UNIVERSIDADE
CATÓLICA
PORTUGUESA

Master's degree in law & business

Navigating Cyberspace: Unraveling the Future of Freedom of Expression in the Age of Digital Platforms

Beatriz Sofia dos Santos Martins de Almeida

143722001

Coordination: Prof. João Confraria & Prof. Giovanni De Gregorio

05/09/2024

“The most important kind of freedom is to be what you really are. You trade in your reality for a role. You trade in your sense for an act. You give up your ability to feel, and in exchange, put on a mask. There can't be any large-scale revolution until there's a personal revolution, on an individual level. It's got to happen inside first.”

Jim Morrison

Aos meus Avós

Ao meu Namorado

Ao meu pai

… e à minha Mãe, que sei que nunca me deixou.

ACKNOWLEDGMENTS

O presente trabalho é fruto de um processo que exigiu alguns sacrifícios, muita força de vontade essencialmente perseverança. A sua concretização só foi possível graças à ajuda de algumas pessoas a quem desejo apresentar os meus mais sinceros agradecimentos.

To my supervisors, Professor Doctor João Confraria and Professor Giovani De Gregorio, for all the support and patience they had with me. I thank them for their good disposition, understanding, advice and total availability at all stages of the work and, above all, for the trust and interest they placed in my topic.

À minha família e Amigos, pela força que sempre me transmitiram para elaborar o projecto e pela total compreensão da minha ausência, em muitos dos eventos de confraternização que tiveram lugar durante este período.

Aos meus avós, pelo apoio e amor incondicional que me deram em todas as fases da minha vida, que se revelou crucial para a elaboração desta dissertação.

Agradeço-lhes a sua compreensão e a habitual motivação com que pude contar ao longo desta jornada.

Ao meu Namorado, André, pela força transmitida, pela paciência e por todo o amor demonstrado em todo o caminho que já percorremos juntos, face a todas as adversidades. Agradecer-te não é um gesto que se coloca em papel, mas algo que se partilha ao longo da vida. Obrigada por tudo.

À minha querida mãe, Anabela, pelo amor imensurável que me deu em vida e pela luz que sempre me guiou após a sua partida. Onde quer que esteja, desejo que saiba que tudo o que alcancei na minha vida e tudo o que sou, lhe devo a si.

RESUMO

A Internet veio revolucionar o mundo das comunicações de modo inovador. O inventor do telefone, rádio, computador e do telegrafo prepararam o mundo para esta nova integração de possibilidades como nunca foi visto outrora[1]. Esta tese de mestrado tem como foco o cenário da liberdade de expressão atual, que é dominado pelas plataformas digitais.

À medida que a sociedade passa por uma mudança sem precedentes no que respeita às interações online e à disseminação de informação, os paradigmas tradicionais de liberdade de expressão enfrentam novos desafios e oportunidades.

Esta pesquisa baseia-se em perspetivas jurídicas, sociológicas e tecnológicas para examinar a natureza evolutiva da liberdade de expressão na esfera digital.

O estudo começa com uma abordagem ao contexto histórico e fundamentos teóricos da liberdade de expressão, na qual destacaremos o seu papel central nas sociedades democráticas. De seguida, examinaremos a influência que as plataformas digitais exerceram e continuam a exercer nos padrões de comunicação bem como o impacto nas diversas vozes, no discurso político e no envolvimento público. Este trabalho pretende avaliar ainda o papel das grandes plataformas digitais enquanto "guardiãs da informação" e os desafios que têm vindo a colocar à visão tradicional da liberdade de expressão.

Além disso, investigaremos o panorama jurídico existente para a liberdade de expressão na era digital e de que forma este se adequa para enfrentar novos desafios, tais como o discurso de ódio online e a desinformação. Em complemento, o estudo examina o papel da governança da plataforma, da moderação de conteúdo e seu impacto no mercado e das considerações éticas associadas a estas práticas.

Um aspeto crucial desta tese é o de fornecer soluções e recomendações políticas para proteger e melhorar a liberdade de expressão na era digital. Tal inclui avaliar a eficácia da

autorregulação, da cooperação internacional e das novas tecnologias na promoção de uma esfera pública digital inclusiva.

Ao fornecer uma análise abrangente do futuro da liberdade de expressão na era das plataformas digitais, este trabalho contribui para os esforços levados a cabo para a existência de uma simbiose entre tecnologia, direito e sociedade.

Por fim, as conclusões visam alertar os decisores políticos, os juristas e os profissionais da tecnologia para a importância de promover uma compreensão diferenciada dos desafios e oportunidades associados à defesa deste direito fundamental.

Palavras-chave: Internet, Liberdade de Expressão, Plataformas Digitais, Comunicações.

ABSTRACT

"The Internet has revolutionized the communications world like nothing before. The invention of the telegraph, telephone, radio, and computer set the stage for this unprecedent integration of capabilities[1]".

This master's thesis focuses on the current scenario of freedom of expression, dominated by digital platforms.

As society undergoes unprecedented changes in terms of online interactions and the dissemination of information, traditional paradigms of freedom of expression face new challenges and opportunities.

This research draws on legal, sociological, and technological perspectives to examine the evolution of freedom of expression in the digital domain.

The study begins with an approach to the historical context and theoretical foundations of freedom of expression, in which we will highlight its central role in democratic societies. We then explore the influence that digital platforms have had and continue to have on communication patterns as well as the impact on diverse voices, political discourse, and public engagement. This work also aims to evaluate the role of large digital platforms as "guardians of information" and the challenges they pose to the traditional view of freedom of expression.

In addition, we will examine the existing legal framework for freedom of expression in the digital age and how it is fit to address new challenges, such as online hate speech and online disinformation. Additionally, the study examines the role of platform governance, content moderation and its implications on the market and ethical considerations associated with these practices.

A key aspect of this research is the analysis of possible policy solutions and recommendations to protect and improve freedom of expression in the digital age. This

includes assessing the effectiveness of self-regulation, international cooperation, and new technologies in promoting an inclusive digital public sphere.

Through a comprehensive analysis of the future of freedom of expression in the age of digital platforms, this work contributes to efforts to create a symbiosis between technology, law and society.

Finally, the conclusions aim to inform policymakers, legal experts and technology experts to promote a nuanced understanding of the challenges and opportunities associated with defending this fundamental right in our increasingly interconnected world.

Keywords: Internet, Freedom of Expression, Digital Platforms, Communications.

ABREVIATIONS

- **DSA** – Digital Services Act
- **UDHR** - Universal Declaration of Human Rights
- **ECHR** - European Convention on Human Rights
- **ICCPR** - International Covenant on Civil and Political Rights
- **API** – Application Programming Interface
- **ISP** – Internet Service Providers
- **DARPA** - Defense Advanced Research Projects Agency
- **UN** – United Nations
- **UGC** – User-generated Content
- **AI** – Artificial Intelligence
- **CWC** – Country-withheld-content
- **DMA** – Digital Market Act
- **VLOPs** – Very Large Online Platforms
- **VLOSEs** – Very Large Online Search Engines
- **GDPR** – General Data Protection Regulation
- **ToS** – Terms of Service

INDEX

1. INTRODUCTION

1.1. The Origins of Freedom of Expression

"Human rights are legally codified norms applying to all human beings, irrespective of national borders. International human rights law obligates states to act in certain ways or refrain from certain acts to protect the human rights of individuals[2]".

The expression of opinion embodies rights, liberty, and freedom. The freedom to hold opinions, think critically and to have access information is the essence of expression. In a democratic framework, freedom of speech and expression allows individuals to freely articulate, disseminate, and publish their perspectives and ideas on any topic. This freedom is essential to nourish and enrich individual identity[3]. This privilege is of immense importance in promoting and maintaining a free and fair society, where governance is based on the consent and understanding of well-informed citizens, committed to protecting the rights of all individuals, including the most marginalized minorities. [3].

The right to freedom of expression is protected by most international human rights instruments, such as: Article 19[4] of the Universal Declaration of Human Rights (UDHR), Article 10[5] of the European Convention on Human Rights (ECHR) and Article 19(2)[6] of the International Covenant on Civil and Political Rights (ICCPR).

Freedom of expression has been influenced by the Internet both negatively and positively:

⇒ On the one hand, it opened the possibility for people to communicate and express their thoughts explicitly. Even if the topic in question is unpopular or subject of general discussion, anyone can express their opinion without fear of being ignored or supressed. This facilitated, therefore, the dissemination of knowledge and concepts that benefited society.

⇒ On the other hand, the Internet has also brought new challenges and risks, as it has made it possible for governments and other stakeholders to censor online

expressions of opinion according with their beliefs. In addition, hate speech, disinformation and harmful content have become the most difficult issues for freedom of speech to address. Minorities and marginalized groups suffer the most from these negative impacts, as they are the most vulnerable to online harassment. This subsequently creates a climate of fear and restricts the diversity of opinions in the online presentation[7].

The right to freedom of expression is not only a fundamental individual right subject to potential interference by public authorities but also a constitutional instrument that promotes autonomy in a democratic society and embodies the principles of dignity that characterise European Constitutionalism[8].

1.2.Origins of the Internet

In 1966, Lawrence G. Roberts went to DARPA[1] to develop the concept of a computer network, which eventually paved the way for the creation of the "ARPANET".

The Internet has revolutionized interpersonal and social communication, exponentially expanded access to information and knowledge, and created a public sphere in which everyone can express ideas, opinions, and disseminate facts[9]. Before the internet, participation in public debates relied on the professional press, which investigated facts, respected journalistic ethics and was liable for damages if it published, knowingly or recklessly, false information. However, this does not imply that world was perfect before the internet[2].

The Internet functions as a global broadcasting platform, a tool for the dissemination of information, and a means of collaboration and interaction between individuals and their

[1] Defense Advanced Research Projects Agency.

[2] The number of media outlets was, and continues to be, limited in quantity and perspectives; journalistic companies have their own interests, and not all of them distinguish fact from opinion with the necessary care.

computers, regardless of their geographical location. This is an example of the significant benefits of continued investment and commitment to research and development in the field of information infrastructure. Since the first research in the field of packet switching, government, industry, and academia have been working together to develop and implement this innovative technology[1].

The Internet is considered multi-jurisdictional in nature, because it allows individuals from different locations[3] to transact with each other[10]. Internet was designed and developed with the objective of facilitating communication. In the mid 1990s, government control over the internet was considered not desirable neither possible.

[3] And, consequently, with different jurisdictions.

2. DIGITAL PLATFORMS AND THE ISSUE OF MODERATION

2.1. The role of digital platforms in contemporary communication

Gillespie[11] refers to platforms as 'online sites and services that host, organize, and circulate users' shared content or social interactions for them, without having produced the bulk of that content themselves, built on an infrastructure for processing data for a range of different purposes including the generation of profit, and which moderate the content and activity of users'.

Digital platforms have fundamentally changed contemporary communication and reshaped the way individuals, organizations and societies interact. Here[12] are some key roles that digital platforms play in contemporary communication:

- Facilitating Instant Communication:
    - Real-Time Interaction: Platforms such as *WhatsApp*, *Messenger*, and *Slack* allow instant messaging, allowing for real-time conversations regardless of geographical boundaries.
    - Video Conferencing: Tools such as *Zoom*, *Teams*, and *Google Meet* have become essential for virtual meetings to shorten distances and enabling face-to-face interaction.

- Expanding Reach and Connectivity:

    - Global Connectivity: Social media platforms such as *Facebook*, *X*, and *Instagram* connect people around the globe and foster a sense of global community.
    - Networking Opportunities: *LinkedIn* and other professional networks allow individuals to build professional relationships, share industry insights, and find job opportunities.

- Enhancing the dissemination of Information:

  - News and Updates: Digital news platforms and social media allow news to spread quickly and provide real-time updates on global events.
  - User-Generated Content: Platforms such as *YouTube*, *TikTok*, and blogs allow individuals to create and share content, thereby influencing public opinion and trends.

- Facilitating Collaboration and Knowledge Sharing:

  - Collaborative Tools: Platforms like *Google Drive*, *Trello*, and *Asana* make collaboration easy and allow teams to work together, regardless of location.
  - Online Communities: Forums and platforms such as *Reddit*, *Stack Overflow*, and *Quora* provide spaces for knowledge sharing and problem-solving within specific interest groups.

- Transforming Marketing and Advertising:

  - Targeted Advertising: Social media and search engines offer highly targeted advertising options, that allow businesses to reach specific demographics with personalized messages.
  - Influencer Marketing: Platforms allow influencers to build a large following and collaborate with brands, creating a new marketing dynamic.

- Strengthening Civic Engagement and Social Movements:
  - Activism and Advocacy: Digital platforms are crucial for organizing social movements, raising awareness, and mobilizing support for various causes (e.g., *#BlackLivesMatter, #MeToo*).
  - Political involvement: They facilitate political campaigns, voter registration campaigns, and public discourse on political issues.

- Entertainment and Leisure:

  - Streaming Services: Platforms such *Netflix*, *Spotify*, and *Twitch* offer a variety of entertainment options, including movies, music, and live game streams.
  - Social Games: Games with built-in social features (e.g., Fortnite, Among Us) combine entertainment with social interaction.

- Support for Education and E-Learning:

  - Online Learning: Platforms such as *Coursera*, *Udemy*, and *Khan Academy* provide access to educational resources and courses from top institutions.
  - Virtual Classrooms: Tools such as *Moodle* and *Blackboard* support online teaching and enable interactive learning experiences.

- Personal Branding and Self-Expression:

  - Digital Portfolios: Platforms like *Behance* and *Dribbble* allow creatives to showcase their work and build their personal brand.
  - Blogs and Vlogs: Personal blogs and vlogs provide opportunities for self-expression and the exchange of personal experiences.

- Financial Services:

  - Buying and selling financial assets: Platforms like *Robinhood*, *E*TRADE*, and *TD Ameritrade* have made it easy for individuals to buy and sell stocks, bonds, and other financial assets from their computers or mobile devices.
    - Lower barriers to entry: these platforms often have lower fees and minimum investment requirements, making the financial markets more accessible to a wider audience.

2.1.1. How does social media affect freedom of expression?

'Over the past few years, social media have contributed positively to individuals' ability to enjoy a broad range of human rights beyond freedom of expression, having a transformative impact on individuals' ability to gather, mobilize, learn, educate, and so on around the globe. However, there is a growing awareness that the digital space also has a negative impact on human rights and can encourage new cases of violence, hatred and discrimination[2].

Social media has a profound impact on freedom of expression, both positive and negative.

Some positive effects include[13][14]:

a) Amplification of voices:
    a. Accessibility: Social media platforms allow anyone with access to the internet to share their views and give a voice to those who may not have access to traditional media.
    b. Global Reach: Ideas and opinions can instantly reach a global audience, allowing cross-culture dialogue and the dissemination of different perspectives.

b) Mobilization and Advocacy:
    a. Social Movements: Social media has been instrumental in organizing and mobilizing social movements (e.g., #BlackLivesMatter, #MeToo), and provides a platform for advocacy and awareness on critical issues.
    b. Activism: Activists use social media to garner support, coordinate protests, and influence public policy, thereby improving democratic participation.
    c. Citizen Journalism:

i. Real-Time Reporting: Individuals can report on events as they happen, providing real-time updates and alternative narratives to traditional media.
ii. Diverse Sources: Citizen journalism increases the diversity of information sources and contributes to a better-informed public.

Negative effects[15]:

a) Censorship and Content Moderation:
   a. Platform Policies: Social media companies have policies in place to remove content that is deemed harmful or inappropriate, which can sometimes lead to excessive censorship and the suppression of legitimate expression.

      1. Excessive censorship refers to the overly strict control or suppression of speech, communication, and other forms of information by governments, private institutions, or individuals. It goes beyond reasonable regulations and often infringes upon freedom of expression and thought[4].

      2. Content moderation by competing platforms does not entirely solve the problems of harmful content for several reasons: different platforms often have varying definitions and policies regarding what constitutes harmful content, leading to gaps where harmful content removed from one platform remains accessible on another; when

[4] Historical examples:

- Throughout history, excessive censorship has been seen in various forms, such as book burning, banning of films and music, and controlling internet access. For instance, regimes like Nazi Germany and the Soviet Union heavily censored all forms of media to suppress dissent and control public opinion.
- In more recent times, countries like China and North Korea enforce strict internet censorship, limiting access to information and restricting online speech to maintain governmental control.

harmful content is moderated and removed from one platform, it often migrates to less regulated or alternative platforms, known as "platform migration"; effective content moderation requires substantial resources and sophisticated algorithms, which smaller or newer platforms may lack, leading to uneven enforcement and the persistence of harmful content; there is an ongoing debate about the balance between content moderation and freedom of expression, with over-moderation leading to accusations of censorship and suppression of legitimate speech, causing platforms to adopt more lenient moderation policies to attract users, thereby allowing more harmful content to proliferate; platforms must navigate diverse cultural norms and legal frameworks globally, complicating the creation and enforcement of universal moderation standards; and the fragmentation of content across competing platforms can lead to the creation of echo chambers, where users are exposed only to information that reinforces their existing beliefs, increasing polarization and the spread of harmful content among like-minded communities without effective moderation interventions.

b. Government Pressure: In some countries, governments are pressuring social media platforms to censor content critical of the regime, thereby undermining freedom of expression. This situation is, in many ways, like what used to happen (and still does) with traditional media. However, there are key differences between both situations[16]: for instance, media platforms operate on a global scale and enable real-time dissemination of information, making censorship efforts more urgent and pervasive compared to traditional media, which had more localized and slower distribution. Secondly, social media is largely driven by user-generated content, making it harder to control and predict what will be shared. In contrast, traditional

media had centralized editorial control, which made it easier for governments to exert pressure on specific entities and control the narrative. Thirdly, social media allows users to post anonymously or under pseudonyms, complicating efforts to track and censor content. Traditional media, with known journalists and clear editorial structures, was easier to monitor and regulate. Lastly, social media platforms use algorithms to prioritize and filter content, which can be manipulated to suppress critical content. Traditional media relied on human editors, making the process of content control more transparent and straightforward. Content on social media can go viral through sharing, liking, and commenting, leading to rapid spread and potentially larger impact. Traditional media lacked this level of interactivity and virality, limiting the immediate reach and influence of censored information. Governments now often employ sophisticated tactics, such as leveraging fake news laws, using cyber armies to flood platforms with propaganda, and pressuring companies to comply with local regulations under the threat of being banned or fined. These methods are more complex and technologically advanced compared to the direct suppression tactics used on traditional media[5].

b) Misinformation and Disinformation:
    a. Spread of misinformation: social media can be a conduit for misinformation and disinformation, which can distort public perception and hinder informed decision-making.

[5]In countries like China and Russia, governments have successfully pressured social media platforms to censor content critical of the regime. For instance, China's Great Firewall blocks access to major social media sites like *Facebook* and X, while Russia has passed laws requiring social media companies to store data locally and remove content deemed illegal. These differences highlight the evolving nature of censorship in the digital age, where the decentralized, rapid, and interactive nature of social media poses new challenges for both censors and defenders of free expression[97], [98], [99].

b. Echo Chambers: Algorithms often reinforce users' pre-existing beliefs by promoting content aligned with their views, limiting exposure to different perspectives and encouraging echo chambers.

c) Online Harassment and abuse:
   a. Trolling and Bullying: anonymity on social media can encourage harassment, bullying and trolling, which can silence individuals and discourage them from speaking up.
   b. Chilling Effect: fear of online abuse can lead to self-censorship, where individuals do not share their opinions to avoid negative reactions.

d) Surveillance and Privacy Concerns:
   a. Data Collection: Social media platforms collect huge amounts of data about users that can be used for surveillance by governments and businesses.
   b. Chilling Effect: Knowledge of being watched can discourage people from expressing controversial or dissenting opinions.

The impact of social media on free speech is a complex balance[17]:

- Content Moderation vs. Freedom of expression: Platforms must moderate harmful content while protecting freedom of expression. This balance is often difficult to achieve, leading to debates about the appropriate scope and methods of content moderation[18].
- Regulation and Oversight: There are calls for greater regulation and oversight of social media platforms to protect users' rights without stifling innovation and freedom of expression. Effective regulation must address issues such as disinformation, harassment and censorship, while respecting the principles of freedom of expression.

To what extent can and should digital platforms be linked to freedom of expression, when they are private companies?

One possible approach to this question is the doctrine of indirect horizontal effect in constitutional law[19]. This doctrine examines whether private actors in the digital space can be bound by fundamental rights[6]. According to this principle, fundamental rights "radiate" into private legal relationships through specific legal provisions. The indirect horizontal effect suggests that individuals have different degrees of obligation to respect fundamental rights, which do not necessarily have to be as stringent as the obligations of the State. The intensity of this commitment may vary; it is not a binary state where you are completely connected or not at all. The magnitude of this horizontal effect can be more pronounced in contexts such as content moderation, depending on several factors. This includes the question of whether private actors provide a space for general communication and social interaction, their dominant position, their orientation and the degree of dependence of users on the platform[20].

2.1.2. How can content be moderated without compromising freedom of expression?

Firstly, a distinction is made between “content moderation” and “content regulation”[21]. Regulating content consists of removing illegal content and ensuring that the limits of freedom of expression are respected. Content moderation is legal and illegal content, as defined by companies in their terms of service and community standards. This means that content can still be moderated if it violates the platform's community guidelines, which often include restrictions on hate speech, harassment, and misinformation. These guidelines are created by the companies to foster a safe and respectful environment for their users.

[6] A recent case of the Federal Constitutional Court states that under certain conditions private parties can increasingly be bound by freedom of expression in the same way as the state.

However, the distinction between what is considered legal or illegal content varies by jurisdiction, and the application of these rules can be subjective and inconsistent.

Human rights norms provide legal standards for the former and limited legal guidelines for the latter.

When it comes to content moderation, the ethical dilemma stems from the painstaking task of balancing the protection of freedom of expression and mitigating potential threats to the public sphere[18]. This means that when we are talking about content moderation policies, it is important to have in mind that they are set according to the objectives of the platforms and, as private businesses, they are profit maximizers. These policies are designed to foster a safe and engaging environment for users, which is crucial for retaining and attracting users, advertisers, and partners. Its objectives include maintaining user trust, ensuring compliance with local and international laws, and protecting the platform's reputation. However, there are limited legal guidelines for content moderation, giving companies significant discretion in setting and enforcing their policies which allows companies to tailor their content moderation strategies to their business goals, user base, and market conditions.

Platforms like *Facebook*, *X*, and *YouTube* develop community guidelines that address issues like hate speech, harassment, misinformation, and graphic violence, which often reflect the platform's values and the expectations of their user communities[7]. As profit-driven entities, social media platforms balance the enforcement of these guidelines with the need to maximize user engagement and revenue. This can lead to conflicts of interest, where the desire to keep users active and engaged might conflict with the need to remove harmful or controversial content. Moreover, the algorithms used to enforce content moderation can sometimes inadvertently suppress legitimate expression or fail to catch

[7] For example, *Facebook's* Community Standards prohibit content that promotes violence, hate speech, and nudity, while X's policies focus on preventing abuse, harassment, and the spread of false information.

harmful content, leading to criticism and calls for greater transparency and accountability in the moderation process.

A particular content regulation policy may affect the ability of billions of users to express their opinions and obtain information. It is therefore unclear whether such a policy would constitute a violation of human rights [2].

With intense competition between online platforms, websites need to attract and retain users to stay relevant. While promoting interactivity is essential to increasing reach and engagement, the overall user experience determines the success of online brands.

Content moderation plays a crucial role in improving the user experience by managing user-generated content (UGC) in three ways[18]:

⇒ Reduce the Spread of Harmful Content: By blocking or removing toxic content, content moderation helps mitigate its negative impact in the real world.

⇒ Creating a Safe and Inclusive Digital Environment: Moderation efforts detect and remove harmful content before it reaches users, promoting a safe and welcoming online space.

⇒ Reflect on the Brand's Values and Ethos: Effective content moderation reduces toxic content and demonstrates the brand's commitment to user well-being, improving its image and giving it a competitive advantage.

Freedom of expression is a cornerstone of democratic societies, and the digital age has expanded this freedom, allowing individuals to express their opinions and beliefs like never before. However, this also brought significant challenges and complexities in content moderation[18]:

⇒ Challenges in Decision-Making:

- Ambiguous Content: While some topics, such as illicit drug use, are clearly recognizable as right or wrong, others are inherently ambiguous. This is especially true for humorous content, which can vary greatly from culture to culture. A joke that is harmless in one culture might be offensive in another, making moderation more complex because it requires a local understanding of the context.
- User-Generated Content (UGC): UGC often includes hate speech, triggered by the anonymity of digital media. This anonymity can entice users or inadvertently lead them to engage in hate speech or discrimination. The challenge is to reconcile freedom of expression with the need to ensure public safety.

⇒ Detection of Misinformation

- Detecting disinformation: Another major challenge is the spread of disinformation. Disinformation can take different forms, and detecting false content is not always easy. While some content has clearly been fabricated or presented by scammers, other forms, such as misleading, satirical, and propagandistic content, fall into gray areas and can slip through moderation efforts because they somehow appear to be true. Therefore, online platforms need to improve their content moderation capabilities to prevent the spread of inaccurate information.

⇒ Take Appropriate Action

- Mitigation: it is ideal to ensure that online information is secure and factual, but mitigating the harms caused by misinformation is a particular challenge for platforms that rely on user engagement. Measures such as removing harmful content, removing websites, or punishing accounts that threaten online safety can sometimes violate freedom of expression. Online platforms such as dating sites, e-businesses, web communities, virtual marketplaces

and online gaming are faced with the pressing task of carefully assessing content to mantain a balance between social responsibility and the defense of freedom of expression.

Here are some real-world examples and case studies[14]:

- X on Donald Trump's ban in January 2021[22];
- Facebook and Cambridge Analytica (2018)[23];
- YouTube and misinformation about Covid-19 [24];
- Chinas TikTok censorship (2020)[25].

2.1.3. Moderation guidelines – who decides and how are they implemented?

Human rights have traditionally regulated the power dynamics between the state and the individual and protect the freedom of the individual against arbitrary interference by state authorities. This protection includes the right to express opinions, which allows individuals to oppose the system. However, if we move from the negative obligation of government to refrain from interfering to a positive duty to protect, it can be argued that States increasingly need to ensure that freedom of expression is defended in cyberspace. Therefore, the Internet should be afforded the same level of protection as the physical public space[26].

As mentioned above, states are responsible not only for human rights violations they themselves commit, but also for those committed by others if they fail to prevent, punish, or remedy such violations. Regarding freedom of expression, state action has traditionally been a decisive factor in alleged human rights violations. If a state commissions a private platform to remove content, this constitutes a violation of this right, unless the order is legally justified and aims to achieve a legitimate aim. On the other hand, if a platform removes content for violating its terms of use, it constitutes a private act and does not fall within the direct scope of human rights law. From a legal point of view, the relationship

between the platform and the user is not regulated by human rights, but by the terms of use (contract law).[8]

2.1.4. Hate speech – is the platform responsible for it?

Currently, there is no uniform definition of what is considered hate speech around the world, so, for this work, we use the definition from the EU Code of Conduct[27]: “all conduct publicly inciting to violence or hatred directed against a group of persons or a member of such group defined by reference to race, color, religion, descent, or national or ethnic origin”[28].

The EU Code of Conduct on countering illegal hate speech online is an agreement between the European Commission and *Facebook*, *Microsoft*, *X* and *YouTube*, that provides for the establishment of internal procedures to ensure that companies investigate reports of illegal removal of hate speech within 24 hours and, if necessary, remove or disable access to such content. This includes working with civil society organizations ("trusted reporters") to identify content that incites violence and hate speech[2].

Rudeness has long been a feature of public discourse, and continues to be so today, despite efforts to overcome it. "Incivility" is defined as elements of a discussion that show unnecessary disrespect for the forum, its participants, or its topics. Although such incivility often arises from opposing ideas, justified or not, it often leads to rude reactions that do not add any significant value to the conversation[29].

A significant modern challenge was highlighted that hampers efforts to create a space for free and respectful exchange of ideas[29]. Initially, the Internet was heralded as a global forum for the presentation, debate, and organization of information[30]. However, this broad interactive media environment, which has broadened the possibilities for public debate, has also facilitated the rapid and widespread spread of incivility[29].

[8] While Facebook’s practices may affect individuals’ ability to exercise freedom of expression, it does not have a legal obligation to protect this right.

Burch[30] notes that before the Internet, hate groups relied on written material to spread their propaganda and connect with like-minded people. The advent of the Internet has simplified this task due to its unique features, such as accuracy and speed, which have provided both access to information and the ability to reach a wider audience. This has resulted in a variety of contents that are as broad as human thought itself. Once these individuals or groups were able to connect effortlessly, their hateful beliefs were mutually reinforced, allowing them to promote hate speech and behaviors cheaply and easily[9].

Burch[30] also highlights the delicate balance between sharing ideas and perpetuating hatred, which can lead to humiliation and violence. Freedom of expression online can easily turn into language that incites violence and poses a real threat. There are two reasons for this conclusion: younger Internet users are more susceptible to hateful ideas and value and hate speech can lead to deadly crimes and destruction.

As an example, we can cite the tragic events of September 11, 2001, in which extremist values led to the deadliest terrorist attacks in United States history[31], suggesting that regulation is necessary to prevent such events. Despite this, the U.S. Supreme Court has consistently upheld the concept of the marketplace of ideas, where both good and bad ideas are presented, with the truth ultimately prevailing[30].

One contributing factor to the preference for a non-interference approach was the original vision of the Internet as a space for democratizing access to information, knowledge and culture, where governments would not have the same sovereignty as in the physical world[32]. Early discussions on regulation of cyberspace emphasized the Internet's reliance on voluntary, cross-border cooperation, and advocated the creation of decentralized laws as an alternative to traditional hierarchical state control. The term "governance" became popular because it suggested a more polycentric order than "government"[33]. For example, Elon Musk expressed a desire to create a dual-version platform for *X*, one with

[9] This can be proven by the fact that various Taliban officials and affiliated entities maintain verified accounts on the platform X[100], [101].

and one without regulations, to manage different content control settings. This ambitious plan is part of his broader vision to transform *X*, into an all-encompassing "everything app" that includes services such as payments and banking. However, implementing this vision faces significant regulatory and logistical challenges[34].

Mueller and Badiei[33] point out that between 2003 and 2009, the Internet was recognized as “the dominance of global governance”. This is meant to reflect how the Internet began to play a crucial role in shaping global political and economic landscapes during that period. Its rise transformed the dissemination of information, communication methods, and the debate and enforcement of international policies. During this period, there was increased recognition of the Internet's central role in national and international policymaking, with issues such as cybersecurity, intellectual property, and online freedom coming to the forefront. The expansion of the Internet fostered greater global connectivity, influencing economic and political dynamics and integrating online platforms and services into governance and international relations. This dominance also led to a shift in power dynamics, as traditional forms of governance were increasingly challenged by new, decentralized forms of organization and activism enabled by the Internet[35]. Essentially, Mueller and Badiei[33] highlight how the Internet's rapid growth and integration into various aspects of life and governance made it a dominant force in shaping global interactions and policies in the early 21st century.

This phenomenon coincided with the emergence of social media platforms such as *Facebook*, *X*, and *Google*. These great global intermediaries have revolutionized discussions about communication policy by providing instantaneous, real-time connections and easy access to information and entertainment.

The unique context of digital communication, where there is a lack of nonverbal cues and physical proximity, has made it prone to rude interactions compared to traditional face-to-face situations[36]. Limited social signals in online communication often led to disinhibition and antisocial behavior[29], which led to hostile exchanges. Anderson et

al[36] also found that online rudeness can polarize opinions and increase the perception of bias in media-related matters.

Burch[30] argues that the global interconnection of jurisdictions through the Internet has made national laws less effective in combating hate crimes in cyberspace. He suggests that to develop new approaches to online hate speech and its unique challenges, it is essential to first understand the existing frameworks to deal with such issues.
While online platforms have a responsibility rather than an obligation to ensure respect for fundamental rights and freedoms, democratic states have a mandate to safeguard those interests to protect the entire democratic system. This duty includes a positive obligation to protect individuals from the acts committed by private persons or entities[37]. Without the preservation of equality, freedom of expression and freedom of assembly, a democratic society would not be possible[8].

2.1.4.1. How to avoid excessive censorship and, at the same time, combat misinformation?

Here are some highlights[38]:

- Practical responses: Empirical research shows several effective strategies to counter disinformation that do not violate freedom of expression or other fundamental rights. Specific fact-checking, content labeling and media literacy techniques have been documented to reduce the impact and spread of disinformation. In addition, countries with diverse and robust independent media tend to be more resilient to disinformation. These strategies do not restrict freedom of expression; instead, they provide the necessary context. They do not restrict independent media; they support and strengthen them. Governments can combat disinformation by promoting these measures through transparency reports that highlight adherence to best practices and by funding independent fact-checking, media literacy, and media initiatives. Similarly, platform companies can take and support these actions and share data

with independent researchers to assess and improve the effectiveness of different responses.

- Legal responses: It is important to apply the same human rights standards to online behavior as to offline behavior, and to ensure that all legal restrictions on online expression are consistent, transparent, and due process is maintained. To protect freedom of expression while addressing disinformation, states should ensure that legal restrictions on expression are clearly defined by law, necessary to protect other fundamental values and proportionate to the specific threat. These conditions are enshrined in the European Convention on Human Rights and have been reaffirmed by interpretations by the European Court of Human Rights, the European Commission of Human Rights, and the Council of Europe. The strict rules of interpretation and the obligation for states to prove that all conditions are fulfilled provide additional safeguards.
- Platform Responses: It is crucial that platforms align their policies and processes with international human rights principles and point out potential conflicts with local laws. As noted by the UN Special Rapporteur, "human rights standards, if implemented transparently and consistently with meaningful input from users and civil society, provide a framework for holding both states and companies accountable to users across national borders." States are obligated under international human rights law to respect, protect, and fulfil human rights locally, and companies must not only obey the law, also respect human rights. The United Nations Guiding Principles on Business and Human Rights provide a clear basis for addressing conflicts between local laws and human rights principles. They state: "the responsibility to respect human rights is a global standard of expected conduct for all business enterprises wherever they operate. It exists independently of states' abilities and/or willingness to fulfill their own human rights obligations[39]."
- Oversight, Transparency, and Due Process: it is crucial to improve transparency in the management of content moderation by platforms, regarding disinformation. This need has been underlined in the Santa Clara Principles on Transparency and

Accountability in Content Moderation[40] and through proposals for independent social media boards as multilateral accountability mechanisms. Transparency should also extend to the use of artificial intelligence (AI) in content moderation. While AI can be useful, it is often unable to accurately identify copyrighted content and pornography, and it is even less reliable when it comes to making political judgments about misinformation. Therefore, AI should be used with caution and under human supervision and independent control. Platforms should also document official orders and requests and disclose their responses. Greater transparency about country-withheld-content (CWC) programs is critical because they are typically used to comply with local laws and are difficult to analyze externally. While platforms often report the number of CWC requests and their compliance rates, more detailed information is needed, especially when governments prefer non-disclosure. This transparency is crucial, as comprehensive legislation on disinformation can give rise to numerous questionable CWC requests. It is important to avoid the transition from opaque and unilateral content moderation by private companies to similar practices by governments. Freedom of expression is under threat around the globe, including by political actors and governments, even in established democracies. Addressing disinformation without compromising freedom of expression and other fundamental rights requires practical interventions and legal responses focused on consistency, transparency, and accountability. While it may be necessary to stop the most egregious examples of disinformation spread by officials and media, such measures should be based on solid evidence and taken with caution so as not to harm freedom of expression. Research can help assess the extent of disinformation and find evidence-based responses that are guided by international human rights law. This approach can encourage governments and platforms to take appropriate action without compromising human rights. Improving oversight, transparency and due process on content moderation will help develop effective solutions to protect against disinformation, while respecting fundamental rights.

3. CHALLENGES

3.1. The impact of the Digital Services Act on freedom of expression

Regulation (EU) No. 2022/2065[41] of the European Parliament and of the Council of 19 October 2022 on a single market for digital services, commonly known as the Digital Services Regulation ("Digital Services Act" or "DSA ") is another important milestone in the European Union's activity in regulating the digital sector[42].

Introduced as part of the Digital Services Act Package launched by the European Commission in December 2020, this Regulation aims to establish common rules to effectively protect consumers using digital services. It sets new obligations for all players in the intermediation of services, ranging from small intermediary providers to major Big Tech companies operating within the European space.

'The DSA regulates online intermediaries and platforms such as marketplaces, social networks, content-sharing platforms, app stores, and online travel and accommodation platforms'[42]. Its main goal is to prevent illegal and harmful activities online and the spread of disinformation. It ensures user safety, protects fundamental rights, and creates a fair and open online platform environment[42]. 'Together with the Digital Markets Act, the DSA plays a critical role in providing a supranational and horizontal regime to mitigate the challenges raised by the power of online platforms in content moderation'[8].

The impact of the Digital Services Acts (DSAs) on freedom of expression is significant and complex. These legislative measures aim to regulate online platforms and ensure that they take responsibility for the content they store, while also protecting fundamental rights such as the right to freedom of expression.

Here's how the DSA can affect freedom of expression[43]:

- Easier reporting of illegal content (article 16 DSA): The DSA obliges online platforms to take action to combat the spread of illegal goods, services or

content[43]. These measures include providing tools to report such content and working with 'trusted flaggers' (article 22 DSA). Attentive users may have already noticed changes in the online environment, such as X's recent addition of a feature to report illegal content. This feature can be accessed by clicking on the three small dots located in the top right corner of each post. *X* is not alone in this effort; other platforms like *Apple*, *Pinterest*, *Facebook*, *Instagram*, and *TikTok* have also introduced new easy-to-use options for reporting illegal content[43].

- More transparency in content moderation and more ways to interact with content (articles 20-21 DSA): Online platforms act as digital spaces where we express ourselves, share our work and stay connected with friends or customers. Therefore, it is frustrating when our content is removed or its reach inexplicably restricted. According to the DSA, intermediaries and hosting service providers, including online platforms, must now explain to users why content has been removed or access to the account has been restricted. This legal requirement ensures clear and specific explanations for content moderation decisions and allows users to challenge them through an out-of-court dispute resolution mechanism. Even platforms like *Facebook* and *Instagram*, which once offered some transparency about content moderation decisions, are expanding the information they offer to users. To promote transparency and allow for a review of content moderation practices under the DSA, the European Commission has launched the DSA Transparency Database. This unique database provides public access to all statements made by online platform providers about their content moderation decisions. Very large online platforms (VLOPs) must start publishing their explanations by the end of August 2023, with a deadline of February 17, 2024, for all other platform providers covered by the DSA[43].

- More knowledge and choice about what we see – and more control over personalization options: DSA requires online platform providers to provide more

transparency and user control over the content displayed in our feeds. This is to help users understand how platforms prioritize content and decide whether to reject personalized recommendations, as VLOPs must provide an option to opt out of personalized content. Similarly, platforms need to improve transparency and control over why certain ads appear in our feeds and label ads accordingly, with VLOPs maintaining a repository that lists paid advertising campaigns in their interfaces. Currently, *TikTok*, *Facebook*, and *Instagram* allow users to disable personalized feeds on their platforms. Member States and the European Commission will work closely together to monitor and enforce the provisions on the transparency of recommendation systems and advertising, including the creation of ad registers[43].

- Zero tolerance for targeting ads to children and young people and targeting ads based on sensitive data: The DSA requires online platforms not to use advertising targeted directed to minors. Very large online platforms (VLOPs) have responded by taking action to comply with these regulations. For example, *Snapchat*, *Google* (including *YouTube*), *Instagram*, and *Facebook* have stopped allowing advertisers to target ads specifically to underage users. Additionally, *TikTok* and *YouTube* have adjusted their settings so that user accounts under the age of 16 are automatically set to private by default. Moreover, the DSA prohibits targeted advertising on online platforms where profiling is created using sensitive categories of personal data, such as ethnicity, political opinions and sexual orientation[43].

    - Example: China has taken strict measures to regulate children's access to content on digital platforms. These regulations aim to limit exposure to potentially harmful or inappropriate material and promote content that aligns with "socialist values" and contributes positively to the development of minors. For example, the "Minor mode" feature restricts children's daily screen time based on age, with younger children (under 8) limited to 40

minutes per day and those under 18 prohibited from using their devices between 10 p.m. and 6 a.m. This mode also includes features to encourage breaks every 30 minutes and parental controls to encourage device usage times. The content allowed in this modality varies according to the age group and focuses on educational and morally positive materials. For children under the age of three, age-appropriate music and audio material are part of the content. As children get older, the content they can access expands to include general educational material, popular science, life skills, and age-appropriate news. For teenagers, the content is focused on fun but positively guiding material that matches their cognitive abilities. These regulations are part of China's broader efforts to combat internet addiction among youth people and integrate political ideologies into everyday online content, reflecting a strong focus on national security and social stability. Some parents have welcomed the increased scrutiny, while critics argue it could limit access to information and impose excessive government oversight[44], [45], [46].

- Protection of children: the new rules state that online platforms that remain accessible to children must protect their safety and privacy, as well as their mental and physical well-being. These measures can be achieved through special privacy and security settings. Platforms may also require age verification measures to control who is viewing content as well as parent controls. For example, *TikTok* and *YouTube* have banned advertising targeted to minors and automatically set minors' profiles to private, meaning that the videos they upload can only be seen by people who have previously shared them[43].

- Electoral integrity: Under the DSA, Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) are needed to assess and mitigate risks related to electoral processes and civic society discourse, while preserving

freedom of expression. The Slovak parliamentary elections held on September 30, 2023[10], marked the first test case for these obligations after the DSA was applied to the VLOP and VLOSE. As a result of the DSA, providers of VLOPs and VLOSEs have adapted their approach to protect the integrity of elections. They reduced response times to content reported by local authorities and trusted partners, introduced clearer procedures to deal with disinformation, improved fact-checking capabilities, and strengthened global resources and capacity[43].

- New obligations on traceability of business users in online marketplaces: The DSA introduces new requirements for online marketplace providers to tackle the distribution of illegal goods. Specifically, these suppliers must now ensure that sellers verify their identity before offering products for sale. They are also required to provide easy identification of the seller responsible for each transaction. If a provider of an online marketplace provider becomes aware that an illegal product or service is being sold, it must inform the users concerned of the nature of the infringement, disclose the seller's identity, and indicate the remedies available. The Commission, in cooperation with Member States, will monitor and enforce these obligations to prevent the proliferation of illegal goods on online marketplaces. [43]. The DSA states that intermediary service providers should not be considered liable for illegal third-party content when they do not initiate or have actual knowledge of the illegal activity or information and are unaware of the facts or circumstances concerning the illegal activity or information (articles 4-6 DSA)[47].

Given that the DSA imposes severe penalties for non-compliance, including fines of up to 6% of the intermediary service provider's annual global revenue and potential temporary bans on operating in the European Union, it is crucial for companies to adhere to the new obligations. Meanwhile, we await the enforcement actions of the newly designated regulatory entities overseeing this Regulation[48].

[10] https://www.osce.org/odihr/elections/slovakia/548812

Therefore, organizations should consider the following actions[48]:

- Determine if their services fall under the DSA;
- Identify the DSA's legal requirements relevant to the nature of their services;
- Assess the applicability of exemptions available for micro and small businesses;
- Implement the necessary measures based on the scope of their services, such as revising and updating terms and conditions, privacy policies, functionalities for reporting illegal content, and other relevant legal documents and technical measures.

Further changes are taking place and the monitoring of the implementation of the DSA measures has begun. The goal of DSA remains unchanged: to create a safer and more inclusive digital world for all.

3.2. Impact of this new regulatory environment on the Market

As stated above, the DSA is a legislative framework enacted by the European Union aimed at creating a safer digital space. For that matter, it introduces a set of regulations and obligations for online platforms to ensure transparency, accountability and protection of user rights.

However, the DSA's approach has significant implications for future policymaking.

"There is a risk that imposing accountability at the threat of fines might increase the power of already dominant intermediaries"[49]. This is more noticeable for content moderation, because DSA's framework threatens to further strengthen the role of Big Tech in determining what is acceptable online speech[49].

First, we should debate whether the DSA can alter the power dynamics underlying innovation and competition in the maker of intermediary services. The DSA aims to

overcome legal fragmentation with a set of harmonized rules and provide "the conditions for innovative digital services to emerge and scale up in the internal market (recital 4)[41]".

Since the adoption of the E-Commerce Directive (ECD), legal fragmentation has been a problem, since it has left some important details to self-regulation and national law. This fragmentation has been increasing because of legislative developments in national law[11]. A more harmonized framework would benefit dominant companies, that have better resources to face a stricter EU regulatory environment, but there is a risk that smaller service providers might not be able to keep up. However, according to the European Comission's Impact assessment[12] accompanying the DSA proposal, said that overcoming fragmentation could also be beneficial for smaller companies, since they would be able to scale up their offerings in a more robust EU market[49].

Regarding costs of compliance, the EC acknowledges that the DSA brings additional costs for hosting service providers, but these costs are expected to be lower than the costs of dealing with a fragmented legal environment. Also, the EC believes that those costs will not be prohibitive for SMEs neither will affect them disproportionately in comparison to VLOPs[49]. However, the estimates for these costs are not comprehensive, since there are missing factors such as out-of-court disputes settlement and the moderation of harmful (but not illegal) content[49].

3.3.Limits of the intervention of private companies that own the platform.

Private companies that own and manage online platforms play a crucial role in moderating content and maintaining the integrity of their services. However, its intervention options are not unlimited. These limits are shaped by legal, ethical and operational considerations to ensure that the balance between freedom of expression and the need to manage harmful content is maintained.

[11] Example: NetzDG.

[12] https://digital-strategy.ec.europa.eu/en/library/impact-assessment-digital-services-act

Platforms must comply with national and international laws, such as the General Data Protection Regulation (GDPR)[50] in the EU, the Digital Services Act (DSA)[42], and other local regulations governing content moderation, privacy and user rights. The terms of Service (ToS) and Community Guidelines[13] define the scope of acceptable behaviour and content. These agreements are legally binding and provide a framework within which platforms can operate. In many countries, especially those with strong free speech protections such as the United States, platforms must navigate the fine line between moderating harmful content and violating users' free speech.

Platforms are increasingly expected to provide clear and specific reasons for content removal or account restrictions. This transparency helps build trust and allows users to understand the reasons behind moderation decisions. Platforms should have mechanisms in place to allow users to appeal moderation decisions. This helps to ensure that moderation is fair and can correct any errors or biases. Platforms should weigh the potential harm caused by certain types of content against the value of freedom of expression. This balance is often contextual and requires nuanced judgment[51].

With billions of users and massive amounts of content generated every minute, platforms face significant challenges when it comes to effectively moderating content at scale. Automated systems and human enablers must work together to accomplish this task. Content that is acceptable in one culture may be offensive in another. Platforms must carefully manage these differences to avoid unintended consequences. The nature of harmful content is constantly changing, so platforms must constantly adapt their moderation strategies [52].

Major platforms like *Facebook*, *X*, and *YouTube* have developed extensive content moderation guidelines that determine what is allowed and what is not. These strategies are frequently updated to respond to new challenges and societal changes. Some platforms

[13] For example: Facebook's community standards, X's rules and policies and YouTube's community guidelines.

have set up independent oversight boards or councils to review and advise on difficult moderation decisions. This can help ensure that moderation practices are not dictated solely by the platform's own interests. Court rulings can also set limits on how platforms can intervene.

By complying with these restrictions, private companies that own platforms can help promote a safer and more respectful online environment, while respecting users' rights to free expression[53].

3.4.Government involvement in content moderation?

Governments are increasingly taking a direct role in moderating online content, a task that used to be left mainly to private companies. While it is appropriate to involve governments, if they are based on clear and precise legislation, with independent oversight, transparency and due process, there is also a risk that such involvement could restrict freedom of expression[14].

Some governments may enact laws that define disinformation as containing criticism of the government or content that contradicts official narratives.
For example, the controversial Pakistan's law[54] lacks a clear definition of fake news, but only states that content that has been labelled as false by the Pakistani regulator should be treated as such. Similarly, Vietnam's Law on Cybersecurity[55] prohibits disinformation, including "distorting history," "denying revolutionary achievements," and "causing confusion among citizens," which can stifle free expression.

Disinformation laws that are too broad or vague pose risks to human rights[38]. They can disrupt legitimate speech and can be used selectively by governments to pressure private companies to monitor content in ways that undermine freedom of expression and limit public debate. Governments can use these laws, or existing laws with vague prohibitions,

[14] Example: Elon Musk's recent tweet on how allegedly European Commission has offered X a deal to censor speech "without telling anyone": https://x.com/elonmusk/status/1811783320839008381?s=48

such as blasphemy or incitement, to force platforms to geo-block or remove content altogether.

While platforms generally follow local laws, they do not always approve all requests for geo-blocking or removal, and sometimes resist government demands. However, their decision-making processes remain opaque and difficult to predict. Despite databases like *Lumen*, transparency in content moderation, especially in responding to government requests, is limited[38].

Governments have several means to pressure platforms to restrict the content under their jurisdictions. In April 2020, *Facebook* agreed to significantly increase compliance with the Vietnamese government's requests to censor anti-state content after the government took *Facebook's* local servers offline for seven weeks, making the platform unusable in Vietnam[56]. Similarly, *YouTube* and *TikTok* complied with Turkey's amended internet law, appointing local representatives and making themselves more susceptible to content removal requests from Turkish authorities[57].

These examples illustrate the influence that governments have on platforms, including imposing fines, blocking access and holding local representatives accountable. To directly influence platforms' content moderation, governments often must request geo-blocking and provide information about objectionable content and relevant laws. This process is time-consuming and is often carried out by civil officials or law enforcement authorities, resulting to inconsistencies and limitations in content restriction.

Inspired by the German Network Enforcement Act of 2017[58], states are trying to close enforcement gaps with new laws. This law obliges platforms with over two million users in Germany to deal with reports of illegal content, making every social media user a legal actor. Platforms must deal with obviously illegal content within 24 hours or investigate and decide within seven days, otherwise they face fines, speeding up content moderation. While the impact of the Network Enforcement Act in Germany may be less severe than feared,

there is still a risk that countries with poor human rights records will align their disinformation laws with it, improving their enforcement capacities and potentially restricting freedom of expression.

These developments show that governments are increasingly encouraging platforms to take on regulatory and policing functions traditionally reserved for public law. This delegation of tasks, unless clearly required by law, is necessary to protect certain values, is proportionate to the issues, is transparent with due process and can confer significant discretion on platforms. This poses risks to individual rights and freedom of expression, including the independence of media.[38]

3.4.1. How can platforms deal with political demands to delete certain content?

Platforms can respond to policy demands to remove certain content by implementing a structured approach that balances legal compliance with the protection of freedom of expression:

⇒ Establish Clear Policies and Procedures

- Transparent Guidelines: Develop and publish detailed content moderation guidelines that specify what content is allowed and what is not. This transparency helps users understand the rules and limits arbitrary enforcement. A good example is *Facebook's* Community Standards[59] and *X* rules[60].
- Consistent enforcement: Ensure that these policies are enforced consistently across all content and users, regardless of political pressures[15].

⇒ Regulatory Compliance[50]

- Compliance with Local Laws: Platforms must comply with the legal requirements of the countries in which they operate, including laws

[15] Donald Trump's example[22].

regarding illegal content. This can include removing content deemed illegal by local authorities.

- Due Diligence: Conduct a thorough legal review of claims to ensure that they are legitimate and comply with international human rights standards.

⇒ Engage in Dialogue

- Government Relations: Maintain open communication with government agencies to discuss the reasons behind content removal requests and seek clarification if necessary.
- Advocacy for Clear Laws: Advocate for clear, concise and fair laws regarding online content to avoid overly broad or vague regulations that may be abusive.

⇒ Transparency and Accountability

- Transparency Reports: Regularly publish transparency reports detailing the number and type of content takedown requests received from governments, the actions taken and the reasons for those actions.
- User Notifications: Inform users when their content is removed due to regulatory requests, providing clear reasons and ways for appeal.

⇒ Implement Robust Review Processes

- Internal Review: Establish a robust internal review process to evaluate the legitimacy and necessity of content removal requests[40]
- Independent Oversight: Consider establishing or participating in independent oversight bodies or boards that can review and make recommendations on contentious content decisions.[61]

⇒ Appeal Mechanisms

- User Appeals: Provide users with accessible and efficient mechanisms to appeal content removal decisions. This ensures that users can contest decisions that they believe are unjust.
- Judicial Recourse: In cases of significant controversy, ensure users can seek judicial review of content removal actions.

⇒ Use of International Standards

- Human Rights Frameworks: Align content moderation practices with international human rights standards, particularly those related to freedom of expression. Refer to guidelines from bodies like the UN or the European Court of Human Rights.
- Global Principles: Adhere to global principles such as the Manila Principles on Intermediary Liability, which provide a framework for protecting free expression while combating illegal content[62].
- Question: what happens if international standards are not consistent with national laws? In content moderation, inconsistencies between international standards and national laws can lead to a range of issues. Platforms might face conflicting guidelines, as international standards might advocate for broad freedom of expression, while national laws impose stricter regulations on certain types of content, such as hate speech or misinformation. This can create confusion and operational challenges for platforms trying to align their policies with both requirements. Platforms may also be exposed to legal liabilities if their content moderation policies do not comply with national laws, potentially facing fines, sanctions, or other legal actions. Inconsistencies can also lead to uneven enforcement of content policies, which affects users differently depending on their location, which can impact their experience and engagement. National laws may require the removal or restriction of certain content that may be allowed under international standards, affecting content availability and access and

potentially leading to disputes. Inconsistent moderation practices can affect user trust, as users may perceive content moderation as biased or unfair. This might cause platforms to face uncertainty about how to balance international standards with national regulations, especially when laws are ambiguous or rapidly evolving. Disputes over content moderation practices can also lead to diplomatic tensions and impact international relations. Because of that, platforms may adopt a hybrid approach, combining international standards with national legal requirements to create content moderation policies that respect local laws while upholding broader global principles[63] [11].

⇒ Technology and Automation

- Smart Moderation Tools: Use advanced technology and automation for initial content reviews, but ensure that final decisions, especially in complex or controversial cases, are made by humans.
- Audit and Improve: Regularly audit automated tools to prevent biases and errors, and continually improve their accuracy and fairness.

⇒ Engage with Civil Society

- Collaborate with NGOs: Work with non-governmental organizations, academic institutions, and civil society groups to understand the broader impact of content removal and to develop best practices.
- User Education: Educate users about their rights and the platform's content policies to foster a more informed and engaged community.

3.5. Self-regulation – ally or enemy of diversity of voice in the online public sphere?

Self-regulation is a combination of standards and codes of appropriate practices, which are necessary to support freedom of expression and guide the monitoring, careful analysis and accountability of media outlets.

The benefits of self-regulation are well known it preserves the independence of the media and protects it against political interference from governments. It can be more efficient than a system of government regulation, since the media understands its environment better than the government does – even though companies in the sector use this knowledge much more to serve their own commercial interests than the public interest .

When the environment becomes global (with the internet and other digital platforms), and jurisdiction issues become more complex, self-regulation also becomes more appropriate in several ways. The government spends less, because it is the industry itself that assumes the adaptation costs, following rules that are much more flexible than those established in state regulation.

The pressure exerted by companies on their peers is another positive factor, as it can act as an incentive for good practices, with respect to standards and rules – although there is evidence that external regulation, or the threat of applying it, is more effective to ensure compliance with standards and rules. Self-regulation can also encourage the improvement of professional standards, as its implementation requires organizations to suggest or even develop elements for their codes of conduct[64].

The communications environment has been profoundly changed by the possibility of transforming sound, image and text into digital code, accessible by various devices, from computers to cell phones. The emergence of the internet meant that the ability to communicate ceased to be something essentially local (whether in a location or country) to become a truly global phenomenon. In its first incarnation, the internet and the web were classified as a new global space, without borders, capable of avoiding traditional censorship[64].

What characteristics of this space have an impact on the right to freedom of expression? As a network of networks, the internet is an international platform without comprehensive administration. No entity governs the entire internet: governance is operated by institutions

that operate in very different jurisdictions. A program can be made in Ukraine, uploaded to a North American server, and downloaded in Ghana[64].

Bodies of international jurisdiction as well as bodies National governments that administer local domains are more concerned with the efficiency of the system and its functionality than with environmental governance, in the sense of the regulation that already exists for the press and radio broadcasting. Consequently, there is a jurisdictional vacuum over the content displayed on the internet. If there is need for interference by the State, it is not clear how such power should be used appropriately, as there is no way to regulate content internationally and there is also no consensus on which standards should be used. For this reason, much emphasis has been placed on the importance of self-regulation by online content providers[64].

But this approach poses risks. There are no self-regulation standards that have been developed for the internet and that are widely accepted. Thus, self-regulation – mainly by companies – takes place in a vacuum where economic interests and political pressure from governments stand out. For example, internet service providers (ISPs), which were originally expected to be mere conduits for the services they offer, are being asked to collect data about their users (as per the European Union Data Retention Directive 2006/24/ ECO) and even monitor browsing histories. These measures are taken through optional agreements with governments, without due analysis regarding their legality. The lack of a clear legal guideline and the understandable caution of providers regarding controversial content leads to a stance of excessive zeal, which results in situations such as the removal of certain content because of a simple complaint. This could be classified as a censorship regime, in contrast to the strict interpretations of the law and careful application of standards in the case of offline media. This framework reinforces the importance of the self-regulation system always being constituted with clear and transparent rules[64].

Self-regulation in the context of online platforms can be seen from both perspectives—as a potential support for the diversity of voices or as potential restriction of that diversity. On

the one hand, self-regulation allows platforms to set their own standards and policies, which can consider a wide range of opinions and viewpoints without direct government intervention. This flexibility could encourage a variety of voices to participate in online discourse, thereby promoting diversity.

On the other hand, self-regulations are also a cause for concern. Platforms may adopt content moderation policies that inadvertently or intentionally suppress certain voices or viewpoints, especially if they prioritize commercial interests or adhere to social norms that may stifle dissent. This could potentially lead to a homogenization of discourse and limit the representation of marginalized or controversial perspectives.

Whether self-regulation acts as an ally or enemy of the diversity of voices in the online public sphere, therefore, depends largely on how effectively platforms balance their responsibility to moderate harmful content with their commitment to upholding freedom of expression and promoting a variety of viewpoints.

3.6.Algorithms

‘Digital media platforms are advertising spaces that match users with news, services and applications, and they deliver personalized content. To be good at what they do, social media platforms need to keep users engaged with their services. This is what characterizes an attention-economy business model: the more a user engages with the platform, the more the platform can adjust the form of addiction to its users. In other words, attention is a scarce resource that platforms need to exploit to the maximum’[47].

Algorithms are pervasive in our lives, from recommending online search results and suggesting new friends on social media, to more significant applications like assessing cancer risks, determining mortgage eligibility, or predicting crimes such as gang violence and burglary. Their presence is so subtle that you might not even realize how deeply they influence our daily experiences. The content we see on *Facebook*, the music we hear on

Spotify, and the movies we watch on *Netflix* all depend on algorithms' predictive modeling. With each interaction, these algorithms refine and personalize our ‘feed’ and marketing to enhance our experience. While many of these applications are driven by commercial interests, algorithms also play a crucial role in public sectors like healthcare, education, criminal justice, and tax administration[65].

An "algorithmic society" refers to a framework where algorithms play a central and pervasive role in shaping various aspects of daily life and decision-making processes[65]. Algorithms influence a wide range of activities, from social media interactions and online shopping to healthcare diagnostics and criminal justice, driving many automated systems and services used daily. In such a society, algorithms assist or make decisions affecting individuals' lives, such as determining credit scores, recommending job candidates, diagnosing medical conditions, or predicting criminal behavior, with significant personal and societal implications. Data is a crucial resource in an algorithmic society, as algorithms rely on vast amounts of data to function and improve over time, making data collection, analysis, and management central to various sectors. Many algorithms are designed to personalize experiences based on individual preferences and behaviors, such as customizing content feeds on social media or personalizing advertisements, and they automate processes that were previously manual, increasing efficiency but sometimes leading to loss of human oversight. Algorithms can shape behavior and preferences by influencing what information, products, or services are presented to individuals, impacting consumer choices, political opinions, and social interactions. The pervasive use of algorithms raises important ethical and privacy issues, including concerns about surveillance, bias, and transparency, as well as how personal data is used and protected. The widespread influence of algorithms necessitates new approaches to governance and regulation, as societies must address how to manage algorithmic decisions, ensure fairness, and protect individuals' rights while fostering innovation[66].

In an algorithmic society, online platforms convey the users' ability to share their opinions and ideas. But this advantage has consequences. At first glance, the digital environment seems to offer a solution to the centralized powers in the media sector[16]. However, a closer look reveals that a similar dynamic of centralization and control over information has been replicated in the digital sphere, creating a quasi-public sphere.[67] The ability of platforms to massively organize or amplify certain voices (and their decisions on how to do so) gives rise to reflection on the future of the online public sphere [8].

Essentially, the promised democratization of information is compromised as power is concentrated in the hands of a few tech companies, replicating the centralization found in traditional media:

⇒ Social media platforms are key examples, as they control vast amounts of user-generated content and have significant influence over what content gets seen through their algorithms[17].

⇒ Search engines, like *Google*. Its algorithms largely determine wat information is accessible to users, as it has a dominant market share in search. The search engine's ranking system can make or break the visibility of websites, which represents a form of control over media consumption.

⇒ Streaming services dominate the distribution of video and music content. Their recommendation algorithm dictates what content is suggested to users, often prioritizing popular or promoted content.

⇒ E-commerce: Amazon's dominance in the e-commerce space is the best example. Its algorithms heavily influence which products are visible to customers and which sellers succeed. This concentrates market power in a single platform.

[16] Comparison to traditional media.

[17] They decide what trends, which posts go viral, and what content is promoted or demoted, centralizing control over information dissemination.

For example, regarding the distribution of content, *Facebook's* algorithm made it so that different reactions to publications had different weights. The "angry" reaction was weighted five times more than the simple "likes". This formula gave advantage to problematic content, as emotionally charged reactions occurred more often in relation to toxic content. Eventually, users started being exposed to less problematic content when *Facebook* set the weight on the angry reaction to zero[47].

3.6.1. Information bubbles – effects on democracy and public debate

The main problem with algorithms in terms of free speech is that they tend to create isolated communities that form isolated subgroups within the public. [68]. Democracy works best when there is a public sphere in which the meeting of ideas and opinions can be a "societal glue"[8]. If this does not occur, individuals may be drawn to extreme and dogmatic viewpoints and neglect the alternative ideas that form the basis of consensus in a democratic society[8].

Humans tend to create relationships with people who share their ideas and values: 'homophily of networks'[18][69][8].

How can individuals be rational users in the algorithmic public sphere when they are subjected to top-down control by online platforms that shape the public sphere through artificial intelligence systems with often inexplicable decision-making processes? In other words, the failure to protect freedom of expression as a negative right[19] to democratic values extends to the liberal vision of the digital public sphere[8]. The liberal vision of the digital public sphere is an ideal where the internet serves as a democratic space for open and free exchange of ideas, where everyone can participate equally and access diverse

[18] This paradox illustrates why social media platforms pledge to protect free speech while simultaneously moderating content to regulate their communities for business purposes. Consequently, a key issue is the compatibility between their private interests and public values[102].

[19] Freedom of expression as a negative right means the right to speak and express oneself without interference, especially from the government. In the context of the algorithmic public sphere, this right is also threatened by the control exerted by private platforms.

viewpoints. However, when platforms exert control through AI systems with non-transparent decision-making, this vision is compromised. Users cannot be fully rational and informed if they are subjected to manipulated information flows and lack understanding of how their information is curated.

The reality we face is that access to more information does not necessarily mean access to better information. Algorithms organize content to engage users based on their preferences and data, leading to the polarization of debate through the creation of 'filter bubbles', which Sunstein defines as 'communication universes in. which we hear only what we choose and only what comforts us and pleases us'[8]. This inevitably leads to the emergence of echo chambers, in which each user is isolated and marginalized from opposing viewpoints due to algorithmic calculations. In this scenario, users do not have access to transparent information about the processes running behind their screens[8].

3.6.2. Transparency in moderation decisions and how can platforms be more transparent?

Moderation can be defined as 'the screening, evaluation, categorization, approval or removal/hiding of online content according to relevant communications and publishing policies[8]. It seeks to support and enforce positive communications behaviour online, and to minimize aggression and anti-social behaviour'[70][8].

The lack of transparency and accountability in online content moderation hampers the exercise of freedoms in the public sphere and leads to a reassessment of the role of freedom of expression as a negative freedom in an algorithmic society. Platforms control the flow of information, defining, enforcing and balancing the right to freedom of expression based on their commercial interests**.**

Content moderation decisions can be fully automated, human-made or a combination of both. Pre-moderation activities such as prioritization, delisting and geo-blocking are

typically automated, while post-moderation often involves a mix of automated and human resources[8].

Moderation can occur before publication (pre-moderation) or after publication (post-moderation). Post-moderation involves organizing content and serves both as a reactive measure to evaluate reported content and as a proactive tool to actively monitor published content. According to Gillespie[11], 'moderation is not an ancillary aspect of what platforms do. It is essential, constitutional, definitional. Not only can platforms not survive without moderation, but they are also not platforms without it'[71]. This system is crucial because as it filters content to protect positive expression, fostering a digital environment where everyone feels safe to share their opinions. A widespread dissemination of harmful content, such as terrorism, can lead to a mass exodus of users, severely harming advertising revenues and negatively affecting companies. Content as data is 'food' to feed the business model of social media with the help of algorithms that tend to show users content related to their algorithmic profile.

But is profit maximizing behavior enough to incentivize appropriate moderation policies? Even though social media companies have algorithms to detect rule violations, they still must hire or subcontract thousands of people to manually review content and enforce their rules. Since moderating on a large scale is resource-intensive, it makes sense to only increase the strictness of a platform's moderation policy until the additional costs equal the additional benefits[72].

One potential benefit is avoiding future regulation. For example, an US regulation called Section 230[73] protects digital platforms from being legally responsible for what gets posted on their sites. Employees at X have said that platforms aren't motivated to keep content up because of regulatory concerns[72]. But it's unlikely that just trying to avoid regulators can cover the costs of content moderation, so this probably isn't the main reason social media companies moderate content. For one thing, social media companies started moderating content well before there was any regulatory threat, often going beyond what

the laws require[72]. Plus, even when regulation is in place, the actual cost to platforms is minimal—*Facebook* only paid 2 million Euros in fines for breaking Germany's anti-hate speech law (NetzDG)[72][74].

Another potential benefit is that better moderation might make advertisers more interested because of improved brand safety. But moderation isn't essential for this since platforms can effectively reduce toxic content and improve brand safety by optimizing their advertising loads[72].

So, theoretically, the biggest benefit that social media companies get from better content moderation is an increase in user engagement, and therefore, more advertising revenue. In practice, small increases in moderation should lead to higher engagement for some users, like those who are attacked or offended by hate speech or don't like misinformation. Also, moderation is more likely to happen when those who benefit from it increase their engagement a lot, while those who are harmed don't decrease their engagement too much[72].

In recent years, companies like *Facebook*, *Instagram*, and *X* have amassed extraordinary power through their digital platforms. Since then, society has struggled to figure out to effectively regulate these influential entities. A major obstacle to these regulatory efforts is the little information available about how they work, and they affect society. Transparency is often proposed as a solution to this problem, but its proponents rarely specify what it entails or how it should be put into practice. Transparency is crucial to improve public understanding of these platforms, ensure they are held accountable and help create a better regulatory framework[75].

It has been suggested in literature[75] that there are essentially four types of platform transparency:

- General Purpose Transparency: 'These are transparency measures for non-sensitive data that are implemented through controlled procedures'[75]. It is the most economical and secure form of visibility on the platform. Examples include how platforms implement content moderation policies, information about government data requests, content removals for copyright infringements and orders from law enforcement and governments, as well as selected data on how recommendation algorithm's function.
- Open-source transparency: This allows individuals and organizations to access non-sensitive data with minimal restrictions. Ideally, the only restrictions are those necessary to protect sensitive data.
- Supervised Transparency**:** This type of transparency is necessary to analyze sensitive data to answer certain questions. It is more expensive and difficult to implement and can be achieved through various methods.
- Finally, transparency might be detrimental if confidential information is accessed unattended. This can result, for example, from abusive use of APIs and web scraping, hacking, abuse of government or law enforcement authority[20].

Since transparency is an instrument of accountability, it depends on who the information should be shared with and what activity is subject to transparency, that is, should digital platforms be transparent to whom and for what?

i. Transparent to whom?

The information may be directed to three groups of stakeholders, according to the classification made by Leerssen[76]: (i) individual users of the platform, to inform them of how their information will be used by the platform and about decisions may be made to remove content or accounts; (ii) government oversight, regulators and other bodies; and

[20] 'The ill-famed case of Cambridge Analytica, where a researcher working at the company unduly collected personal data of millions of Facebook users by exploiting a vulnerability in the system of the platform makes for a perfect exemplification of what can happen when the wrong data ends up in the wrong hands'[75].

(iii) civil society, the public, and independent researchers. As the author explains, these three types of transparency are not mutually exclusive, but add up in line with the idea that platform governance requires a multistakeholder approach.

ii. Users

"The disclosure of information to users aims to reduce the information asymmetry between them and the platforms so that they can be understand: (i) what type of information reaches them based on the analysis of their collected data; (ii) if the content or accounts have been taken down; (iii) the reason that led to the removal, which platform rule was violated; (iv) whether they have the right to appeal the moderation decision in the event of an error or misinterpretation; (v) if they are digital influencers, if their publications are not recommended by the platform (*shadowban*[21]) or how their content is monetized[77]".

iii. State supervision

"State supervision may be under the jurisdiction of a specific body that is required to oversee the decision-making process in content moderation, in accordance with standards previously established by state regulation. The disclosure of information to the State also leads to an asymmetry of information between regulators and platforms, allowing the formulation of a more informed public policy, and the implementation of a decision-making process closer to the public interest[77]".

iv. Civil society

The dissemination of information to civil society aims to enable research into the decision-making process of platforms. The partnership between digital platforms and research groups in automated moderation systems can even be beneficial for the platforms. Collaboration between intermediaries, public representatives, civil society and independent

[21] As a political example, we could refer CHEGA'S (Portuguese political party) *shadowban*, made by Facebook: https://observador.pt/2024/04/14/conta-do-chega-no-facebook-restringida-durante-10-anos-ventura-fala-em-censura-politica-e-vai-levar-caso-ao-parlamento-e-a-justica/

researchers can lead to the formulation of monitoring mechanisms based on the investigation of the decision-making process can increase transparency, understood as observability, and increase the possibility of assessing the results of the moderation process [62][77].

v. Transparency for what?

As diverse as the recipients of the information disclosed by the platforms are, so are the activities about which transparency is required. Digital platforms may be asked to provide information on content recommendation techniques[76][78], optout, pricing, standard-setting[79], or on so-called "moderation as medicine"[80]. In the moderation decision-making process, three distinct moments can be defined, which involve the provision of information by the platforms: (i) the establishment of moderation policies and rules; (ii) make a moderation decision against a particular user's content or account; and (iii) the results (enforcement) of moderation policies and rules in the social media environment[77].

While transparency, as discussed here, is an important element for accountability and therefore the legitimacy of platforms actions, there are some limitations to the transparency requirements that need to be recognized. In view of what can be observed from the information already disclosed, what kind of information can be extracted from the platforms that enable accountability? In other words, what kind of information can be shared with its recipients in a visible and understandable way to allow external control of the actions of the platforms and, ultimately, their responsibility for inappropriate behavior? These limitations can be found in the context of digital platforms, especially regarding the performance of algorithmic automated moderation systems. Mechanisms for disseminating information through algorithms, codes with machine learning protocols, can make certain information visible, but do not increase the recipient's ability to understand the information about the system.

Ananny and Crawford[81] explain that, even in the digital context, mechanisms of information dissemination refer to networks of interactions between human and non-human actors, which creates specific conditions for the visibility and observation of this context. The authors also argue that for the system to be understood, and not just seen, the focus of transparency must be on seeing "across a system". Because algorithms are "black boxes", digital platforms do not have full control over the decisions made. Even access to the code may not be enough to understand automated decisions.

Examining the inner workings of a machine-learning system is fundamentally different from conducting accounting audits, for example[82]. Therefore, the possibility of holding platforms accountable must consider what can be made available by the algorithms.

For this reason, it has been argued that the focus of regulation should be on improving researchers' access to platform data, so that they can extract relevant information about moderation systems from this data[82][83]. This extraction of relevant information from the survey is the difference between transparency, that simply provides aggregated data, and transparency, which can increase accountability.

4. STUDY CASES

⇒ L'Oréal v. eBay (2011)[84]

Case Summary: *L'Oréal*, a French cosmetics company, holds several national trademarks in the UK and community trademarks across the European Union. After discovering widespread trademark infringement on *eBay*, *L'Oréal* filed legal actions against the online marketplace, its European subsidiaries, and individual sellers of counterfeit products mimicking its brand names. *L'Oréal* contended that *eBay* was responsible for the misuse of its trademarks through the display of these marks on its site and in sponsored search engine advertisements, such as those on *Google*. In 2009, the UK Chancery Division of the High Court of Justice halted the proceedings to seek a preliminary ruling from the European

Court of Justice (ECJ) on relevant EU directives. The ECJ ruled that, according to EU Directive 89/104[85] and Regulation 40/94[86], trademark holders can prevent online marketplace operators from advertising goods without permission if those goods are intended for consumers in the EU. Furthermore, the Court decided that *eBay* could not avoid liability under Article 14(1) of Directive 2000/31[87] if it actively participated in the sale of the goods by enhancing their presentation or promoting them[88].

Economic Implications:

- Compliance and Monitoring Costs: E-commerce platforms had to invest in better monitoring and compliance systems to detect and prevent the sale of counterfeit goods.
- Operational Changes: Platforms needed to implement stricter seller verification processes and proactive measures to prevent infringements.
- Impact on Freedom of Expression: The ruling affects how user content is managed, potentially leading to over-censorship to avoid liability.

⇒ **YouTube and Article 17 of the EU Copyright Directive**[89][90]

Case Summary: The EU Copyright Directive[91], particularly Article 17, required online platforms like *YouTube* to ensure that uploaded content does not infringe on copyright, mandating the use of upload filters.

In terms of content moderation, this means that online platforms are required to take proactive measures to prevent the upload of copyrighted content without permission. This often involves implementing upload filters or content recognition technologies.

To be compliant with the law**,** Platforms must negotiate licensing agreements with copyright holders to ensure that content can be legally used and monetized.

However, the directive provides some exemptions for smaller platforms with an annual turnover below EUR 10 million, easing their compliance burdens (article 17(6)).

Economic Implications for *YouTube*:

1. Increased Operational Costs:
    - Implementation of Filters**:** *YouTube* had to invest heavily in developing and refining Content ID, its proprietary content recognition system, to comply with the directive[92]. This system identifies copyrighted material and manages licensing and revenue-sharing.
    - Administrative Costs: Managing and negotiating licensing agreements with a wide range of copyright holders adds significant administrative overhead.
2. Impact on User-Generated Content:
    - Increased Censorship: With stricter content controls, some legitimate user-generated content may be inadvertently blocked or removed, impacting the diversity and volume of content available on *YouTube*.
    - User Experience: The effectiveness of content filters can affect user experience, as overly aggressive filtering might lead to the removal of content that falls under fair use or transformative use.
3. Legal and Compliance Risks:
    - Litigation Risks**:** Non-compliance with the new regulations can result in legal challenges and financial penalties. The complexity of copyright law and the scale of content on *YouTube* pose ongoing compliance challenges.
    - Reputation Management: Balancing copyright enforcement with user rights is crucial for maintaining *YouTube's* reputation as a platform that supports content creators while respecting copyright laws.
4. Market Dynamics:

- Competitive Impact: The costs and operational changes required to comply with Article 17 may affect *YouTube's* competitive position relative to smaller or newer platforms that can take advantage of the exemptions provided for smaller services.
- Content Availability: The directive could lead to changes in how content is made available online, potentially limiting access to certain works and impacting the overall content ecosystem.

In response to the directive, *YouTube* and other major platforms have had to enhance their content moderation systems and engage in extensive negotiations with copyright holders. The implementation of stricter content filters and increased licensing fees has altered the dynamics of online content distribution and monetization[93][94].

In conclusion**,** Article 17 of the EU Copyright Directive has had a profound economic impact on *YouTube* by increasing operational costs, influencing user experience, and altering market dynamics. Compliance with these new regulations requires substantial investment in technology and legal resources, which can affect the platform's competitiveness and the availability of content for users.

⇒ **Delfi AS v. Estonia**[95]

Background: The Delfi AS case is a landmark legal battle concerning the liability of online platforms for user-generated content. Delfi AS, an Estonian news portal, was sued over defamatory comments posted by anonymous users on one of its articles. The case went through various levels of the Estonian courts before reaching the European Court of Human Rights (ECHR).

In January 2006, Delfi published an article about a ferry company[96]. The article received several anonymous comments, some of which were defamatory towards the company's owner. The company's owner sued Delfi AS for defamation, arguing that the platform was

responsible for the user comments. The Estonian courts ruled in favor of the plaintiff, holding Delfi liable for the defamatory comments. Delfi AS appealed the decision, arguing that holding the platform liable violated its freedom of expression under Article 10 of the European Convention on Human Rights[96].

European Court of Human Rights Decision:

- Initial Ruling (2013)**:** The ECHR upheld the Estonian court's decision, stating that Delfi AS was responsible for the comments because it had not taken sufficient measures to prevent defamatory content from being published.
- Grand Chamber Ruling (2015)**:** Upon further appeal, the ECHR Grand Chamber confirmed the initial ruling, emphasizing that Delfi AS exercised a substantial degree of control over the comments section and profited from the content.

The Court noted that Delfi had a high level of control over the comments, as it automatically filtered certain words, had a notice-and-take-down system, and allowed anonymous comments. It was also observed that Delfi derived financial benefit from the comments section, which contributed to its liability. When balancing freedom of expression with the right to reputation of individuals, the Court concluded that the latter took precedence in this case.

Economic Implications:

1. Increased Operational Costs: Platforms like Delfi AS may need to invest more in content moderation technologies and staff to avoid liability for user-generated content.
2. Legal Risks and Insurance: Online platforms could face higher legal risks, necessitating increased legal compliance efforts and potentially higher insurance premiums against defamation claims.

3. Impact on User Engagement: Stricter moderation might deter user engagement and reduce the volume of comments and user interactions, affecting advertising revenues.
4. Platform Liability: This case set a precedent in Europe, influencing how other online platforms handle user-generated content and the extent of their liability.

5. CONCLUSION

Digital platforms are now central to modern communication, influencing every aspect of the way we connect, share, and interact. While they provide unimaginable opportunities for networking and collaboration, it is important to keep in mind that they also require careful analysis of the challenges involved.

The interaction between the right to freedom of expression and regulation in the digital age reveals a complex and evolving scenario, where traditional legal frameworks are increasingly defied by rapid technological progress.

While regulation is important and necessary to address issues such as disinformation, hate speech and data privacy, it also poses significant risks to the fundamental right to freedom of expression, especially when implemented without transparency, proportionality or liability.

The impact of the new regulatory framework on the market is profound, since companies operating in the digital sphere must create a balance between compliance and innovation, especially in the case of small businesses that, as we have seen, lack the resources to adapt as easily to stringent requirements. On the other hand, underregulating is also a situation to avoid because it can lead to market monopolization by dominant platforms, which complicates the aim to protect user rights and the promotion of diversity of opinions.

In the end, the challenge lies in creating a regulatory framework that safeguards both freedom of expression and market competitiveness, ensuring that the digital space remains a vibrant forum for ideas, while protecting its users from harm. This balance is not only essential for the sake of democratic discourse, but also for the sustainable growth of the digital economy in the future.

The upcoming years should aim for ongoing dialogue between authorities, businesses, civil society and users in order to promote a regulatory environment that is fair and adaptable in a time of rapid change.

6. BIBLIOGRAPHY

[1] B. M. Leiner *et al.*, "Brief History of the Internet 1997." [Online]. Available: http://www.acm.org

[2] R. F. Jørgensen and L. Zuleta, "Private Governance of Freedom of Expression on Social Media Platforms EU content regulation through the lens of human rights standards," *Nordicom Review*, vol. 41, no. 1, pp. 51–67, Jan. 2020, doi: 10.2478/nor-2020-0003.

[3] M. Sharma and P. Bhandarkar, "INTERNATIONAL JOURNAL OF LAW MANAGEMENT & HUMANITIES Freedom of Speech and Expression on Internet: An Emerging Right," *International Journal of Law Management & Humanities*, vol. 3, 2020, [Online]. Available: https://www.vidhiaagaz.com

[4] "Universal Declaration of Human Rights Preamble."

[5] E. - Cedh, "European Convention on Human Rights." [Online]. Available: www.conventions.coe.int.

[6] "International Covenant on Civil and Political Rights."

[7] A. Das, "Effect of internet on freedom of speech & expression ." Accessed: May 07, 2024. [Online]. Available: https://legalvidhiya.com/effect-of-internet-on-freedom-of-speech-expression/

[8] G. De Gregorio, *Digital Constitutionalism in Europe*. Cambridge University Press, 2022. doi: 10.1017/9781009071215.

[9] L. Roberto Barroso and L. van Brussel Barroso, "Former President of the Brazilian Superior Electoral Court," Masters in Public Law, 1989.

[10] F. Damania, "THE INTERNET: EQUALIZER OF FREEDOM OF SPEECH? A DISCUSSION ON FREEDOM OF SPEECH ON THE INTERNET IN THE UNITED STATES AND INDIA." [Online]. Available: http://www.nua.ie/surveys/how-manyonlineln_america.html

[11] T. Gillespie, *Custodians of the Internet*. Yale University Press, 2019. doi: 10.12987/9780300235029.

[12] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Bus Horiz*, vol. 53, no. 1, pp. 59–68, Jan. 2010, doi: 10.1016/j.bushor.2009.09.003.

[13] Gallen Ray Canillo, "Should Free Speech Be Allowed On Social Media?," Oct. 2022, Accessed: Jul. 11, 2024. [Online]. Available: https://webfriendly.com/social-media-free-speech/

[14] Chekkee, "What Are the Pros and Cons of Censorship on Social Media?" Accessed: Jul. 11, 2024. [Online]. Available: https://chekkee.com/the-pros-and-cons-of-censorship/

[15] J. Barata, "Freedom of Expression and Privacy on Social Media: The Blurred Line Between the Private and the Public Sphere." [Online]. Available: https://rm.coe.int/1680790e14

[16] J. Horowitz, “The First Amendment, Censorship, and Private Companies: What Does ‘Free Speech’ Really Mean?,” Carnegie Library. Accessed: Jul. 23, 2024. [Online]. Available: https://www.carnegielibrary.org/the-first-amendment-and-censorship/

[17] K. Bontcheva *et al.*, *Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression Broadband Commission research report on “Freedom of Expression and Addressing Disinformation on the Internet” 2*. 2020.

[18] Chekkee, “The Ethics of Content Moderation: Balancing Free Speech and Harm Prevention.” Accessed: Jul. 11, 2024. [Online]. Available: https://chekkee.com/the-ethics-of-content-moderation-balancing-free-speech-and-harm-prevention/

[19] S. Gardbaum, “Michigan Law Review Michigan Law Review The ‘Horizontal Effect’ of Constitutional Rights The ‘Horizontal Effect’ of Constitutional Rights Recommended Citation Recommended Citation THE ‘HORIZONTAL EFFECT’ OF CONSTITUTIONAL RIGHTS.” [Online]. Available: https://repository.law.umich.edu/mlrhttps://repository.law.umich.edu/mlr/vol102/iss3/2

[20] “Content moderation on digital platforms: A more intensive horizontal effect of freedom of expression?,” *Alexander Von Humboldt Institut Fur Internet und Gesellschaft*. Accessed: May 07, 2024. [Online]. Available: https://www.hiig.de/en/content-moderation-on-digital-platforms/

[21] D. Kwieciński, “Measures of Competitive Intensity – Analysis Based on Literature Review,” *Journal of Management and Business Administration. Central Europe*, vol. 25, no. 1, pp. 53–77, Jan. 2017, doi: 10.7206/jmba.ce.2450-7814.189.

[22] X, “Permanent suspension of @realDonaldTrump.” Accessed: Jul. 12, 2024. [Online]. Available: https://blog.x.com/en_us/topics/company/2020/suspension

[23] Nicholas Confessore, “Cambridge Analytica and Facebook: The Scandal and the Fallout So Far,” *New York Times*, Apr. 2018, Accessed: Jul. 12, 2024. [Online]. Available: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

[24] H. O. Y. Li, E. Pastukhova, O. Brandts-Longtin, M. G. Tan, and M. G. Kirchhof, “YouTube as a source of misinformation on COVID-19 vaccination: A systematic analysis,” *BMJ Glob Health*, vol. 7, no. 3, Mar. 2022, doi: 10.1136/bmjgh-2021-008334.

[25] Max Zahn, “Is TikTok different in China? Here’s what to know,” *abc News*, Mar. 2024, Accessed: Jul. 12, 2024. [Online]. Available: https://abcnews.go.com/Business/tiktok-china/story?id=108111708

[26] R. F. Jørgensen© and G. Noll, “EUROPEAN MASTER DEGREE IN HUMAN RIGHTS AND DEMOCRATISATION 2000-2001 RAOUL WALLENBERG INSTITUTE Internet and Freedom of expression.”

[27] *“Hate speech” explained : a toolkit*. Article 19, 2015.

[28] “REPORT FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL on the implementation of Council Framework Decision 2008/913/JHA on combating certain forms and expressions of racism and

xenophobia by means of criminal law /* COM/2014/027 final */", Accessed: May 21, 2024. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A52014DC0027

[29] K. Coe, K. Kenski, and S. A. Rains, "Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments," *Journal of Communication*, vol. 64, no. 4, pp. 658–679, Aug. 2014, doi: 10.1111/jcom.12104.

[30] E. Burch and E. Burch', "Censoring Hate Speech in Cyberspace: A New Debate in New America Comment: Censoring Hate Speech In Cyberspace: A New Debate in a New America," 2001. [Online]. Available: http://scholarship.law.unc.edu/ncjolt/vol3/iss1/10

[31] K. Huiskes and UVA | MILLER CENTER, "The September 11 Terrorist Attacks: the day that defined the beginning of the 21st Century for Americans. ." Accessed: Jul. 12, 2024. [Online]. Available: https://millercenter.org/remembering-september-11/september-11-terrorist-attacks

[32] L. Belli and J. Venturini, "Private ordering and the rise of terms of service as cyber-regulation," *Internet Policy Review*, vol. 5, no. 4, Dec. 2016, doi: 10.14763/2016.4.441.

[33] M. L. Mueller and F. Badiei, "Inventing Internet Governance: The Historical Trajectory of the Phenomenon and the Field," in *Researching Internet Governance*, The MIT Press, 2020, pp. 59–84. doi: 10.7551/mitpress/12400.003.0004.

[34] R. Kern, S. Sutton, R. Reader, and T. Snyder, "Why Elon Musk's 'X App' could be an even bigger headache for D.C. than Twitter," *POLITICO*, Mar. 2023, Accessed: Jul. 15, 2024. [Online]. Available: https://www.politico.com/news/2023/01/03/elon-musk-x-app-collides-washington-00075467

[35] R. Radu, "Privatization and Globalization of the Internet".

[36] A. A. Anderson, S. K. Yeo, D. Brossard, D. A. Scheufele, and M. A. Xenos, "Toxic Talk: How Online Incivility Can Undermine Perceptions of Media," *Int J Public Opin Res*, vol. 30, no. 1, pp. 156–168, Mar. 2018, doi: 10.1093/ijpor/edw022.

[37] Distr, "UNITED NATIONS CCPR International covenant on civil and political rights The Nature of the General Legal Obligation Imposed on States Parties to the Covenant," 2004.

[38] Prof. Rasmus Kleis Nielsen and Dr. Mackenzie F. Common, "How to respond to disinformation while protecting free speech ," *Reuters Institute for the study of journalism*, Feb. 2021, Accessed: Jun. 24, 2024. [Online]. Available: https://drive.google.com/file/d/1bUfwO50LR7laFIeh9W42LU-98znYdbSl/view

[39] "GuidinG PrinciPles on Business and Human riGHts Implementing the United Nations 'Protect, Respect and Remedy' Framework," 2011.

[40] "The Santa Clara Principles On Transparency and Accountability in Content Moderation." Accessed: Jul. 11, 2024. [Online]. Available: https://santaclaraprinciples.org

[41] "I (Legislative acts) REGULATIONS REGULATION (EU) 2022/2065 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 19 October 2022 on a

Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance)."
[42] Europe Commission, "The Digital Services Act." Accessed: Jul. 10, 2024. [Online]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en
[43] Europe Commission, "The impact of the Digital Services Act on digital platforms." Accessed: Jul. 10, 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/dsa-impact-platforms
[44] Z. Yang, "China is escalating its war on kids' screen time," MIT Technology Review. Accessed: Jul. 15, 2024. [Online]. Available: https://www.technologyreview.com/2023/08/09/1077567/china-children-screen-time-regulation/
[45] S. Sharwood, "China floats strict screentime limits and content crimps for kids," *The Register*, Aug. 03, 2023. Accessed: Jul. 15, 2024. [Online]. Available: https://www.theregister.com/2023/08/03/china_kids_internet_restrictions_plan/
[46] Y. Li, "Why does China want kids to spend less time on their phones?," Taipei, Feb. 09, 2023. Accessed: Jul. 15, 2024. [Online]. Available: https://www.dw.com/en/why-does-china-want-kids-to-spend-less-time-on-their-phones/a-66694947
[47] M. Maroni, "'Mediated transparency': The Digital Services Act and the legitimisation of platform power," in *(In)visible European Government: Critical Approaches to Transparency as an Ideal and a Practice*, Taylor and Francis, 2023, pp. 305–326. doi: 10.4324/9781003257936-19.
[48] J. M. Agostinho, L. P. Marques, and L. B. Espindola, "DSA: o novo regime para os serviços intermediários," CUATRE CASAS. Accessed: Jul. 25, 2024. [Online]. Available: https://www.cuatrecasas.com/pt/portugal/propriedade-intelectual/art/dsa-o-novo-regime-para-os-servicos-intermediarios-1
[49] I. Buri and J. van Hoboken, "The DSA Proposal's Impact on Digital Dominance," 2021. doi: 10.17176/20210830-112903-0.
[50] Intersoft Consulting, "General Data Protection Regulation", Accessed: Jul. 11, 2024. [Online]. Available: https://gdpr-info.eu
[51] C. Clune and E. McDaid, "Content moderation on social media: constructing accountability in the digital space," *Accounting, Auditing & Accountability Journal*, vol. 37, no. 1, pp. 257–279, Jan. 2024, doi: 10.1108/AAAJ-11-2022-6119.
[52] J. M. Balkin, "Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation," 2018.
[53] K. Klonick, "THE NEW GOVERNORS: THE PEOPLE, RULES, AND PROCESSES GOVERNING ONLINE SPEECH." [Online]. Available: http://www.startribune.com/
[54] C. Yongmei and J. Afzal, "Impact of Enactment of 'The Prevention of Electronic Crimes Act, 2016' as Legal Support in Pakistan," *Academy of Education and Social Sciences Review*, vol. 3, no. 2, pp. 203–212, May 2023, doi: 10.48112/aessr.v3i2.500.

[55] "Law on Cyber Security 2018".
[56] James Pearson, "Exclusive: Facebook agreed to censor posts after Vietnam slowed traffic - sources," *Reuters*, Apr. 2020, Accessed: Jul. 11, 2024. [Online]. Available: https://www.reuters.com/article/world/uk/exclusive-facebook-agreed-to-censor-posts-after-vietnam-slowed-traffic-source-idUSKCN2232K2/
[57] "Türkiye: Big Tech Should Protect Online Expression, Resist Censorship Rigorous transparency measures needed as government pressure grows," *HRW*, Accessed: Jul. 11, 2024. [Online]. Available: https://www.hrw.org/news/2024/03/04/turkiye-big-tech-should-protect-online-expression-resist-censorship
[58] "Network Enforcement Act Regulatory Fining Guidelines Guidelines on setting regulatory fines within the scope of the Network Enforcement Act (Netzwerkdurchsetzungsgesetz-NetzDG)," 2018.
[59] Meta, "Facebook Community Standards." Accessed: Jul. 11, 2024. [Online]. Available: https://transparency.meta.com/pt-pt/policies/community-standards/?source=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards%2F
[60] X, "THE X RULES." Accessed: Jul. 11, 2024. [Online]. Available: https://help.x.com/en/rules-and-policies/x-rules
[61] Oversight Board, "IMPROVING HOW META TREATS PEOPLE AND COMMUNITIES AROUND THE WORLD." Accessed: Jul. 11, 2024. [Online]. Available: https://www.oversightboard.com
[62] "Manila Principles on Intermediary Liability Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation MANILA PRINCIPLES ON INTERMEDIARY LIABILITY • MANILAPRINCIPLES.ORG Manila Principles on Intermediary Liability."
[63] R. Gorwa, R. Binns, and C. Katzenbach, "Algorithmic content moderation: Technical and political challenges in the automation of platform governance," *Big Data Soc*, vol. 7, no. 1, p. 205395171989794, Jan. 2020, doi: 10.1177/2053951719897945.
[64] A. Puddephatt, "A Importância da autorregulação da mídia para a defesa da liberdade de expressão; Series CI debates: communication and information; Vol.:9; 2011," 2011, [Online]. Available: www.fordfound.org
[65] R. Peeters and M. Schuilenburg, "The algorithmic society An introduction."
[66] L. Freund, "Open Peer Review on Qeios The Age of the Algorithmic Society-A Girardian Analysis of Mimesis, Rivalry, and Identity in the Age of Artificial Intelligence," 2023, doi: 10.32388/C9FER2.
[67] "Policing Content in the Quasi-Public Sphere." [Online]. Available: http://opennet.net.
[68] H. Level Group, "Final Report of the High Level Group on Media Freedom and Pluralism."
[69] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a Feather: Homophily in Social Networks," *Annu Rev Sociol*, vol. 27, no. 1, pp. 415–444, Aug. 2001, doi: 10.1146/annurev.soc.27.1.415.

[70] T. Flew, F. Martin, and N. Suzor, “Internet regulation as media policy: Rethinking the question of digital com- munication platform governance.”, Accessed: May 24, 2024. [Online]. Available: https://eprints.qut.edu.au/126702/1/05_JDTV_10_1_art_Flew_Martin_Suzor.pdf
[71] Global Digital Cultures, “Platform Empires: Navigating the Terrain of Data Colonialism and the Movement Towards Digital Decolonization.” Accessed: Jul. 30, 2024. [Online]. Available: https://globaldigitalcultures.org/2024/03/05/platform-empires-navigating-the-terrain-of-data-colonialism-and-the-movement-towards-digital-decolonization/
[72] R. Duran-Jiménez, “The Economics of Content Moderation on Social Media,” *Promarket*, Nov. 2022, Accessed: Jul. 24, 2024. [Online]. Available: https://www.promarket.org/2022/11/10/the-economics-of-content-moderation-on-social-media/
[73] “Section 230: An Overview.” [Online]. Available: https://crsreports.congress.gov
[74] J. Delcker, “Germany fines Facebook €2M for violating hate speech law,” *POLITICO*, Jul. 2019, Accessed: Jul. 24, 2024. [Online]. Available: https://www.politico.eu/article/germany-fines-facebook-e2-million-for-violating-hate-speech-law/
[75] A. Leone de Castris, “Types of Platform Transparency: An Analysis of Discourse Around Transparency and Global Digital Platforms,” *Public Integrity*, 2024, doi: 10.1080/10999922.2024.2304741.
[76] P. Leerssen, “The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems,” 2020.
[77] C. Leitão, D. E. Almeida, and R. De Janeiro, “FUNDAÇÃO GETULIO VARGAS FGV DIREITO RIO REGULAÇÃO DA TRANSPARÊNCIA EM PLATAFORMAS DIGITAIS E LEGITIMIDADE NA MODERAÇÃO DE CONTEÚDO.”
[78] Ben. Wagner, *Global free expression : governing the boundaries of Internet content*. Springer, 2016.
[79] C. Law *et al.*, “The Virtues of Moderation,” 2015. [Online]. Available: http://scholarship.law.cornell.edu/facpub
[80] E. Goldman, “CONTENT MODERATION REMEDIES.” [Online]. Available: http://www.ericgoldman.org.
[81] M. Ananny and K. Crawford, “Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability,” *New Media Soc*, vol. 20, no. 3, pp. 973–989, Mar. 2018, doi: 10.1177/1461444816676645.
[82] B. Rieder and J. Hofmann, “Towards platform observability,” *Internet Policy Review*, vol. 9, no. 4, Dec. 2020, doi: 10.14763/2020.4.1535.
[83] J. Ausloos and P. Leerssen, “Operationalizing Research Access in Platform Governance What to learn from other industries? GOVERNING PLATFORMS,” 2020. [Online]. Available: https://www.ivir.nl;

[84] S. Sagar and T. Hoffmann, "Intermediary Liability in the EU Digital Common Market – from the E-Commerce Directive to the Digital Services Act", [Online]. Available: https://idp.uoc.edu
[85] Directive 89/104, "Official journal of the European Communities."
[86] Regulation No 40/94, "Official journal of the European Communities."
[87] "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)."
[88] Global Freedom of Expression - Columbia University, "L'0réal SA v. eBay International AG," Global Freedom of Expression - Columbia University. Accessed: Jul. 29, 2024. [Online]. Available: Global Freedom of Expression - Columbia University
[89] Bateman. Robert, "Complying with Article 17 of the EU Copyright Directive." Accessed: Jul. 29, 2024. [Online]. Available: https://www.termsfeed.com/blog/eu-copyright-directive-article-17/
[90] Alexander. Julia, "YouTube creators are still trying to fight back against European copyright vote / 'With Article 13, things can only get even worse,'" *The Verge*, Mar. 27, 2019. Accessed: Jul. 29, 2024. [Online]. Available: https://www.theverge.com/2019/3/27/18283800/youtube-copyright-directive-article-13-memes-grandayy-philip-defranco-european-union
[91] Directive 2019/790, "Official journal of the European Communities."
[92] Music Bed, "YouTube Content ID Explained for Video Creators." Accessed: Jul. 29, 2024. [Online]. Available: https://www.musicbed.com/articles/resources/youtube-content-id/
[93] J. B. Nordemann, "YouTube's first Copyright Transparency Report 2021 – A step towards 'factfulness.'" Accessed: Jul. 29, 2024. [Online]. Available: https://copyrightblog.kluweriplaw.com/2022/01/20/youtubes-first-copyright-transparency-report-2021-a-step-towards-factfulness/
[94] Youtube Team, "Access for all, a balanced ecosystem, and powerful tools." Accessed: Jul. 29, 2024. [Online]. Available: https://blog.youtube/news-and-events/access-all-balanced-ecosystem-and-powerful-tools/
[95] Global Freedom of Expression - Columbia University, "Delfi AS v. Estonia." Accessed: Jul. 29, 2024. [Online]. Available: https://globalfreedomofexpression.columbia.edu/cases/delfi-as-v-estonia/
[96] "Making an Internet news portal liable for the offensive online comments of its readers was justified", Accessed: Jul. 29, 2024. [Online]. Available: https://www.indexoncensorship.org/wp-content/uploads/2013/10/udgment-Delfi-AS-v.-Estonia-making-an-online-news-site-liable-for-its-readers-comments-was-justifie.pdf
[97] R. Standish, "Interview: Will The Russian Internet Resemble China's 'Great Firewall'?," RadioFreeEurope Radio Liberty. Accessed: Jul. 23, 2024. [Online].

Available: https://www.rferl.org/a/russia-internet-china-great-firewall-censorship-meta-facebook-instagram/31765408.html
[98] C. H. Smith, J. Merkley, S. Oregon, J. Collins, I. Risch, and S. D. Harris, "CONGRESSIONAL-EXECUTIVE COMMISSION ON CHINA LEGISLATIVE BRANCH COMMISSIONERS House Senate."
[99] P. Mozur, "China Presses Its Internet Censorship Efforts Across the Globe," *The New York Times*, Mar. 2018, Accessed: Jul. 23, 2024. [Online]. Available: https://www.nytimes.com/2018/03/02/technology/china-technology-censorship-borders-expansion.html
[100] Euronews with AFP, "Twitter's free speech approach backed by Taliban official," Jul. 11, 2023. Accessed: Jul. 15, 2024. [Online]. Available: https://www.euronews.com/2023/07/11/twitters-approach-to-free-speech-backed-by-taliban-official
[101] Afghan Witness, "Surge in Verified Taliban Accounts on X," Afghan Witness. Accessed: Jul. 15, 2024. [Online]. Available: https://www.afghanwitness.org/reports/surge-in-verified-taliban-accounts-on-x
[102] B. Keskin, "Van Dijk, Poell, and de Wall, The Platform Society: Public Values in a Connective World (2018)," *Markets, Globalization & Development Review*, vol. 03, no. 03, 2018, doi: 10.23860/mgdr-2018-03-03-08.

Navigating Cyberspace: Unraveling the Future of Freedom of Expression in the Age of Digital Platforms