Outsourced justice: The case of the Facebook Oversight Board

Riku Neuvonen

University of Helsinki and Tampere University

CONTRIBUTOR DETAILS

LL.D. Riku Neuvonen is a senior lecturer of public law at the Tampere University and associate professor of media law at the University of Helsinki. He has worked as researcher and teacher in Finnish universities and for the Finnish government as an expert. Neuvonen has been involved in several national and international research projects and various networks. He has published ten books and hundreds of articles.

https://orcid.org/0000-0003-1722-461X

Esa Sirkkunen

Tampere University

MSocSc, researcher Esa Sirkkunen works at the Research Centre COMET. His research interests include journalism research, journalism ethics, social media research, platform regulation and social theory. Sirkkunen has published over 70 books, articles and reports and led and participated in more than twenty research projects at the Centre.

https://orcid.org/0000-0003-1243-1740

Received 25 March 2022; Accepted 21 July 2022

Abstract

In this article, we explore the possibilities for the self-regulation of online platforms, here by using Facebook's Oversight Board (OB) as an example. First, we analyse and systematize how the OB fits in the mosaic of internet regulation. Our analysis shows that the OB has tried to lay the foundation for global self-regulation, but because of its limited jurisdiction and indicative nature, it falls short of becoming a real 'supreme court' of Facebook. In addition, although the OB is a positive attempt to deal with many problems, it does not seem to be able to process enough cases, relies on idiosyncratic standards instead of general rules and principles and has problems deciding which human rights principles it should follow. Additionally, the OB is not compatible with the Digital Services Act (DSA) of the European Union or with the recent initiatives for social media councils.

Keywords: platforms, platformization, regulation, freedom of speech, communicative rights, digital rights, social media, rule of law, Digital Services Act

Introduction

The early internet contained many small websites that were mostly controlled by developers and users (Zittrain 2006). The gap between administrators – who were often the owners – and moderators and users was relatively small. A fundamental change was marked by the rise of social media platforms (DeNardis and Hackl 2015). The number of users increased, and the gap between users and moderators grew. The moderators were no longer peers but employees of these platforms which requires stricter self-regulation. The new business model for social media is based on selling advertisements and keeping users on platforms (Gillespie 2018; Zittrain 2008). This requires not only the moderation of harmful content but also other ways to manage content and keep users engaged. As a result, the moderation of big platforms needs resources and rules that are developed in alignment with the business model (Roberts 2019).

In the platformization process (Poell et al. 2019), social media sites have become an infrastructure for social interaction, communication and self-expression – seemingly free of charge for users. The platform business is based on keeping users in a company's ecosystem to gather as much data from them as possible (van Dijck et al. 2019) and to then build user profiles and target advertising to these users. In essence, the users are paying for the use of the platforms with their data without having the right to decide how their data are used. In this process, we can see so-called surveillance capitalism (Zuboff 2015) in action.

The fundamental problems with platformization, such as the violations of the privacy of users, the dissemination of unlawful and harmful content and the political manipulation of selected groups of users – for example, during the US presidential elections in 2016 and 2020 – have become more evident. It has also been shown that Facebook's problems in handling moderation are crucial, especially in language areas other than the English-speaking world (Wijeratne 2020; Roberts 2019). For example, Facebook was not capable of regulating Spanish-language disinformation about COVID-19 (Paul 2021). Similarly, Facebook has its history of moderation scandals, from banning the Terror of War (i.e. the napalm girl) picture to removing breastfeeding images (Gillespie 2018). Furthermore, automated, algorithmic tools have created problems because of their inability to detect and identify illegal or harmful content because of a lack of understanding of the social context of the expressions. On the other hand, the pressure to remove illegal or harmful content has compelled platforms to remove even the most slightly suspicious content, thereby needlessly harming rule-abiding users' freedom of speech. Users have also expressed a constant critique that appeal processes have been slow and haphazard, if available at all.

Moderation is only part of platforms' content management – in other words, curation (Klonick 2018; Gillespie 2018). Curated content helps keep users engaged on platforms. Especially controversial content, such as QAnon-related narratives around 2021, can create debates and keep users engaged. More engaged users mean more users who see advertising and spend more time and do more activities, which means more data can be gathered on users. Essentially, the platforms' incentive to manage content differs radically from that of their users or the requirements of the public good.

The problems are clear, but developing remedies through regulation has been difficult. Currently, internet regulation is a patchwork of laws and treaties that partially contradict each other.

The internet is global, but its regulation is bounded by jurisdictions that are national or regional (e.g. the European Union). The only truly global regulators of the internet are self-regulatory

bodies, such as ICANN or W3C (Kettemann 2020; Maroni 2022); these bodies regulate the internet's critical infrastructure but only at a minimum level. Therefore, there are few global rules or principles for content regulation or frameworks to promote cooperation between regulators (Radu 2019: 90–92). The only area in which most industrialized nations have come to an agreement is cybercrime. Here, the United States, Canada, South Africa and Japan, among many others, signed the Convention of Cybercrime, which was adopted by the Council of Europe in 2001. However, this convention focuses only on child pornography and copyright infringements (Freedman and Rorive 2002).

Despite the lack of clear and enforced global rules, many platforms have highlighted self-regulation as a solution to regulative problems (Jørgensen and Pedersen 2017). Facebook's Oversight Board (OB) is perhaps the most prominent attempt to formulate a 'supreme court' for evaluating the content decisions of one company. In the current article, we carry out a policy analysis on how the OB fits into this obscure patchwork of internet regulation and how it could fulfil growing regulatory demands. First, we explore the evolution of internet regulation. Second, we place the OB in the framework of platform regulation and analyse and systematize its functions as part of the self-regulation of the internet. Third, we draw conclusions about whether the OB is able to keep its promises, especially in the context of recent EU regulation and other new proposals for self-regulation, for example, social media councils (SMCs). Especially the Discussion section we compare the various legal proposals of European Union and the OB's status and operating methods to identify possible problematic and conflicting areas. Our research method is careful close reading of the existing legislation, various legislative bills and the documents and reports that Facebook or OB has published. Our viewpoint is regulatory, combining arguments from policy and legal analysis. The key question is why, from the regulatory angle, there is a need for organs like the OB and how well the OB functions as part of contemporary regulation of the internet and platforms.

History: First, there was an ethos of freedom

The early internet was formed by many small websites, Usenet newsgroups, discussion forums and web communities; in the beginning, social media were often understood as a place for more extensive freedom of speech. The ethos of freedom in the early internet also affected regulations, and most of the owners and leaders of big tech companies were influenced by the American approach to fundamental rights. In the United States, the Information Technology Act of 2000 (earlier, the Communications Decency Act 1996), Section 230, guarantees an exemption of liability to intermediaries from third-party acts if the intermediary acts in good faith. The US court went even further and granted immunity, even if the host is aware of the unlawful nature of the content (Freedman and Rorive 2002).

In the European Union, Article 14 of the E-Commerce Directive of 2000 (ECD) states that intermediaries – or digital or online platforms – are not legally responsible for hosting illegal content. Only the courts can order intermediaries to remove illegal content; however, what circumstances constitute social media platforms as intermediaries is unclear. At this stage, the pressure to regulate the internet was fragmented into the fields of antipiracy, competition and, for example, child pornography. These were the areas in which it was possible to achieve supranational regulation. Similarly, legal studies were first focused on the jurisdictions and competence of regulators, as well as the liabilities of intermediaries (Pollicino 2021). This optimism was grounded based on the fact that several content providers were stable software companies (Marsden 2011).

Therefore, the internet was a dynamic and innovative environment, which remained the case until tech companies grew to become massive corporations.

In the early days, there were fewer internet users than today. Illegal activities were mainly piracy and the dissemination of questionable content. Hate speech, misogyny, fake news or defamation were common phenomena in digital environments like Bulletin Board Software (BBS) or news groups. These environments were quite open and fragmented, meaning there was simultaneously more competition and choices for users but also difficulties in regulating and monitoring this content. The second wave of debate – in the 2010s – discussed the rights of users and even digital constitutionalism, which were both affected by a rise in the importance of privacy (De Gregorio 2022; Pollicino 2021).

However, the rise of social media giants as we know them in the 2020s has diluted the principles of an open environment and free speech. Big tech companies have created their own ecosystems in different areas of the digital world. Content, software and infrastructure are increasingly exclusive and only available from certain service providers (Zittrain 2008). Therefore, the internet and different services it comprises are controlled by a limited number of gatekeepers.

More recently, freedom of speech has been used as a justification against piracy or privacy-related regulation, as well as to justify the self-regulatory actions of platforms (De Gregorio 2021). Over the past decade, the role of social media has been especially emphasized in debates about free speech. Now, in the 2020s, at the dawn of web 3.0, the focus of the discussion has moved more to structures, the transparency of platform activities and the liability of platforms in general.

At present: A mosaic of norms and regulations

One basis of social media jurisdiction is the contract between the user and company that is providing the services (Radin 2004; Bygrave 2015). By accepting the terms of service, users have transferred most of their rights to platform companies. The contract is at the core of private law, whereas the protection of human rights is within the realm of public law, as well as some parts of consumer protection and competition law. It is still unclear whether private companies could be held responsible for violating users' human rights and other rights under international law (Callamard 2019). Therefore, as of now, human rights standards cannot be considered a part of the binding legal norms of platforms in a contractual relationship.

Nevertheless, human rights treaties have had some influence on the development of the regulatory framework of the internet. In particular, UN special advisers have given statements on how the internet should be governed based on global human rights (Kaye 2016). This approach has been criticized as an attempt to secure a liberal interpretation of human rights on the internet and, therefore, reduce the possibilities for regulation (Maroni 2022). It is also clear that human rights should be applied on the internet and that what is illegal outside the net is also illegal on the internet (Kettemann 2020). The crucial issue is how these norms are monitored and enforced in digital environments.

At the regional level, the European Convention on Human Rights (ECHR, drafted in 1950) is one of the oldest such regional treaties and is interpreted by the European Court of Human Rights (ECtHR). Additionally, since the adoption of the EU Charter of Fundamental Rights of the European Union, the European Court of Justice (ECJ) has played a role in setting human rights standards in Europe. In the Americas, the Organization of American States has established the Inter-American Court of Human Rights to interpret the provisions of the American Convention on

Human Rights (ACHR). The role of the American court is more adjudicatory and advisory than decisive. It should also be noted that the United States and Canada are not members of the ACHR or the Inter-American Court. The United States constitutes one regional actor, but most of the major internet platforms are from the United States. Therefore, both commercial and content regulations in the United States have global effects.

The growing significance of platforms has increased the pressure to address unclear regulations (Tambini and Marsden 2007). Solutions to this situation in Europe have been taken at both the national and regional levels. One of the first national laws was Sweden's Electronic Bulletin Boards Responsibility Act of 1998. A more recent example is Germany's Network Enforcement Act (Netzwerkdurchsetzungsgesetz, NetzDG) of 2018. Both national laws reference criminal law and oblige platforms to moderate and remove content. In addition, the NGO 'Freiwillige Selbstkontrolle Multimedia-Diensteanbieter' (FSM), which has been set up by YouTube and Facebook, has been certified as a self-regulation institution under the NetzDG by the Federal Office of Justice. Here, platforms can ask the FSM to decide on content removal cases in Germany (Holznagel 2022).

The European Commission has stimulated the voluntary removal of content in, for example, the 2016 Code of Conduct on Hate Speech, the 2017 Communication on Tackling Illegal Content and the 2018 Recommendation on Measures to Effectively Tackle Illegal Content Online. In 2019, the European Parliament also adopted a report (COM/2018/640) pushing for content monitoring to be outsourced to hosting services under the pretext of the fight against terrorism. All of these regulations gave companies the competence to review the legality of content, at least to some extent. In addition to private law, the nature of a contract between the user and company has led to a situation where companies can remove content and ban users under the umbrella of legality. However, to some extent, the national courts in Europe, especially in Germany, have made decisions in which the platform's decision to remove content or ban users was considered unfair in terms of freedom of speech (Holznagel 2022).

The Digital Services Act (DSA) and the Digital Market Act (DMA) are the next steps in the move from soft regulation via communication or codes to harder regulation via more binding regulation.

These proposals, which will enter into force in 2024, suggest special obligations for large gatekeepers such as Facebook; these duties include tightening current moderation practices. New organs have also been suggested as a way to handle the enforcement of these new rules. National digital services coordinators (DSCs) in the EU countries will oversee this enforcement and will have the power to investigate, fine and impose restrictions on platforms, as well as to coordinate with other DSCs to conduct cross-border investigations. High penalties have also been created for sanctioning non-compliance with the DSA, which may result in fines of up to 6 per cent of a service provider's or platform's global annual revenues. The failure to provide information or submit to an inspection can result in a fine of 1 per cent of a service provider's or platform's annual turnover. The DMA should offer consumers the choice to use the core services of big tech companies, such as browsers, search engines or messaging, all without losing control over their data. The approach of the DMA is more competition and market based, whereas the DSA focuses on content regulation and services.

The DSA distinguishes between very large platforms (45 million European users), other platforms and small platforms. The small platforms are exempt from most obligations – for example, setting up compliance mechanisms. The DSA focuses on illegal content, but in principle, its provisions can be applied to other harmful materials as well. Notifications of illegal or even harmful content must

proceed, and the competent authorities should be informed. This process could solve one of the main problems of platform regulation: the platforms' total power to refuse to delete or republish already deleted content.

Article 18 of the DSA will allow for the certification of out-of-court dispute settlement bodies. Anyone who can show independence and expertise in content moderation matters can apply to be certified by authorities. Once certification has been granted, users can request that the body review their dispute over a moderation decision. It is noteworthy that most of the platform companies are American and that the US regulation doctrine is quite binary. Therefore, similar attempts at coregulation or involuntary voluntary self-regulation are constitutionally impossible in the United States because of the freedom of speech doctrine and because of the antitrust regulation, in which coregulation could be seen as an obstruction for competition cases (Marsden 2011). In the United States, regulation comes in the form of hard regulation based on laws or is self-regulation in purest form without (federal) state activity.

The DSA protects freedom of speech, but it is still a matter of open debate as to how it will guarantee other human rights. According to the DSA, member states must establish independent dispute settlement bodies. Whether each member state should establish its own body or cooperate with others is unclear. Although these bodies' functions are extrajudicial, members should be legal experts.

The birth of the OB

As recently as 2009, Mark Zuckerberg invited Facebook users to vote on some features of the terms of service, but the attempt turned out to be unrealistic (Suzor 2018). The threshold of users for a valid result was unrealistic, and it is unclear whether it was a genuine attempt to hear users' concerns. However, this implies that there has been a trend on Facebook to seek out different forms of legitimation and justification, including from the user level.

In 2018, Zuckerberg stated that Facebook was looking to create some kind of structure, almost like a supreme court, to make the final judgement calls on appeals (Klonick 2020: 2425). It is noteworthy that before 2018, Facebook only had an appeals system for suspended accounts and removed pages, not for single content items. Therefore, Facebook and Instagram have only had a broader appeal mechanism for a short time. In addition, the competence of the OB does not include commenting or evaluating the algorithms that organize and display user content to other users.

The major platforms already started to publish transparency reports on the information requests, removal requests and government requests for access to user data to add transparency and accountability in their inner processes (more e.g. in Puddephatt 2021). They have also started to develop tentative self-regulative practices. The OB can be considered an articulation of Facebook's intentions to meet the claims of accountability through self-regulation. Facebook (Meta) committed $130 million to fund the OB's costs for the first six years (Culliford 2019). A trust and the OB Limited Liability Corporation were founded to govern the OB. The OB is set in the charter, and its detailed procedural norms are called the bylaws. The OB has a membership committee that works together with Facebook (Meta) to interview and recruit new members. Facebook selects cochairs of the OB who will act as officers of the OB. Recommendations are operated by the US law firm Baker McKenzie. The idea is that members represent several regions: the United States and Canada, Latin America and the Caribbean, Europe, Sub-Saharan Africa, Middle East and North Africa, Central and South Asia and Asia Pacific and Oceania.

The OB was launched with relatively high media attention in 2020, and its members include several well-known politicians, journalists and experts in international law, such as former Prime Minister of Denmark Helle Thorning-Schmidt, former Guardian editor-in-chief Alan Rusbridger, PEN America chief executive officer Suzanne Nossel and Nobel Peace Prize Laureate Tawakkol Karman.1 In the spring of 2022, the OB has 23 members, but it is planned to have 40 members from around the world when fully staffed. As of the spring of 2022, six members are from the United States, and some other members have backgrounds or other connections with US universities or other institutions. Three members represent Europe, and one member represents both Cameroon and France. Other regions are represented as follows: one Mexico, one North Africa, three and half Sub-Saharan Africa and Middle East, two South America, one Israeli, one Australia and four Asia.

On its website, the OB presents itself as follows:

The board uses its independent judgment to support people's right to free expression and ensure that those rights are being adequately respected. The board's decisions to uphold or reverse Facebook's content decisions will be binding, meaning Facebook will have to implement them unless doing so could violate the law.

(Oversight Board 2022)

A little later, its profile is presented somewhat differently:

The board is not designed to be a simple extension of Facebook's existing content review process. Rather, it will review a select number of highly emblematic cases and determine if decisions were made in accordance with Facebook's stated values and policies.

(Oversight Board 2022)

How the OB works

The cases are selected by the case selection committee. Once a case is selected, it will be assigned to a board panel of five members. One member is assigned randomly from among those who are from the region of the case, and the other four are assigned randomly from all members (bylaws 31.3). According to the first transparency report from the three quarters between October 2020 and the end of June 2021, the OB received more than half a million appeals from users (Oversight Board 2021c).2 Out of this number, the OB selected 21 cases to review and was able to come to a decision in eleven cases, overturning Facebook's decision eight times and upholding it three times (Oversight Board 2021c). As part of these decisions, the OB made 52 recommendations to Facebook. The OB also sent 156 questions about decisions to Facebook, out of which Facebook answered 130, partially answered twelve and declined to answer fourteen (Oversight Board 2021c).

According to the report, two-thirds of the appeals were related to hate speech (36%) or bullying and report harassment (31%). The rest of the appeals dealt with violence and incitement (13%), adult nudity and sexual activity (9%) and dangerous individuals and organizations (6%) (Oversight Board 2021c). By the end of June 2021, the OB estimates that nearly half of the cases submitted (46%) came from the United States and Canada, while 22% of cases came from Europe, 16% from Latin America and the Caribbean, 8% from the Asia Pacific and Oceania region, 4% from the Middle East and North Africa, 2% from Central and South Asia and 2% from Sub-Saharan Africa (Oversight Board 2021c). These figures of submitted cases show a strong bias towards North America and Europe because India has the most Facebook users, with over 349 million, followed

by the United States (194 million), Indonesia (143 million), Brazil (127 million) and Mexico (96 million) (Statista 2021).

The OB also brought out a special case of failed moderation practices: Facebook's cross-check programme (Oversight Board 2021c). The Wall Street Journal revealed in September 2021 (Horwitz 2021) that the programme shielded millions of VIP users – including Donald Trump – from Facebook's normal content moderation rules. In a Special Section of the report, the OB openly criticizes Facebook for being 'not fully forthcoming' regarding its cross-check system. The criticism caused Facebook to ask for an advisory opinion on cross-check from the OB, which the OB has agreed to give. The OB has also reacted to other public criticism around Facebook. In October 2021, the OB announced that it had invited Frances Haugen, a former employee of Facebook who leaked a massive amount of data on Facebook's misconduct to the media, to discuss the company's ethical state (Oversight Board 2021b). These incidents show that the OB is ready to criticize and challenge Facebook (Meta) in public and to establish a relatively independent position towards the company.

Since the OB has not published transparency reports after the third quarter of 2021, we analysed the 23 decisions made by the OB up to February 2022, taking into account the community rules on which they were based (Figure 1), how the global regions were targeted (Figure 2) and how the number of decisions was split between Facebook and Instagram (Figure 3).

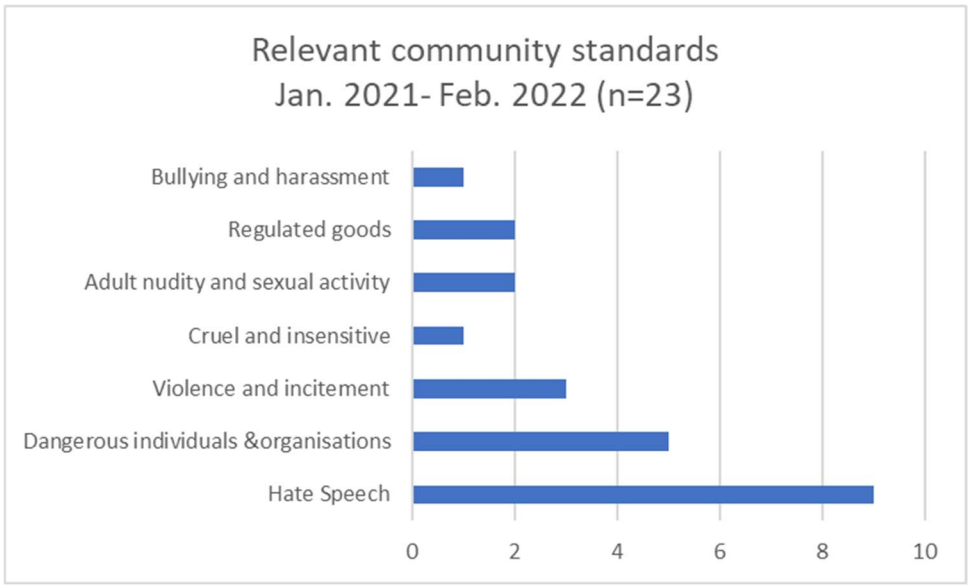

Figure 1: Hate speech is the most common standard used in nine decisions, following dangerous individuals and organizations used in five decisions, with the rest of the standards used in one to three cases.

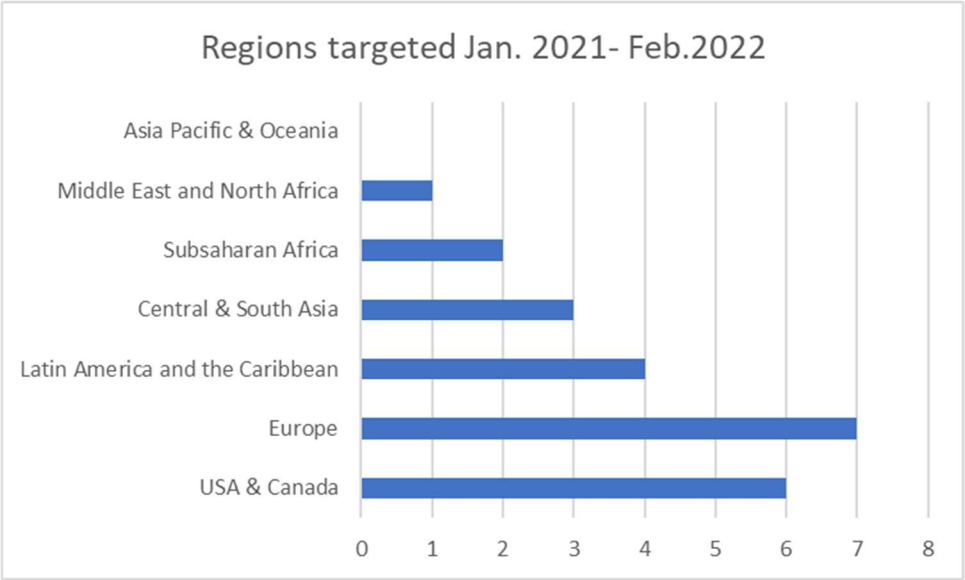

Figure 2: The regions targeted the most in the decisions are Europe (7) and the United States and Canada (6), followed by Latin America and the Caribbean (4).

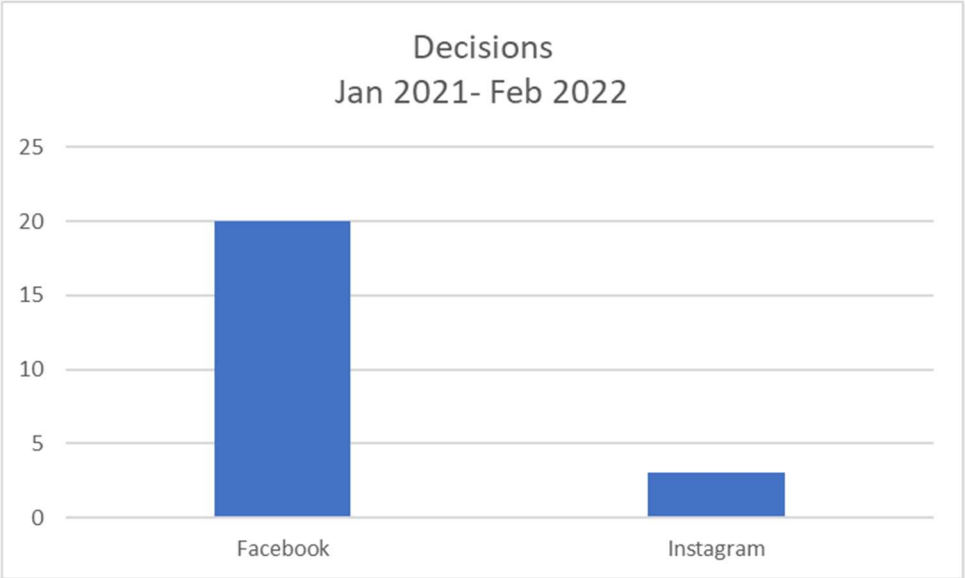

Figure 3: The vast majority of decisions (20) were directed at Facebook.

These figures show that the trends reported in the first transparency report seem to be continuing also in 2020. Hate speech seems to be the most common standard used in the decisions, the bias towards Europe and North America as regions where most cases come from continues and Facebook is treated much more than Instagram.

Facebook states that it takes seriously its role as a communication service for the global community. In a 2020 US Senate hearing, CEO Mark Zuckerberg noted that the company's products 'enabled more than 3 billion people around the world to share ideas, offer support, and discuss important issues' and reaffirmed a commitment to keeping users safe. However, Facebook Files (Rochefort et al. 2022) shows that Facebook allocates 87 per cent of its budget for combating misinformation to issues and users based in the United States, even though these users make up just about 10 per cent of the platform's daily active users. Finland is a good example as a small nation and language with fewer speakers. A Facebook data leak (Yle 2021) has revealed that Finnish-language moderation

rests with a handful of moderators: about ten people working in Berlin. Moderation for languages such as Finnish is often half-baked, a fact apparent in thousands of pages of leaked internal Facebook documents obtained by the Finnish Broadcasting Company (Yle 2021). These documents show that Facebook has not developed Finnish-language automated moderation for things such as hate speech, violence and nudity.

The principles of 'Lex Facebook'

The rules of Facebook – the 'Lex Facebook', as Bygrave (2015) has named them – are a hierarchically structured normative system. Their principles comprise Facebook Values at the top tier, Community Standards at the second tier and Internal Implementation Standards and AI Protocols as the interpretation of these abstract rules (Klonick 2018). Facebook Values and Community Standards are public documents. The standards and protocols are specific, non-public instructions for moderators and those developing the protocols of the algorithms. In several cases, moderators have leaked these internal rules and instructions but because these rules and instructions change all the time, it remains unknown how up-to-date these leaked versions are. In any case, Lex Facebook is only partially based on public documents and policies, which makes a complete assessment of the company's practices impossible.

The OB's core function is to review content enforcement decisions, to determine whether they are consistent with Facebook's content policies and values and to interpret decisions vis-à-vis Facebook's articulated values (OB charter, 1.4.2). Therefore, all decisions must be based on Lex Facebook. On the other hand, past decisions are not binding such as decisions by the US Supreme Court (art 2.2). In that sense, the OB is more like a European court than Anglo-American one. The charter demands that the OB balance free expression with other rights (art 3). The OB has the option to offer advice on how Lex Facebook should be developed, but the company has no obligation to follow this advice. Additionally, the OB's bylaws state that the OB's duty is to implement reviews and decisions in accordance with company policies and values (art 1.3): 'The board will have no authority or powers beyond those expressly defined by this charter' (art 1.4). However, the charter allows the OB to consider human rights norms that protect free expression. Still, these external norms complement the Lex Facebook, and the norms regulating the OB's process do not define which human rights norms the OB should follow.

According to the bylaws, the OB reviews cases when people have exhausted the other appeals processes of Facebook, when Facebook refers the case to the OB and when the OB decides that the case has wider significance in interpreting the rules of Facebook. After selection by the selection committee, the OB has 90 days to give a final verdict (bylaws article 1 3.1).

First, the case is reviewed by a panel of five members, of which four are randomly assigned and one represents the region of the case in hand. It should be noted that not all content of platforms is available for review by the OB. This content includes spam, messages, dating, marketplace and other services owned by Facebook (article 2 1.2.1). In addition, if content is removed based on illegality, the OB has no competence to review its removal. In theory, this allows both platforms the option to avoid the review by OB stating that the content is illegal without reviewing illegality in court or in other instances.

The Lex Facebook constitutes the core structure of norms upon which the OB bases its decisions. After its establishment, the OB announced that it also considered international human rights norms and standards (Gradoni 2021). The Rulebook for Case Review and Policy Guidance is a framework

for the OB, and it declares that it aligns with the United Nations' Guiding Principles on Business and Human Rights (UNGP). The UNGP guidelines call on companies and states to prevent and remedy human rights abuse in business. It comprises three pillars and 31 principles. The UNGP pillars are a state's duty to protect human rights, a corporate responsibility to respect human rights and individual access to a remedy if human rights are not respected or protected. However, the UNGP is soft law by nature, lacking an enforcement mechanism.

In its cases, the OB has referred not only to the UNGP but also to the International Covenant on Civil and Political Rights (ICCPR), which is part of the core structure of the United Nations' human rights system. It should be noted that the ICCPR and its interpretation are closer to European standards of weighing and balancing than US absolute freedom of speech doctrine. Therefore, the OB does not idiosyncratically follow only the Lex Facebook; it also considers freedom of speech and other human rights formulations in key human rights treaties. The structure is almost identical in each decision: first, the OB refers to the Community Standards and Facebook Values, and then, it interprets human rights. Lorenzo Gradoni (2021) even suggests that the OB has de facto neutralized these values and enthroned human rights. However, the OB has somewhat problematically made special reference to article 19 of the ICCPR to overturn removal decisions in most of its first decisions. Thus, the OB seemingly protects freedom of speech over other conflicting human rights, which aligns with Facebook's business model: more content means more users, more traffic and more money (Douek 2021; Montero Regules 2021).

One of the OB's most prominent cases so far has been the case of former US President Donald Trump. The CEO of Facebook, Mark Zuckerberg, announced on 7 January 2021 that the Trump ban would apply indefinitely for posts related to his supporters' attack on the US Capitol (Romm and Dwoskin 2021). The OB's decision to uphold Facebook's ban on Trump attracted worldwide attention. The OB found the platform to have wrongly banned Trump 'indefinitely', insisting that the company 'apply and justify a defined penalty' and allowing Facebook six months to review its initial decision in May 2021 (Tidman and O'Connell 2021). In June 2021, Facebook announced that it had defined the ban of Donald Trump as lasting two more years (Isaac and Frenkel 2021).

The OB's decision justified the ban, but for technical and procedural reasons, the OB returned the case to Facebook. One of the reasons was that the Lex Facebook did not allow a permanent ban; therefore, Facebook needed to decide how to continue. This case might be one of the watershed moments of the OB. In 2018, Kate Klonick and other critics assumed that the OB was just a scapegoat for controversial decisions (Klonick 2018). Opposing this notion, the OB dodged the bullet and returned the case to Facebook. Therefore, the Trump ban was a decision made by Facebook, not by the OB. However, in the Trump case, the OB asked Facebook 46 questions, but Facebook declined to answer two partially and seven entirely. These questions were related mostly to algorithms, and the rationale of the questions was to find out whether there were less severe measures than permanent ban.

Discussion: The OB's work assessed

Mark Zuckerberg has suggested that the OB should eventually become the 'Supreme Court of Facebook' (Klein 2018). Therefore, its decisions should establish precedents (Gradoni 2021), its decisions should be built on a robust hierarchy of international law and treaties, and it should be compatible with other organs of international law. Here, based on the data of the cases and our legal analysis and we have collected some arguments against the requirements mentioned above.

Too few cases

When assessing the OB's normative and legal bases, it is a step forward in the sense that Facebook is now admitting to having some responsibility for the content available on its services. The OB is, of course, an achievement as such, and its occasional willingness to challenge the Lex Facebook comes as a pleasant surprise. However, the OB handles only a small fraction of possible cases. By the end of February 2022, the OB had only made decisions in 23 cases. This small number of decisions indicates that the OB serves more as a legal consultant, making decisions in 'emblematic cases', rather than as a real court that would serve all of those who feel they and their content have been mistreated. However, this does not necessarily mean that it will not play an important role in guiding the company to adopt more ethical procedures and create discussions around freedom of speech online.

Idiosyncratic system

The OB was intended to be an external monitoring body for Facebook and Instagram. The OB relies on the Lex Facebook and human rights treaties that are relatively ambiguous, and in practice, no court refers to the same sources. Therefore, the OB and the Lex Facebook remain idiosyncratic systems, and their decisions are valid only within this system. Additionally, the OB has not yet to practise any real weighing and balancing between different rights in the manner practised by courts – especially the US Supreme Court or the ECtHR. Initially, the OB has been very free speech oriented, which well suits the business models of Facebook and other social media companies. The OB also lacks a contextual approach, which is essential for hard human rights cases (Montero Regules 2021).

As stated, the Lex Facebook is an idiosyncratic system of standards, some of which are not public. One of the criteria for appealing to the OB stipulates that the content should not be removed based on its illegality. The OB decides this solely based on the Lex Facebook and the OB's bylaws. In its first decisions, the OB referred to the ICCPR and UNGP – but what is the added value of these references? The OB does not interpret laws or human rights treaties; rather, it picks one guideline (UNGP) and one treaty (ICCPR) to give its decisions juridical camouflage. Thus far, its decisions represent a very liberal and US-based interpretation of freedom of speech, with a hint of the European style of proportionality (Pollicino et al. 2021). However, one of the fundamental problems is not that platforms remove content but rather that users depend on these platforms and that appealing the removal or restoration of content is almost impossible (Ghosh and Hendrix 2021).

Unclear justifications for human rights

The ICCPR and UNGP are global human rights treaties. Among their first related cases were decisions made in Europe and South America. However, both regions have their own treaties to protect human rights. The OB's decisions may be questioned from the perspective of such regional human rights doctrines. Additionally, since the adoption of the EU Charter of Fundamental Rights of the European Union, the ECJ has played a role in setting human rights standards in Europe, and the European Union and ECJ have been very active in cases related to the internet and platforms. The OB's decisions may be questioned from the perspective of such regional human rights doctrines. Especially in light of ECtHR case law, one could argue that an opposite outcome to similar cases would have occurred in human rights courts.

Not compatible with the DSA Act

The concept of light regulation has recently faced heavy criticism because of the clearly criminal or harmful content that has circulated on platforms. The codes and communications of the European Union have been insufficient, and the solution to this criticism is more binding regulation in the style of the DSA and DMA. In response to many concerns, the DSA proposal includes demands for transparency and an independent dispute settlement body. However, the new authorities outlined in the proposal would operate at the national or regional level, and the body's final form has yet to be determined.

One should also assess how the OB fulfils its need to improve the basic rights of users and citizens, especially in the context of the DSA proposal. Interestingly, for the OB's legitimacy, an advisory group – the European Board for Digital Services – will issue guidance and help ensure consistency for very large platforms such as Facebook. The question that remains is as follows: how does this planned European board relate to Facebook's OB?

The legal body described in the current proposal for the DSA would provide very formal answers to problems that are constantly changing. This perspective suggests that self-regulation bodies or SMCs could offer a softer solution to the global and borderless nature of the issues surrounding platforms. However, this kind of softer regulation requires a normative structure and, as we have argued, clear principles for interpretation. After all, the normative structures of the internet and human rights law are not a buffet table from which bodies like the OB can pick suitable norms for each of their purposes.

The OB is not a SMC

The latest proposals to promote human rights in social media content moderation are SMCs. The most discussed proposal is by the NGO, Article 19. In this proposal, the SMC would be a transparent and independent mechanism to address the problems of content moderation. This model is focused on making moderation accountable to international human rights standards, which means UN treaties. The proposal for SMCs promotes freedom of speech and urges that its members should represent all groups of society. UN Special Rapporteur on freedom of expression David Kaye first endorsed this proposal in 2018. In 2019, Article 19 changed its focus from global SMCs to national SMCs, and a pilot SMC is planned to be installed in Ireland.

Matthias C. Kettemann and Martin Fertmann (2021) argue that the OB follows the SMC proposal, but it is more of a council for the platforms of Facebook and Instagram. That assessment is fair, and compared with the idea of a more inclusive and effective SMC, the scope of OB is narrower, and so far, it has only made a handful of decisions. However, the creation of the OB is still much more than other social media companies have done, and at the moment, it is the only active 'council' in this field.

Conclusion

The early internet was influenced by the ideas and politics of freedom and openness, which also affected regulation principles. The regulation of social media platforms started developing according to the idea of platforms serving as intermediaries with no obligation to control the content that they mediate.

The vast number of complaints about illegal and harmful content – and the hundreds of thousand appeals to the OB – show that the moderation by Facebook and Instagram is not sufficient for the task. However, the platforms are owned by Meta, which operates at a global level. Regional or

national laws do not apply fully to Meta, and the global normative order is relatively soft compared with regional and national normative orders. Additionally, the DSA by the European Union is a legal body; hence, the tension between legal requirements for moderation and requirements in the platforms' terms of service remains. However, as an organ for only one company, the OB cannot be part of the current initiatives for SMCs. The OB is sui generis by nature.

As we have stated in the present article, the OB is a step forward because it steers and guides the moderation practices of Meta's platforms towards a slightly more transparent and human rights-friendly direction. At the same time, it insufficiently fulfils the regulatory needs expressed, for example, in the DSA. The OB can be assessed as more of a pseudo-regulative organ than a real solution to the lack of solid legislation for internet platforms. It remains to be seen how the OB will handle its biggest problem – the arbitrariness of moderation – depending on language, region and the user's social status. For example, Frances Haugen's revelations have shown that currently, Meta's platforms moderate its North American content at the highest intensity and, with somewhat lower intensity, the European content. Most of the world, including most of Facebook's and Instagram's users, especially in the Global South and regions of less common languages, are almost entirely neglected.

Finally, the OB maintains the focus of public discussion on content issues and diverts it away from the other problems caused by Meta's policies and business model. The discourse around illegal and harmful content may obscure other fundamental yet unrealized rights, such as equal access to platforms, equal availability of reliable information and rights to privacy and control over personal data for all citizens (Sirkkunen et al. 2021). These rights must also be addressed through international law.

Funding

Both authors are members of the research project Communications Rights in the Age of Digital Disruption (CORDI) funded by the Academy of Finland.

References

Bygrave, L. A. (2015), Internet Governance by Contract, Oxford: Oxford University Press.

Callamard, A. (2019), 'The human rights obligations of non-state actors', in R. Jörgensen (ed.), Human Rights in the Age of Platforms, Cambridge, MA: MIT Press, pp. 191–226.

Culliford, E. (2019), 'Facebook pledges $130 million to content Oversight Board, delays naming members', Reuters, 12 December, https://www.reuters.com/article/us-facebook-oversight-idUKKBN1YG1ZG. Accessed 20 April 2022.

DeNardis, L. and Hackl, A. M. (2015), 'Internet governance by social media platforms', Telecommunications Policy, 39:9, pp. 761–70.

Douek, E. (2021), 'The Facebook Oversight Board's first decisions: Ambitious, and perhaps impractical', Lawfare, 28 January, https://www.lawfareblog.com/facebook-oversight-boards-first-decisions-ambitious-and-perhaps-impractical. Accessed 5 March 2022.

Freedman, B. and Rorive, I. (2002), 'Regulating internet content through intermediaries in Europe and the USA', Zeitschrift für Rechtssoziologie, 23:1, pp. 41–60, https://doi.org/10.1515/zfrs-2002-0104. Accessed 5 June 2022.

Ghosh, D. and Hendrix, J. (2021), 'Facebook's Oversight Board just announced its first cases, but it already needs an overhaul', VerfBlog, 19 December, https://verfassungsblog.de/fob-first-cases/. Accessed 5 April 2021.

Gillespie, T. (2018), Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media, London: Yale University Press.

Gradoni, L. (2021), 'Constitutional review via Facebook's Oversight Board: How platform governance had its Marbury v Madison', VerfBlog, 10 February, https://verfassungsblog.de/fob-marbury-v-madison/. Accessed 5 April 2021.

Gregorio, G. De (2021), 'The rise of digital constitutionalism in the European Union', International Journal of Constitutional Law, 19:1, pp. 41–70.

Gregorio, G. De (2022), Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society, Cambridge: Cambridge University Press.

Holznagel, D. (2022), 'A self-regulatory race to the bottom through Art. 18 Digital Services Act: How the DSA will introduce competition for the Meta Oversight Board (and the German FSM) and why we should be worried about this', VerfBlog, 16 March, https://verfassungsblog.de/a-self-regulatory-race-to-the-bottom-through-art-18-digital-services-act/. Accessed 20 June 2022.

Horwitz, J. (2021), 'Facebook says its rules apply to all: Company documents reveal a secret elite that's exempt', The Wall Street Journal, 13 September, https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353. Accessed 20 September 2021.

Isaac, M. and Frenkel, S. (2021), 'Facebook says Trump's ban will last at least 2 years', New York Times, 4 June, https://www.nytimes.com/2021/06/04/technology/facebook-trump-ban.html?smid=em-share. Accessed 20 November 2021.

Jørgensen, R. and Pedersen, A. (2017), 'Online service producers as human rights arbiters', in L. Floridi and M. Taddeo (eds), The Responsibilities of Online Service Providers, Berlin: Springer, pp. 179–99.

Kaye, D. (2016), Report of the Special Rapporteur to the Human Rights Council on Freedom of Expression, States and the Private Sector in the Digital Age, A/HRC/32/38 (11 May 2016), New York: United Nations General Assembly.

Kettemann, M. (2020), The Normative Order of the Internet: A Theory of Rule and Regulation Online, Oxford: Oxford University Press.

Kettemann, M. C. and Fertmann, M. (2021), 'Making platforms rules more democratic: Are social media councils the way to go?', [online] 19 May 2021, doi: 10.5281/zenodo.4773130, https://doi.org/10.5281/zenodo.4773130.

Klein, E. (2018), 'Mark Zuckerberg on Facebook's hardest year, and what comes next', Vox, 2 April, https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-interview-fake-news-bots-cambridge. Accessed 20 October 2021.

Klonick, K. (2018), 'The new governors: The people, rules, and processes governing online speech', Harvard Law Review, 131:6, pp. 1598–670.

Klonick, K. (2020), 'The Facebook Oversight Board: Creating an independent institution to adjudicate online free expression', Yale Law Journal, 129:2418, pp. 2418–99.

Maroni, M. (2022), The Right to Access the Internet: A Critical Analysis of the Constitutionalisation of the Internet, Helsinki: Law Faculty, University of Helsinki.

Marsden, C. T. (2011), Internet Co-regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace, Cambridge: Cambridge University Press.

Montero Regules, J. (2021), 'The Facebook Oversight Board and "context": Analyzing the first decisions on hate speech', VerfBlog, 16 February, https://verfassungsblog.de/fob-context/. Accessed 7 April 2021.

Oversight Board (2021a), 'Oversight Board publishes transparency report for third quarter of 2021', https://www.oversightboard.com/news/640697330273796-oversight-board-publishes-transparency-report-for-third-quarter-of-2021/. Accessed 23 June 2022.

Oversight Board (2021b), 'Oversight Board to meet with Frances Haugen', https://www.oversightboard.com/news/1232363373906301-oversight-board-to-meet-with-frances-haugen/. Accessed 23 June 2022.

Oversight Board (2021c), 'Oversight Board transparency reports: Q4 2020, Q1 and Q2 2021', https://www.oversightboard.com/news/215139350722703-oversight-board-demands-more-transparency-from-facebook/. Accessed 23 June 2022.

Oversight Board (2022), 'The website', https://www.oversightboard.com/. Accessed 21 February 2022.

Paul, K. (2021), 'Facebook must tackle "Spanish-language disinformation crisis", lawmakers say', The Guardian, 16 March, https://www.theguardian.com/technology/2021/mar/16/facebook-spanish-language-disinformation-congress. Accessed 7 April 2021.

Poell, T., Nieborg, D. and van Dijck, J. (2019), 'Platformization', Internet Policy Review, [online] 8:4, https://doi.org/10.14763/2019.4.1425. Accessed 17 April 2021.

Pollicino, O. (2021), Judicial Protection of Fundamental Rights on the Internet: A Road Towards Digital Constitutionalism?, Oxford: Bloomsbury Publishing.

Pollicino, O., Gregorio, G. De and Bassini, M. (2021), 'Trump's indefinite ban: Shifting the Facebook Oversight Board away from the First Amendment doctrine', VerfBlog, 11 May, https://verfassungsblog.de/fob-trump-2/. Accessed 7 June 2021.

Puddephatt, A. (2021), 'Letting the sun shine in: Transparency and accountability in the Digital Age', Unesco, https://unesdoc.unesco.org/ark:/48223/pf0000377231. Accessed 26 June 2022.

Radin, M. J. (2004), 'Regulation by contract, regulation by machine', Journal of Institutional and Theoretical Economics, 160:1, pp. 142–56.

Radu, R. (2019), Negotiating Internet Governance, Oxford: Oxford University Press.

Roberts, S. T. (2019), Behind the Screen: Content Moderation in the Shadows of Social Media, New Haven, CT: Yale University Press.

Rochefort, A. and Rogoff, Z.; RDR Staff (2022), 'Cross-checking Facebook: Five lies revealed by Frances Haugen', Ranking Digital Rights, https://rankingdigitalrights.org/2021/10/14/cross-checking-facebook-frances-haugen/. Accessed 15 June 2022.

Romm, T. and Dwoskin, E. (2021), 'Trump banned from Facebook indefinitely, CEO Mark Zuckerberg says', The Washington Post, 7 January, https://www.washingtonpost.com/technology/2021/01/07/trump-twitter-ban/. Accessed 7 January 2021.

Sirkkunen, E., Horowitz, M., Nieminen, H. and Grigor, I. (2021), Media Platformisation and Finland, University of Helsinki and Tampere University, https://urn.fi/URN:ISBN:978-952-03-2110-9.

Statista (2021), 'Leading countries based on Facebook audience size as of October 2021', October, https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/.

Suzor, N. (2018), 'Digital constitutionalism: Using the rule of law to evaluate the legitimacy of governance by platforms', Social Media + Society, 4:3, pp. 1–11.

Tambini, D. and Marsden, C. T. (2007), Codifying Cyberspace: Communications Self-Regulation in the Age of Internet Convergence, London: Routledge.

The Citizens (2020), 'The Real Facebook Oversight Board', https://the-citizens.com/real-facebook-oversight/about-us/. Accessed 7 April 2021.

Tidman, Z. and O'Connell, O. (2021), 'Trump v Facebook: Ex-president rages at ban as White House says tech has responsibility to public', The Independent, 5 May, https://www.independent.co.uk/news/world/americas/us-politics/trump-facebook-latest-news-desk-b1842687.html. Accessed 10 May 2021.

van Dijck, J., Nieborg, D. and Poell, T. (2019), 'Reframing platform power', Internet Policy Review, [online] 8:2, https://doi.org/10.14763/2019.2.1414.

Wijeratne, W. (2020), 'Facebook, language, and the difficulty of moderating hate speech', LSE Blog, 23 July, https://blogs.lse.ac.uk/medialse/2020/07/23/facebook-language-and-the-difficulty-of-moderating-hate-speech/. Accessed 8 April 2021.

Yle (2021), 'Facebook failing at Finnish moderation', https://yle.fi/news/3-12238371. Accessed 17 June 2022.

Zittrain, J. L. (2006), 'A history of online gatekeeping', Harvard Journal of Law and Technology, 19:253, pp. 253–98.

Zittrain, J. L. (2008), The Future of the Internet: And How to Stop It, London: Yale University Press.

Zuboff, S. (2015), 'Big other: Surveillance capitalism and the prospects of an information civilization', Journal of Information Technology, 30:1, pp. 75–89.

SUGGESTED CITATION

Neuvonen, Riku and Sirkkunen, Esa (2022), 'Outsourced justice: The case of the Facebook Oversight Board', Journal of Digital Media & Policy, X:X, pp. 00–00, https://doi.org/10.1386/jdmp10.1386/jdmp_00108_1.

Notes

1. The non-profit organization The Citizens was formed soon after the foundation of the OB. The Citizens founded The Real Facebook Oversight Board as a shadow board to respond to the critical threats posed by Facebook's unchecked power during the 2020 US presidential elections. The shadow board's members are also prominent figures – for example, researcher and author Shoshana Zuboff, author Timothy Snyder and former President of Estonia Toomas Hendrik Ilves (The Citizens 2020).

2. For some reason, the OB has not published the transparency reports after the third quarter of 2021 (Oversight Board 2021a) by the end of June 2022. The OB published first Annual Report 22 June 2022 which summarize information of quarter reports.