitalism." This model is characterised the commodification of human experience, which it turns into recorded and analysed behaviour, i.e., which it transforms into marketable opportunities to anticipate and influence [16]. In fact, a personal data economy has emerged, where all kind of consumer data, individual and aggregated are collected and exchanged by the largest platforms, but also by all kinds of service providers, sellers, media companies, and public bodies. The data, or rather services based on them (e.g., advertising services), may be sold for a profit, the exchange value of such data being based on its possible uses to anticipate and modify the behaviour of people (and in particular of consumers).

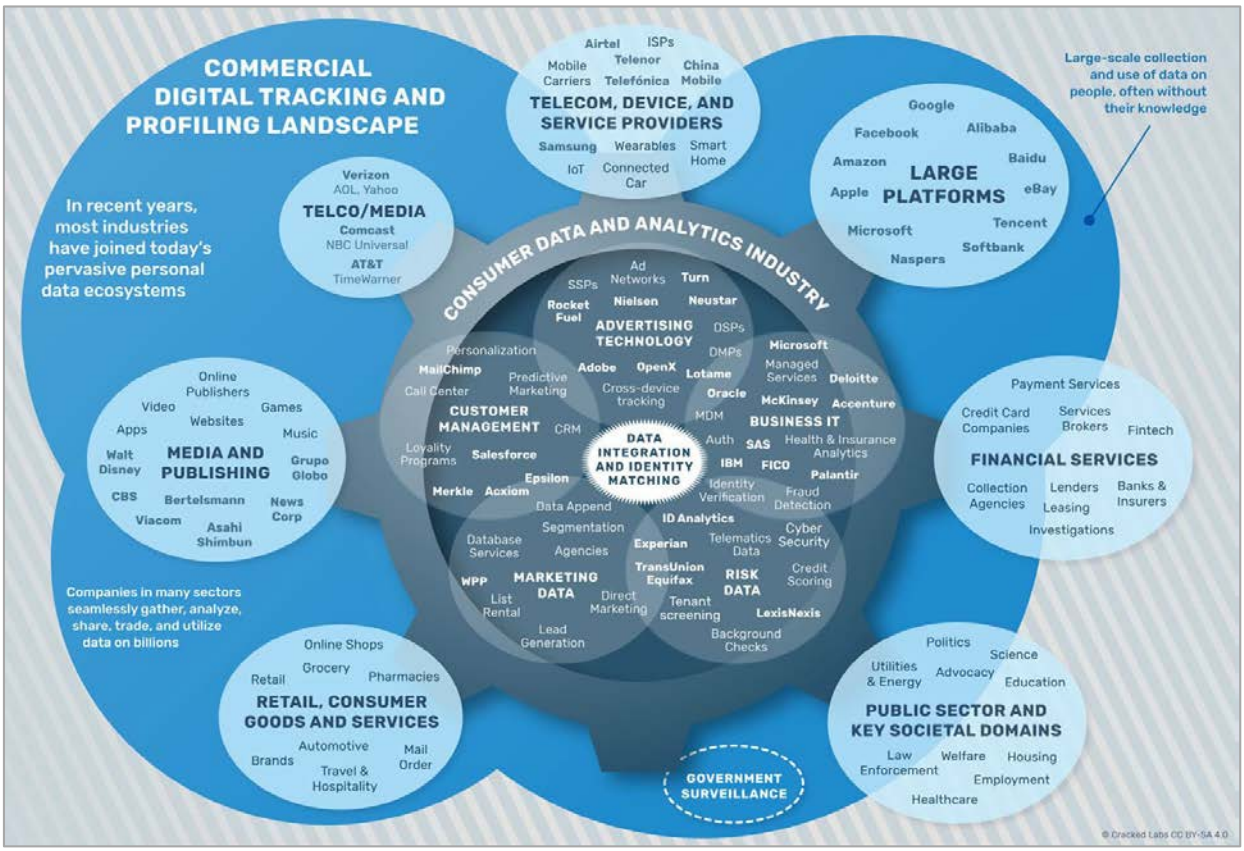
The model introduced relatively to advertising does not need to remain limited to the economic domain. The same methods for prediction and influence that can be applied to the purchasing of products can also be applied to electoral campaigns and more generally to any effort to influence the

[15] Stiglitz (2019, 115).

[16] Zuboff (2019).

public opinion. As the Cambridge Analytica case clearly exemplified, the data collected from social networks, merged with data from other sources, can be used to understand people character, interest, and political views, and consequently to target them with messages meant to change their voting behaviour.

Figure 4: Data services (from Cracked Labs 2017)



Source: Cracked Labs 2017

Figure 4 shows the multiple actors involved in the consumer tracking industry, in which online platforms play a key role, but which also includes retail suppliers, media and publishing industries, telco and telecom services.

6. MONETISING VS NON-MONETISING CONSUMER DATA

KEY FINDINGS

Presently, consumers' personal data are most often extracted from online services at no cost, and then used and exchanged to the benefit of providers. One way out of this predicament consists in accepting that personal data are a tradable commodity, but ensuring that data subjects as well draw some benefit from the use made of their data, while also exercising some control over these data. The other way out consists in excluding that personal data can be a tradable commodity, i.e., in barring vendors from offering services or benefits in exchange for personal data. On the latter approach, personal data should be used only when necessary to deliver a service requested by consumers, not as something given in exchange for a different service. EU law has not yet chosen between these two models, nor has it found a way to reconcile them.

There is the need to overcome the current situation, where consumers' personal data are extracted at no cost from online services and used to modify consumers' behaviour. Progress can take two different directions, among which I believe a choice will have to be made, or at least a balance will have to be struck: either personal data are viewed as a valuable tradable property, or they are viewed as an inalienable asset, one that cannot be traded for a consideration.

The first alternative, i.e., viewing personal data as a tradable property[17], consists in accepting the view of that a market for personal data exists and is here to stay. Thus, the ability to influence people, in consumption as well as in other dimensions of their individual and social life — which can be obtained by processing personal data — is to be viewed as a tradable commodity in the information economy[18]. The policy goal should therefore be to prevent abusive exchanges and provide a fairer allocation of the profits resulting from such a market. In particular, it should be ensured that a share of such profits also goes to the data subjects who accept being tracked and targeted. Following this idea, consumer data should be viewed as a property that informed consumers may exchange for a fair consideration, which may consist either in services or in money. Arguably, this would mean an improvement of consumers' situation. They would move from a situation in which their data are available "on the wild", being free for the taking — or are valued no-more that the provision of a service (which usually has zero marginal cost for the provider) — to a situation in which, in exchange for their data and attention, they obtain a monetary or other benefit. We could also imagine that organisations would emerge to which consumers transfer the management of their data, according to their preferences, with the task of bargaining with providers and advertisers and of extracting advantageous deals for consumers. Possibly participation in such an organisation would provide consumer with some information about and control over the way in which their data are used, though this might involve further collection and duplication of personal data, with additional privacy risks[19].

17 As suggested by Lanier (2014, 2018).

18 As extensively argued by Zuboff (2018).

19 This issue was already addressed by Schwartz (2004). See also Ayres and Funk (2003).

A view of privacy as a tradable asset may seem to be assumed by Article 3 Directive (EU) 2019/770[20] and Article 3 of Directive 2011/83[21] (as amended by Directive 2019/2161[22]), according to which the same directives "shall also apply where the trader supplies or undertakes to supply digital content [...] or a digital service to the consumer and the consumer provides or undertakes to provide personal data to the trader." This norm indeed assumes that consumers may agree to provide traders with data that are not needed for delivering goods or services to such consumers. Thus, such data — unless they are meant to be used for a goal that is shared by the consumer, or are collected without a real consent by consumer — must have been provided as a counter-performance for services or accessory monetary or other benefits provided by the trader.

The second alternative would on the contrary consist in ruling out that personal data can be a tradable property. On this approach, traders would be barred from offering services or benefits in exchange for personal data. First of all, there should be no "tracking walls", namely, it should not be the case the use of "free" services be conditioned on the users' consent to being tracked: users should not be faced with the choice between being tracked while using a service, and being unable to use the service[23]. Under such conditions consent would not be considered to be free, and therefore the processing would be unlawful. Following a rigorous view of privacy and data protection as fundamental rights, even the exchange between personal data and a monetary or other advantage (e.g. a discounted or free access to a paid service) should be excluded. On this view, the collection of data by a commercial operator would only be permissible when the data are necessary for a service that the user desires for its intrinsic merit (e.g. a health service, or also recommendations on purchases) or for the pursuit of an individual or social goal the user endorses (e.g. fighting an epidemics or supporting scientific research,). This view assumes that an unbundled and granular set of services is proposed to consumers, so that they can accept the main service while refusing further services that require tracking (e.g., recommendations based on viewed products). A fortiori, all commercial practices in which consumers are lured into providing their personal data in exchange for coupons or participation in lotteries would be excluded. This perspective is to some extent supported by the GDPR, which at Recital 43 of the GDPR specifies that "Consent is presumed not to be freely given (...) if the performance of a contract, including the provision of a service, is dependent on the consent despite such consent not being necessary for such performance." It is also supported by the revocability of consent under the GDPR.

The clash between the two perspectives has emerged in the Opinion 8/18 by European Data Protection Supervisor (EDPS) on the legislative package "A New Deal for Consumers". The EDPS observed that a synergy exists between data protection and consumer protection. The two regimes share the common goals of correcting imbalances of informational and market power, and, along with competition law, they contribute to ensuring that people are treated fairly. However, according to the EDPS the provision of personal data should be considered a counter performance to be exchanged for the delivery of a service or other consideration. According to the EDPS, consent to the processing of personal data should not be bundled with acceptance of terms and conditions, it should be "separate from the consent needed for the conclusion of the contract." Therefore, the EDPS criticised the

[20] Directive (EU) 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services.

[21] Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights.

[22] Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules.

[23] As argued by the EDPS Opinion 3/18. See Borgesius et al. (2018).

assimilation, under article 3, of the contracts in which the consumer pays price and those in which "the consumer provides or undertakes to provide personal data to the trader."

It seems to me that although the opposition between two different visions of consumer privacy, and two alternative regulatory perspectives is clear, the formulation of Article 3 of Directives 2019/770 and Article 3 of Directive 2011/83 can be reconciled with the view that Privacy and Data protection are inalienable fundamental rights. This result can be obtained by considering that the directive on consumer rights does not override the GDPR: if the consumers' consent to the processing of their data will not satisfy the GDPR requirements, such a consent will be invalid and the processing by the providers/traders will be unlawful. However, the consumers whose data have been unlawfully processed should still enjoy the consumer rights under Directives 2019/770 and Directive 2011/83.

7. CONSUMERS' PROFILING AND CONSENT

KEY FINDINGS

The AI-based processing of consumer data for the purpose of personalised advertising falls under the concept of profiling which, according to the GDPR should also be arranged in such a way as to provide security and avoid discriminatory effects.
The effective protection of consumer privacy requires that consumers are not tricked by "design tricks" or "dark pattern," that deceivingly induce them into consenting to the processing of their data.

The AI-based processing of consumer data for the purpose of personalised advertising falls under the concept of profiling according to GDPR — i.e., on the use of personal data for inferring further information and taking decisions — and thus is subject to the corresponding regulatory framework. All considerations concerning the legitimacy of profiling, namely, of the processing of personal data for the purpose of making inferences concerning individuals and adopting consequential actions (e.g., sending or not sending an offer or and ad) apply to it.

In particular it is highly relevant that there be a possibility to opt out on personal grounds, as recognised in general whenever consent provides the legal basis of the processing. Profiling in the context of direct marketing is addressed in Article 21 (2) GDPR, which recognises an unconditioned right to object to profiling for the purpose of direct marketing. Finally, researchers have also pointed out that automated inferences need to be "reasonable"[24], which may include both statistical accuracy and normative acceptability. This idea that can be found in Recital (71) of GDPR, according to which profiling should not only be based on sound mathematical-statistical principles and on accurate data, but should also be arranged in such a way as to provide security and avoid discriminatory effects. Moreover, the limitations concerning automated decision making according to Article 22 GDPR may apply to automated determinations having a serious impact on the interests of consumers. This will be the case when an application is rejected which is critical for the consumer's life prospects (e.g., the application for a loan or a house rent) or also when systematic systemic discrimination or exclusion from market opportunities is at stake.

The effective protection of consumer privacy requires that consumers are not tricked by "design tricks" or "dark pattern," that deceivingly induce them into consenting to the processing of their data[25]. Simple and clear information should be given on how to opt-in or opt-out relative to critical processing, such as the tracking of users or the transmission of data to third parties. An interesting example is provided by the new California Data Privacy Act, which requires companies to include in their website a link with the words "do not sell my data" (or a corresponding logo-button) to enable users to exclude transmission of their data to third parties. It could be argued that the choice of rejecting all processing of personal data that are not needed for a service being delivered should be clearly offered a default option to all users, at least in those cases in which access to a service cannot be conditional to tracking users and processing their data.

[24] Wachter et al (2017), Edwards and Veal (2019).

[25] Norwegian Research Council (2018).

8. FROM DATA PROTECTION TO CONSUMER PROTECTION

KEY FINDINGS

The AI-based processing of consumer data is relevant to the main goals of consumer protection law: protection of the weaker party, regulated autonomy, and non-discrimination. First, as noted above, the use of AI by vendors/retailers and service providers may introduce further imbalances between the supply side and the demand side. Second, the manipulative us of big data and AI may limit the independence of consumers. Third, automated decisions may work to the disadvantage of certain individuals and groups, and without any acceptable rationale. In this regard, some clarification is also needed, possibly through soft law instruments.

The AI-based processing of consumer data is relevant not only to data protection law, but also to consumer protection law. Such a processing may indeed affect the key values that underly EU consumer protection, such as the protection of the weaker party, regulated autonomy, and non-discrimination[26].

First of all, as noted above, the use of AI by providers/traders may introduce further imbalances between traders and consumers. Traders, by using big-data and AI system, may gain much more information about consumers than consumers have about traders. Moreover, each individual consumer can only devote a limited effort to collecting information and reasoning with it, while traders can rely on the incessant processing done by a vast networks of computer system, which deploy their huge computational power over vast datasets, and use the latest AI technologies. As noted above adaptive systems constantly improves their performance, on the basis of new data and persistent experimentation.

Secondly, the use of big data and AI may limit the autonomy of consumers. Their ability to make adequately informed choices in light of their reasoned preferences is challenged by the possibility to influence consumers' choices, possibly without them being aware of such influence[27]. Consumer may be "hyper-nudged" by targeted advertising and adaptive manipulative design into choices they will possibly regret: this can be achieved by profiting of (mis)perceptions and vulnerabilities[28]. Thus, people may be induced to purchase goods they do not need, to overspend, to engage in risky financial transactions, to indulge in their weaknesses (e.g. gambling or drug addiction). The opportunity for undue influence is emphasised by psychographic techniques, which enable psychological attitudes to be inferred from behaviour, thereby disclosing opportunities for manipulation[29]. Moreover, the use of image and voice recognition technologies allow businesses to capture emotional response to advertising and to exploit such knowledge in consumer transaction to trigger desired behaviour. The concepts of misleading and aggressive commercial practices need to be updated to take into account development of AI.

Third, the use of AI is relevant to discrimination, as AI and big data allow new, much more refined and systematic, forms of stereotyping and differentiation to takes place. On the one hand, the individuals whose data support the same prediction or who are subject to the same influence patterns (e.g., same expected compliance rate, or responses to ads) will be considered and treated in the same way. AI

[26] See Jablonowska et al (2018).

[27] Calo (2013), Helberger (2016).

[28] Bar-Gill (2018); Calo (2014); Mik (2016); Yeung (2018).

[29] Burr and Cristianini (2019).

represents a challenge for discrimination law, as individuals may unfairly subject to differential treatment based on criteria that do not directly match prohibited discriminations under EU law (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). Moreover, the discrimination may result from different defects of the AI systems being deployed, pertaining in particular to its training set, and it may be difficult to provide explanations of the behaviour of such systems[30].

In the application of AI in the consumer domain intentional discrimination is rare, but still automated decisions may be disadvantageous for certain individuals and categories, without an acceptable rationale. For instance, as noted above the practice of price-discrimination may negatively affect consumer welfare as well as the efficiency of markets. While it may be reasonable to provide different prices when this reflects different risks or costs, in other cases traders offer higher prices exploiting people's situation of need or misconceptions. Certain categories of people, may be systemically excluded from certain marked opportunities (e.g., loans or jobs), based on individualised automated assessment. We should consider, however, that AI and big data system may also provide more objective and fair outcomes that biased or defective human decision-making[31]. Whenever, as in the case of lending or recruiting, decisions were discretionally done by individuals, even a biased system can in some cases provide fairer results than even more biased human decision makers.

[30] Barocas et at (2016), Kleinberg et al (2018).

[31] Kleinberg et al (2019).

9. FROM CONSUMER PROTECTION TO CONSUMER EMPOWERMENT

KEY FINDINGS

AI can support citizens and their organizations so that they may not only make better use of the opportunities available in the market, but may also resist and respond to unfair and unlawful behaviour by AI-powered companies. Consumer-empowering AI technologies can support consumers in protecting themselves from unwanted ads and spam; they can enable consumers to identify cases where unnecessary or excessive data is being collected or where fake and untrustworthy information is provided; ; they can or to support consumers and their organisations in detecting violations of the law, assessing compliance, and obtaining redress. The public could support and incentivise the creation and distribution of AI tools for the benefit of consumers, as data subjects and citizens.

An adequate regulatory framework is an essential element to direct artificial intelligence towards the good of individuals, groups and society as a whole. However, regulatory instruments and their implementation by public bodies may be insufficient. Indeed, AI and big data are employed in domains already characterized by a vast power imbalance, which they may contribute to exacerbate. In fact, these technologies create new knowledge (analytical and forecasting abilities) and powers (control and influence capacities) and make them available to those in control, and in particular to big companies and public authorities. As note above big data and AI can increase the imbalance of power between providers/traders on the one side, and consumers on the other.

A possible, albeit partial, remedy can be identified by establishing a parallelism between the dynamics of power underlying the development of artificial intelligence and those related to the industrial society and the mass consumption society. In both cases, a limit to the market excesses, was found in social movements, such as those of workers and consumers. To ensure an adequate protection of citizens, beside regulation and public enforcement, also the countervailing power of civil society[32] is needed, to detect abuses, inform the public, activate enforcement, etc. In particular, consumer organizations have traditionally provided an important contribution to inform the public, promote the application of consumer protection law, and exercise forms of collective pressure. In the AI era, effective countervailing powers needs to be supported by AI: if citizens and their organizations are able to use AI to their advantage, they can better resist, and respond to the unfair and unlawful behaviour by AI-powered companies.

A few examples of consumer-empowering technologies are already with us, such as ad-blocking systems as well as more traditional anti-spam software and anti-phishing techniques. The growing interest in data protection and consumer protection has resulted in several proposals for further tools for consumers, often based on AI technologies.

Systems can be built to automatically extract, categorize and summarise information from privacy documents and consumer contacts, and assist users in processing[33] and understanding their contents. Machine learning and natural language processing methods can also be used to identify cases where unnecessary or excessive data is collected, or where fake and untrusty information is delivered[34]. AI can

[32] Galbraith (1956).

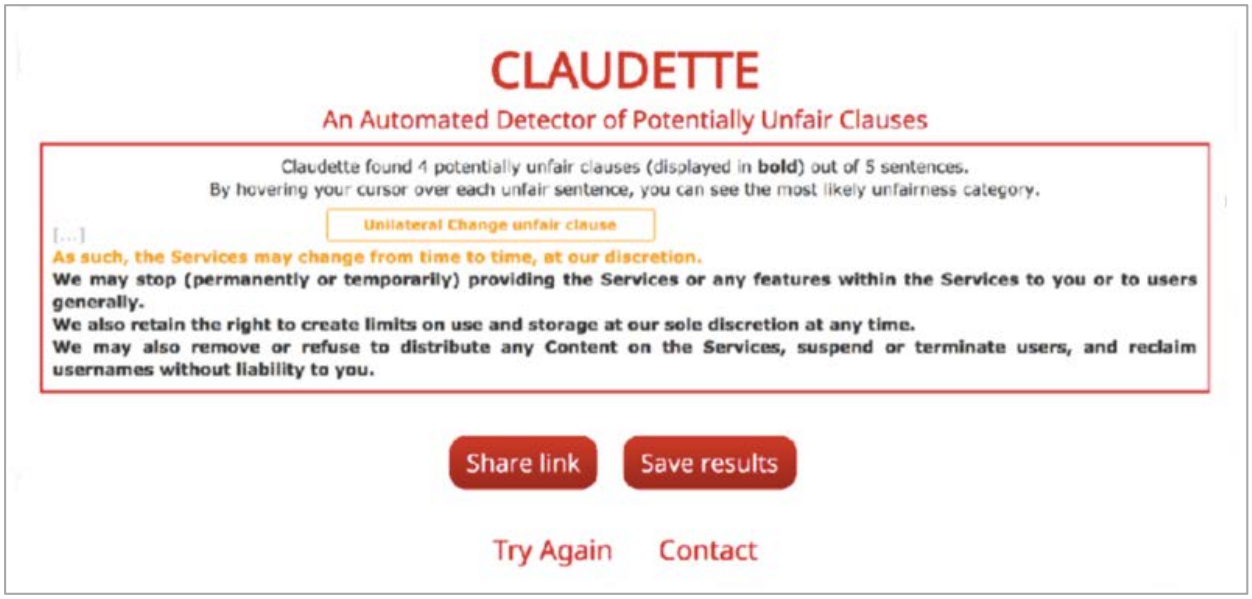
[33] See Costante et al (2012),I Gordon et al (2015).

[34] Bodó et al (2017).

contribute to address such consumer's "information overload"[35] by enabling consumers to isolate and understand relevant parts and act upon.

AI may directly contribute to the implementation of consumer protection law by addressing the detection of law infringements and the assessment of compliance, as well as by supporting consumers in the exercise of their rights. For instance, AI may help consumers and their organisations in determining whether contractual clauses are unfair, whether a privacy policy violates legal requirements, whether an advertisement is potentially misleading or aggressive, etc. One example in this direction is offered by Claudette (Figure 5)[36], an online system which uses machine learning techniques to automatically detect unfair clauses in online contracts and in privacy policies[37]. The Claudette system for online contracts has been trained on a data set of more than one hundred online contracts, in which unfair clauses have been identified and classified (terms of service). On this basis, using multiple algorithms for machine learning, the system has acquired the capacity to detect an classify unfair clauses in new contracts. A similar methodology has been used to provide Claudette with the capacity to identify unlawful clauses in privacy policies.

Figure 5: The Claudette system



Source: http://claudette.eui.eu/

AI-based system could also support consumers and their association, in grouping similar cases, and building legal documents.

Considerable effort has also been devoted to the development of data mining techniques for detecting discrimination[38], as with the aim to build supporting tools that could identify prejudice and unfair treatments in decisions, such as those related to loan and job applications, and the granting of social benefits.

Multiple AI methods to support data protection and consumer protection could be merged in integrated PDA-CDA (privacy digital assistants/consumer digital assistants), meant to prevent excessive/unwanted/unlawful collection of personal data, as well as to protect from manipulation and

[35] Elshout et al Elsen (2016).

[36] https://claudette.eui.eu/.

[37] Contissa et al (2018a, 2018b).

[38] Ruggeri et al (2010).

fraud. More generally, AI guardians have been recently proposed[39] as AI programs that examine other AI program, "ethically-bounded" AI systems[40]. Recent research has envisioned the development of the so-called algorithmic guardians, i.e., software tools completely under human control, able to protect us from undesirable behaviours of third party algorithms[41].

It may be worth considering how the public could support and incentivise the creation and distribution of AI tools to the benefit of data subject and citizens. Such tools would provide new opportunities for research, development, and entrepreneurship. They would contribute to reduce unfair and unlawful market behaviour and favour the development of legally and ethically sound business models. In conclusion, citizen-empowering technologies would support the involvement of civil society in monitoring and assessing the behaviour of public and private actors and of the technologies deployed by the latter, encouraging active citizenship, as a complement to the regulatory and law-enforcement activity of public bodies.

[39] Etzioni & Etzioni (2016).

[40] Rossi & Mattei, (2019).

[41] Zambonelli et al (2018).

10. PROVIDERS LIABILITY AND THE DIGITAL SERVICE ACT

KEY FINDINGS

The regulation on providers' liability contained in the 2000 eCommerce Directive should be revisited, to adapt them to the present circumstances. Leading online companies can sustain the financial burdens for content moderation, and technologies are available, in particular based on AI, that reduce the cost of monitoring online behaviour and content. However, the risk remains that providers, for fear of sanctions, unduly restrict users' freedom of expression and right to information. The complexities involved in detecting the different kinds unlawful and inappropriate behaviour require that machines complement, but not substitute human judgement, and that the circumstances of smaller providers are taken into account

The use of AI in the context of e-commerce is not limited to the processing of consumer data. It is also relevant in the context of the debate on how to upgrade the 2000 eCommerce Directive, in particular concerning obligations and liabilities of Internet platforms. One of the key points of the Commission's agenda is indeed to upgrade EU rules for digital platforms, services and products. As early as April 2020, the discussion over a new Digital Service Act has been set forth[42]. A probable target of the reform will be the eCommerce Directive[43], with regards to liability and safety rules for online platform. The intention is to address providers obligations in areas such as content moderations, political advertising online, and collaborative platforms.

The eCommerce Directive, exactly that 20 years ago, introduced in particular limitation for the secondary liability of host providers, namely, their liability for the unlawful behaviour of their users[44]. In particular, according to Article 14 a host provider is exempted from liability when

a) the provider does not have actual knowledge of illegal activity or information and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or information is apparent; or

b) the provider, upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information.

Another key rule of the Directive (Article 15) specifies that intermediaries may be ordered, by competent authorities, to terminate or prevent infringements by their users, but they may not be subject to any "general obligation to monitor the information which they transmit or store" nor to "actively to seek facts or circumstances indicating illegal activity".

Among the rationales for such provisions (which parallel the US 1996 Communication Decency and the 1998 Digital Copyright Act) the following can be mentioned: (a) supporting the development of Internet services, as liabilities may negatively interfere with the intermediaries' capacity to maintain and develop their activity; (b) preserving current business models, in particular the offer of "free" services, (c) avoiding that, for fear of liabilities, providers may impede or obstruct lawful activities of their users (infringing their freedom of expression and information)[45].

---

42 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data, COM(2020), 66 final.

43 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').

44 On liabilities of Internet service providers see Sartor (2017).

45 See Balkin (2014).

Today, in a radically changed economic and social context, there is the need to rethink the limitation to providers' liability, as these rationales and the ways to implement them have to reappreciated.

Firstly, some budding start-ups of 20 year ago are now leading global companies, having vast financial and technological resources and often enjoying a quasi-monopoly position. Thus, such companies could sustain the financial burdens for effective content moderations. In fact, while 20 years ago preventing unlawful activities by the users would require very costly human intervention, today new technologies, in particular based on AI, are considerably reduce such costs.

Secondly, it is now clear that the provision of "free services" may take place under different models. The for-profit two-side market business model, in which providers of free services are paid by advertisers have to be distinguished from no-profit models in which the provider's costs are covered by donations and voluntary work. When the provider's activity delivers profits, it makes sense to require that a part such profits can be devoted to covering the harm caused by the use of the platform.

These general considerations have different implications for different sources and conditions of liability, different kinds of sanctions, and different kinds of intermediaries. Secondary liability works better, as a mechanism for secondary regulation, for illegalities that can be detected with ease, cost-effectiveness and precision. This is the case, for instance, for communication to the public of copies of entire copyrighted works. The illegal nature of this activity is apparent, and software tools exist (content recognition systems), often also including technologies, that can detect what copies of a given work occur on a platform and recognise attempts to upload new copies.

Other instance of unlawful or harmful user-generated content raise very different concerns. Consider, for instance, defamation or hate speech. In such cases, it may be difficult for intermediaries to identify in advance — before receiving a notice — what messages by their users may be affected by these grounds of illegality. Even when the inquiry is focused on a specific message, consequent on a complaint, doubts may remain on whether the message is illegal, or whether, on the contrary, it is legal, and possibly even socially beneficial. The parties directly involved — on the one hand the issuer of the message and on the other hand the alleged victim of it — may have opposing views: the first may view the message as a way of expressing a legitimate opinion, advancing an individual interest or even a valuable social cause; the second may view the same message as being illegal and harmful. The intermediary concerned, in taking the decision to terminate or maintain access to the message has to act like a judge between two opposite parties, a judge whose interests are to some extent involved in the case (in particular, the interest in reducing its own potential liabilities). In such contexts, as we will consider above AI technologies should only operate to improve human oversight.

Liability risks may have a different impact on different kinds of intermediaries. It may be argued that big players will have less incentive to exceed in censorship than small players. In fact, the largest commercial intermediaries can effectively limit their legal risks by investing in legal reliable assessment processes and can absorb the cost of possible sanctions. Moreover, in some cases — when they consider that the communication at issue is more likely to be viewed as legal by the competent authorities — they might willingly accept the risk of the liabilities resulting from an unfavourable authoritative decision, in exchange for the possibility of obtaining a decision favourable to them. The latter decision will benefit such intermediaries not only relatively to the particular communication at issue, but also relatively to the many similar communications they are and will be enabling. Smaller intermediaries or those whose business model does not provide them with large resources, will have to take a much more cautious attitude, and acquiesce in removal requests. In addition, smaller providers may be unable to access the technological resources (data and software) that are needed to develop and deploy advanced tools for the detection of inappropriate or unlawful content.

11. ISSUES ON INTERMEDIARY LIABILITY

KEY FINDINGS

Among the innovations to be considered in this regard are the following: clarifying that liability limitations also apply to search engines, online repositories, and social networks; breaking down the distinction between active and passive intermediaries; clarifying that the scope of the prohibition on imposing on service providers a "general" obligation to engage in monitoring refers to the non-availability of effective (AI) technologies and to the possible impact on users' freedoms. The limitations on providers' secondary liability should not apply when providers have contributed to the unlawful behaviour of their users by failing to take reasonable measures that could have prevented that behaviour or mitigated its effects. This failure may also depend on not having adopted the most effective technologies.

In the context of the new situation described in the previous section, we may wonder whether the regulation of provider's liability should be reconsidered.

First, we might wonder whether the EU should also have a separate set of rules on providers' secondary liability for copyright infringements (as in the US model). In fact, many instances of copyright infringement can be more easily detected and assessed than other kinds of violations, while notice and action procedures for copyright violations are already in place in various EU countries. The 2019 Directive on Copyright in the Digital Single Market at Article 11, already requires providers hosting large amounts of content uploaded by their users to adopt measures in order to prevent the availability of such content, in collaboration with rightsholders.

Second, to be relevant today, the scope of any regulation of intermediaries need to go beyond the three categories should have a broad personal scope, including in particular the main content intermediaries of our times, namely, search engines, online repositories and social networks, still in their infancy in 2000.

Third, the practice and social function of content intermediaries in today's Internet ecology challenge the view that the passivity should be viewed as a necessary condition for liability limitations (a view that has been endorsed in some national case law). The idea that merely passive providers are to be covered by such limitations, to the exclusion of those providers who play an active role, should be overcome, since most on line platforms todays engage in the presentation and access to user generated content. Thus, the use of "active" AI tools by a provider — for searching, providing ads, filtering content, etc. — should not, as such, lead to the provider to being excluded from liability limitations

Liability limitations should be maintained whenever intermediaries, while not creating or selecting content, intervene actively, through automated or non-automated means, to enable and shape third party communications, meeting the preferences of its users. This includes cases where intermediaries in good faith prevents access to objectionable material or activity (the so-called good Samaritan clause), in order to provide a service that meets legal and social standards.

Fourth, doubts remain concerning the admissibility of broadly scoped orders to monitor an users' content, an issue that was recently addresses by the EU Court of Justice in the 2019 Glawischnig-Piesczek decision (Case C-18/18). It seems to me that the prohibition on overbroad blocking/removal order by competent authorities still makes sense. However it should be clarified that the excessive broadness does not just depend on the generality of the obligation, but rather on the technological

possibility of implementing the obligation in a sustainable and cost-effective way — which also depends as we shall see, on the available technologies — and on its impacts on users' rights.

Finally, it should be clarified that exemptions from secondary liability (with regard to civil liability, and proportionate administrative sanctions) no longer apply when it can be established that providers have violated their duties of care, i.e., that they have contributed to the unlawful behaviour of their users by failing to reasonable measures which could have prevented that behaviour or mitigated its effect[46]. The assessment concerning whether a provider fails to exercise due care by omitting measures that could prevented illegal behaviour should be based on several factors, such as:

- the gravity of the risk of unlawful user behaviour that the omission of the measures would entail, this risk including both the probability of such behaviour and the seriousness of the damage it may cause;
- the technologies that are available for implementing such measures;
- the economic sustainability of the measure, given the (socially beneficial) business model adopted by the intermediary;
- the way in which such measures may affect the rights and interests of the users of the intermediary, as well as social values.

The duties of care of intermediaries may indeed be viewed as pertaining to a general design responsibility, namely, to their responsibility for harms that could have been avoided by adopting a better technological and organisational design (including, for instance, state-of the art technologies for filtering or moderating users' generated content, in combination with human intervention)[47]. In particular, the ascription of a legal liability to intermediaries should require an assessment on the availability of effective and proportionate risk-reduction measures. In other terms, the liability of intermediaries for users' behaviour should presuppose that proportionate risk-reduction measures — e.g., screening or removing users content, monitoring users' behaviour, banning users from engaging in inappropriate behaviour, disabling or making less usable those features that may most probably be unlawful action, etc. — were available. The proportionality of such measures would consist in the fact that their implementation would not lead to harms — with regard to legally protected interests and values, such as free speech, right to information or even economic initiative — more important than the harms to third parties that such measures could successfully prevent. If the intermediary has omitted proportioned measure, it may indeed be claimed that the intermediary has exposed third parties to unnecessary risks, i.e., risks that were not necessary to achieve the benefits that the platform provides to its users and to society, and consequently, it may be liable for resulting harm.

[46] On duties of care of intermediaries, see Valcke et al., (2017).

[47] See Helberger et al (2018).

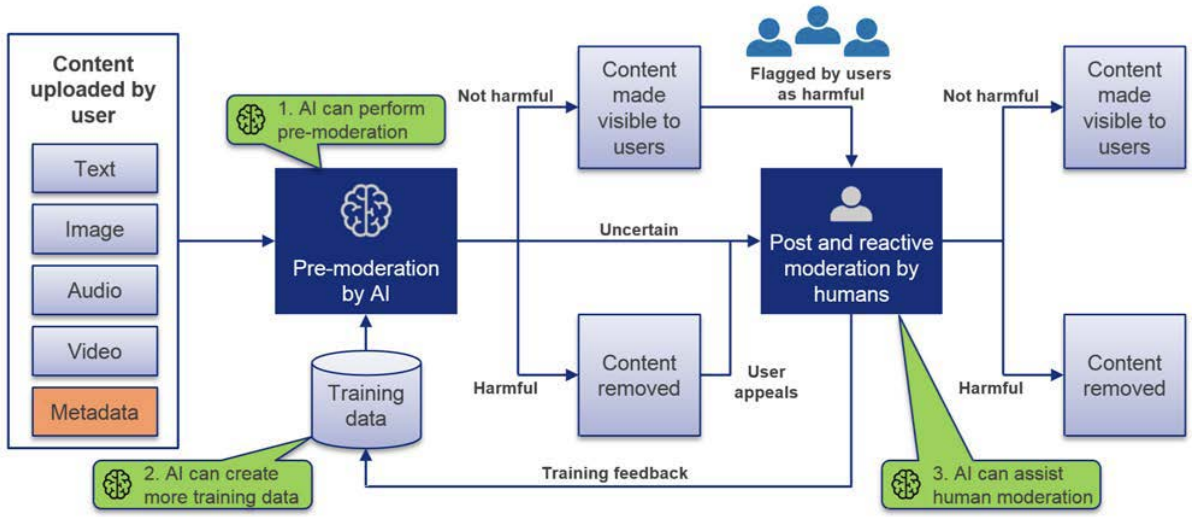
12. AI IN CONTENT FILTERING/MODERATION AND CONSUMER PROTECTION

KEY FINDINGS

AI can indeed provide effective filtering and moderation tools. These tools can improve the prefiltering/moderation stage and flag content for review by humans, increasing the accuracy of the moderation process. AI can also assist human moderators by increasing their productivity. It may also reduce the potentially harmful effects of content moderation on individual moderators. However, automated filtering/moderation should only be used in combination with human judgment, and the final decision to block any particular content or activity should in general be entrusted to the human in this loop. The use of such tools may become a necessary aspect of compliance, with an obligation of providers do exercise due care. Third-party filtering/moderation should be encouraged so as to broaden access, and so should the sharing of datasets (to train AI classifiers) and software, so that both are accessible to small companies as well.

The connection between AI technologies for filtering/moderation is relevant to consumer protection in online platforms, since such technologies determine what content the users of such platforms will be able to access, what content will they be exposed to against their preferences, and what content they will be able to post online and keep visible to others.

Figure 6: AI in the online content moderation workflow



Source: Cambridge Consultants 2019

Figure 5[48] shows the way in which AI can intervene in the process. The picture concerns moderation for harmful content — child abuse material; incitements to violence and terrorism; harm to people, suicide, self-harm, animal abuse; hate speech and harassment; pornography; bullying, aggressive/defamatory comments, etc. — but the same consideration can extend to other kind of unlawful content, such as in particular unauthorized distribution of copyrighted content.

AI can be used to improve the pre-moderation stage and flag content for review by humans, increasing accuracy of the moderation process. AI can also assist human moderators by increasing their

[48] From Cambridge Consultants (2019).

productivity. It may also be used to reduce the potentially harmful effects of content moderation on individual moderators (e.g., by hiding portions of picture or movies showing disgusting of atrocious acts).

Concerning the use AI tools to support online filtering/moderation, a careful approach has to be adopted, keeping in mind that in most cases only an adequate mix of human and AI technologies can provide the best results. Relaying only on technological solutions would lead over over-filtering (many false positive, i.e., rejected acceptable items) or to under-filtering (many false negatives, i.e., undetected inacceptable items). In fact, AI systems works as statistical machines, applying to new cases the solutions adopted relative to similar cases in the past, and may fail to appreciate what makes a difference to human understanding (e.g., the difference between insult and satire, or between terrorist content and videos documenting atrocities, between educational and unlawful uses of copyrighted material, between an apology of violence or torture and a speech condemning them, etc.). Thus, automated filtering/moderation should only be used in combination with human judgment, and the final decision to block a particular content or activity should be given to the human in the loop. Respect for the dignity and autonomy of users, as content originators and readers, also requires that they can contest the decisions of providers on the accessibility of such content. An appeal to human decision-makers should be made available to those who have posted content being banned or removed.

Even under these constraints, AI-enabled content filtering/moderation tools will play an important role in providers' response to harmful and unlawful online content. The use of such tools may indeed become a necessary aspect of compliance with the due care obligation of providers. Thus, access to such tools should and tools must be accessible to organisations of all sizes. The provision of filtering/moderation by third party should be encouraged, as well as the sharing of data sets (to train AI classifiers) and software, so that it both are also accessible to small companies.

Ideally, the duty of care of providers concerning the identification and removal of unlawful content should be understood as a best effort obligation. As with other risk mitigation measures as best effort obligations, where what is requested has to be proportional to the scale of the processing, and possibly also to the resources available to the controller. Certification and auditing of moderation systems may contribute to generate trust on such systems.

13. RECOMMENDATIONS

KEY FINDINGS

a) Consumers should have the option not to be tracked and (micro)-targeted6 and should have an easy way to express their preferences.

b) The grounds on which service providers and traders cannot price-discriminate should be specified.

c) It should be considered how discrimination in ad targeting is to be addressed.

d) Guidance should be given concerning what algorithmic practices count as instances of aggressive advertising.

e) Guidance should be given concerning cases in which consumers have a right to contest a decision that undermines their interests.

f) Consumers should be provided with information on whether and for what purposes they are tracked and on whether they are receiving information for advertising purposes.

g) Protection of consumer privacy requires preventive risk-mitigation measures in combination with collective redress.

h) The development of consumer-friendly AI-technologies should be encouraged and supported. Service providers should be prevented from blocking legitimate tools for the exercise of consumer rights.

i) Liability limitations for online providers should also apply to "active" providers, such as search engines, online repositories, and social networks, regardless of whether user-generated content is organised, presented and moderated by humans, by algorithms or both.

j) Limitations on providers' secondary liability should not apply when providers have failed to adopt take reasonable precautionary measures that could have prevented that behaviour or mitigated its effects. This failure may also depend on not having adopted the most effective AI technologies.

k) The availability of AI technologies for detecting unlawful online content and behaviour should be encouraged, in combination with human judgment.

l) Third-party filtering/moderation should be encouraged so as to broaden access, and so should the sharing of datasets (to train AI classifiers) and software, so that both are accessible to small companies as well.

In this section I will select some policy issues that emerge out of the scenarios presented above[49].

[49] This analysis is based on Jabłonowska et al 2018.

a. Should consumers have the option not to be tracked and (micro)-targeted?

As noted in Section 5, it is unclear to what extent consumers have a right to enjoy "free" online services without being tracked and targeted with commercial information. It is also unclear to what extent providing personal data and being tracked can count as consideration for money or services. Clarity would benefit consumers, who in situations of uncertainty are in difficulty, given their limited bargaining power.

To enhance consumer privacy, it should be specified that consumers always have the right not to be tracked unless this is needed for a service they specifically request. It should also be clarified that consumers should maintain access to "free" services, paid by advertising, even if they refuse to be tracked. In this case they may be exposed to non-targeted ads (context advertising).

As noted above, consumers are often induced to consent to unwanted processing or are prevented from adopting privacy-preserving options by misleading interfaces. "Design tricks" and "dark patterns" serving that purpose should be prohibited. Consumers should be offered clear and standardised ways in which to communicate their choices concerning the processing of their data. Refusing unnecessary processing of personal data should not be more difficult or require more steps that accepting such processing. Privacy-preserving options should ideally be the default whenever consumers have the right to choose such options.

b. Should price discrimination be outlawed, and in what contexts?

The application of different prices to different consumers is not as such contrary to consumer protection law. However, in the context of AI and big data it becomes problematic, since it can be deployed on a much larger scale, and in a systematic way, profiting from vast knowledge about consumers, their preferences, their weaknesses, and their misconceptions. If algorithmic pricing should not be prohibited outright, at least the grounds under which it is impermissible should be specified. More generally, if consumers are to be protected from being unfairly or in any way disadvantageously treated because of their individual characteristics, they could be granted a right to engage in commercial transactions anonymously and be identified only when all elements of their transactions have been determined.

c. Should discrimination in commercial offers be addressed?

AI and big data offer the opportunity to engage in granular mass differentiation of individual consumers. In some cases the unfavourable differential treatment of consumers, based on their individual features, may count as an instance of unlawful or unfair discrimination. Discrimination may pertain not only to the content of offers or to the rejection of applications, but also to the selection of the individuals to whom ads are addressed, or rather in the exclusion of certain individuals and groups from certain offers (e.g., job or loan offers). This selection may result from the working of AI systems, even when they are not engineered to achieve discriminatory goals, but rather pursue the goal of optimising the uptake of ads on the basis of the information to which they have access. Lawmakers need to consider whether consumer protection or antidiscrimination law should also address this new kind of discrimination (which was impossible when ads could only be aimed at the general public).

d. Should the concept of aggressive and misleading commercial practices be updated?

The Unfair Commercial Practices Directive[50] contains broad notions of what counts as a misleading action, a misleading omission, or an aggressive commercial practice, which are clearly also applicable

[50] Directive 2005/29/EC of the European Parliament and of the Council concerning unfair business-to-consumer commercial practices in the internal market (Unfair Commercial Practices Directive).

when the action, omission, or practice results from the deployment of an automated system, and even when the system itself has learned that behavior. However, to provide guidance to traders, consumers, and decisionmakers, exemplification of how this may apply to AI-based algorithmic practices may be useful. In particular, with regard to aggressive practices, it should be considered when the use of certain information (e.g., about the data subject's vulnerabilities) may lead to an "undue influence" affecting consumers' freedom of choice (Article 8 of the Directive).

Under the GDPR, the processing of personal data is subject to a standard of necessity, minimization, and proportionality. We may wonder whether, regardless of the data subject's consent, certain kinds of data about consumers may be legitimately collected of inferred for the purpose of advertising and consumer management. This concerns, for instance, the use of face recognition to recognise individual consumers in public spaces or of psychographic methods, to determine consumers' psychological attitudes and weaknesses on the basis of their behaviour.

e. Should consumers have the right to human intervention/contestation?

The GDPR, at Article 22, provides for a right to human intervention and contestation in the case of automated decisions having legal effect or significantly affecting the data subject. This also applies to consumers. It should be clarified to what extent this applies to cases in which consumers are subject to a decision that undermines their interests (e.g., being denied a loan, being excluded from a social network, being systematically subject to substandard treatment)

f. Should consumers have a right to transparency over profiling and automated decision-making

Consumers may be tracked, profiled, and nudged into making purchases (by being fed selected information or being presented with a curated menu of options), but they are not necessarily aware of that fact. They should have a right to know whether they are being tracked, and whether the information pushed onto them and the offers they are presented with are based on commercially driven profiling. A new aspect of the right to information is provided by Article 7(4a) of Directive 2011/83/EU on consumer rights (as amended by Directive 2019/2161): consumers must be informed of the "main parameters determining the ranking of products presented to the consumer as a result of the search query and the relative importance of those parameters, as opposed to other parameters." The issue of information on consumers is also addressed in the same Annex I at point 11a of Directive 2005/29/EC (as amended by Directive 2019/2161), which extends the list of unfair commercial practices to include search results presented "in response to a consumer's online search query without clearly disclosing any paid advertisement or payment specifically for achieving higher ranking of products within the search results." This approach should be expanded to other cases in which information is selectively pushed by providers onto consumers (rather than pulled by the latter) so as to induce a certain purchasing behaviour.

g. What preventive measures should be adopted?

The GDPR provides for preventive measures to anticipate and mitigate data protection risks. It does so under the heading of privacy by design and by default. It requires (Article 35) that the controller carries out a data protection impact assessment when the processing is "likely to result in a high risk to the rights and freedoms of natural persons." It also requires prior consultation with a competent data protection authority (Article 35) when "the processing would result in a high risk in the absence of measures taken by the controller to mitigate the risk". It should be clarified what risk-preventing measure are required in processing consumer data, and in particular under what conditions the risks is such as to require a data protection impact assessment and prior consultation.

The proposed directive on the collective redress of consumers will enable consumer organisations[51] to undertake representative actions for the protection of the collective interests of consumers. This is much needed to ensure effective consumer protection relative to violations related to the use of big data and AI, given the imbalance in knowledge and power between individual consumers, on the demand side, and service providers or sellers/vendors, on the supply side.

h. Should the development of consumer-friendly AI-technologies be supported?

As argued above, consumer-friendly AI technologies could support consumers and their organization in accessing the market, resisting unwanted tracking and data collection as well as in detecting unlawful or unfair behavior and responding to it. Public institutions could support these initiatives by financing projects and providing incentives. It should also be clarified that when consumers have a right not to be tracked, service providers and retailers/vendors/sellers should not block the use of antitracking tools. AI tools supporting consumer may contribute to the achievement of the goal of Article 5 of Directive 2019/2161, under which the Commission should provide citizens with information on consumer rights, and consumers should have the ability to file complaints through the online dispute-resolution platform established under Regulation (EU) No 524/2013, and also to file such complaints with the competent centre of the European Consumer Centres Network.

i. Should liability limitation for online providers be updated, and applied also to "active" providers

Liability limitation for online providers should be maintained, so as to ensure a uniform legal discipline across the EU. However, it should be clarified that such limitations also apply to "active" providers, such as search engines, online repositories, and social networks, regardless of whether user-generated content is organised, presented and moderated by humans or by (AI) algorithms or both.

j. Is AI relevant to the revision of rules on the liability of Internet service providers?

Limitations on providers' secondary liability should not apply when providers have failed to take reasonable precautionary measures that could have prevented that behaviour or mitigated its effects. In this connection, the availability of AI becomes highly relevant: AI enhances providers' ability to identify, and respond to, inappropriate and unlawful online activities. Therefore, the availability of AI may expand the scope of providers' obligation to take reasonable due care to prevent and respond to unlawful behaviour by their users.

The availability of AI solutions should lead to more stringent duties on care of on providers. This, however, should be a best-effort obligation, taking into account the state of the art, as well as the scale of the processing at stake and the financial and technological resources available to providers. Users of platforms should be granted a right to object to decisions affecting the accessibility of content and should be able to obtain human intervention for the final verdict on such objection.

k. Should AI technologies be used to detect and respond to inappropriate and unlawful content?

AI technologies will play a key role in identifying unlawful and inappropriate online content. They can contribute to detect critical content and activities, and support the assessment of human moderators. Their use should be encouraged, though a final human assessment should be required before blocking activities or removing content, unless the case is such that an automated response may ensure

[51] Proposal for a directive of the European Parliament and of the Council on representative actions for the protection of the collective interests of consumers, COM/2018/0184 final - 2018/089 (COD).

sufficient accuracy. In fact, AI performance to this effect differs in different domains. For instance, it is more accurate in detecting copyright infringement than in engaging with hate speech and defamation.

l. Should the use of AI technologies for content moderations be promoted?

Initiatives should be taken to make AI tools for content moderation available to smaller companies as well, by promoting the third-party provision of moderation services, and by encouraging the sharing of datasets and the development of accessible software. The sharing of datasets (to train AI classifiers) should also be supported as well as the development of open-source software, so that both data and basic software tools are accessible to small companies as well.

REFERENCES

- Balkin, J. M. (2014). Old school/new school speech regulation. *Harvard Law Review,* pp. 2296–2342
- Bar-Gill, O. (2019). Algorithmic price discrimination when demand is a function of both preferences and (mis)perceptions. *The University of Chicago Law Review 86*, pp. 217–254. Barocas, S. and A. D. Selbst (2016). Big data's disparate impact. *California Law Review 104*, pp. 671–732
- Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedoms*, Yale University Press
- Borgesius, F. J. Z., S. Kruikemeier, S. C. Boerman, and N. Helberger (2018). Tracking walls, take-it-or-leave-it choices, the GDPR, and the ePrivacy regulation. *European Data Protection Law Review 3*, pp. 353–368
- Burr, C. and N. Cristianini (2019). Can machines read our minds? *Minds and Machines 29*, pp. 461-494
- Calo, R. (2013). Digital market manipulation. *George Washington Law Review*, 82:995
- Castells, M. (2001). *The Internet Galaxy*. Oxford University Press
- Cambridge Consultants (2019). Use of AI in online content moderation. 2019 report produced on behalf of OFcom
- Cohen, J. D. (2019). *Between Truth and Power. The Legal Constructions of Informational Capitalism*. Oxford University Press
- Edwards, L. and M. Veale (2019). Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke Law and Technology Review, pp. 16- 84*
- Hagiu, A. (2009). Two-sided platforms: Product variety and pricing structures. *Journal of Economics and Management Strategy 18*, pp. 1011–1043
- Helberger, N. (2016). Profiling and targeting consumers in the internet of things – a new challenge for consumer law. In R. Schulze and D. Staudenmayer (Eds.), *Digital Revolution: Challenges for Contract Law in Practice*. Nomos
- Helberger, N., J. Piersonb, and T. Poell (2018,). Governing online platforms: From contested to cooperative responsibility. *The Information society 34*, pp. 1–14
- Jabłonowska, A., M. Kuziemski, A. M. Nowak, H.-W. Micklitz, P. Pałka, and G. Sartor (2018). Consumer law and artificial intelligence. Challenges to the EU consumer law and policy stemming from the business's use of artificial intelligence final report of the artsy project1. Technical report, European University Institute
- Kleinberg, J., J. Ludwig, S. Mullainathan, and C. R. Sunstein (2018). Discrimination in the age of algorithm. *Journal of Legal Analysis 10*, pp. 113–174
- Lanier, J. (2014). *Who owns the future*. Simon and Schuster
- Lanier, J. (2018). *Ten Arguments for Deleting your Social Media Accounts Right Now*. Holt
- Lessig, L. (2001). *The Future of Ideas: The Fate of the Commons in a Connected World*. Random House
- Mik, E. (2016). The erosion of autonomy in online consumer transactions. *Law, Innova- tion and Technology 8*, pp. 1–38

- Norwegian Consumer Council (2018). *Deceived by design. how tech companies use dark patterns to discourage us from exercising our rights to privacy*
- OCDE (2010, April). *The Economic and Social Role of Internet Intermediaries. Report DSTI/ICCP(2009)9/FINAL*
- Pariser, E. (2011). *The Filter Bubble.* Penguin
- Rochet, J.-C. and J. Tirole (2003). Journal of the European economic association. *Journal of the European Economic Association 1*, pp. 990–1029
- Sartor, G. (2017). *Providers Liability: From the eCommerce Directive to the future. In-Depth Analysis for the IMCO committee.* European Parliament
- Sunstein, C. R. (2001). *Republic.com.* Princeton University Press
- Valcke, P., A. Kuczerawy, and P.-J. Ombelet (2017). Did the romans get it right? what Delfi, Google, eBay, and UPC Telekabel Wien have in common. In L. Floridi and M. Taddeo (Eds.), *The Responsibility of Online Service Providers*, pp. 101–15. Springer
- Wachter, S. and B. Mittelstadt (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review,* pp. 1–130
- Yeung, K. (2018). 'Hypernudge': Big data as a mode of regulation by design. *Communication and Society 20*, pp. 118–36
- Zuboff, S. (2019). *The Age of Surveillance Capitalism.* Hachette

The study addresses the new challenges and opportunities for digital services that are provided by artificial intelligence, in particular which regard to consumer protection, data protection, and providers' liability. The discussion addresses the way in which digital services rely on AI for processing consumer data and for targeting consumers with ads and other messages, with a focus on risks to consumer privacy and autonomy, as well as on the possibility of developing consumer-friendly AI applications. Also addressed is the relevance of AI for the liability of service providers in connection with the use of AI systems for detecting and responding to unlawful and harmful content.

This document was provided by the Policy Department for Economic, Scientific and Quality of Life Policies at the request of the committee on the Internal Market and Consumer Protection (IMCO).

PE 648.790
IP/A/IMCO/2020-14

Print ISBN 978-92-846-6554-9 | doi:10.2861/910168 | QA-02-20-298-EN-C
PDF ISBN 978-92-846-6553-2 | doi:10.2861/56468 | QA-02-20-298-EN-N