THE UNIVERSITY OF CHICAGO

Types of Platform Transparency: An analysis of digital platforms and policymakers discourse on Big Tech governance and transparency

By

Arcangelo Leone de Castris

August 2022

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts Program in the Social Sciences

Faculty Advisor: Prof. Geoffrey Stone
Preceptor: Dr. Yan Xu

Table of Contents

Introduction 3

Methodology 6

Part 1 – A Genealogy of Platform Transparency 7

1.1 – Transparency as a doctrine of democratic governance: making government and corporations accountable to the public 8

1.2 – From the promise of radical transparency to contemporary data enclosures: Transparency in the Digital Age 11

1.3 – Transparency and digital platforms 13

Part 2 – Defining features of the emerging concept of platform transparency: Discourse analysis and the extraction of semantic patterns 19

2.1 – Transparency in the public discourse of major digital platforms 20

2.2 – Transparency in the public discourse of the U.S. Congres 25

PART 3 – Platform transparency: A conceptual framework 31

3.1 – Key dimensions and types of platform transparency 31

Conclusion 36

Bibliography 38

List of Tables

Figure 1 – Digital platforms' discourse on transparency……20

Figure 2 – Most distinctive words: Tf-Idf scores……..……..22

Figure 3 – Topic modelling of platforms transparency communications……...……..23

Figure 4 – U.S. Congress discourse on platform transparency……..……..26

Figure 5 – Most distinctive words: U.S. Congress……..28

Figure 6 – Higher word probability by topic……..29

Figure 7 – Typology of Platform Transparency……..32

Introduction

Today, after almost thirty years from the onset of the internet revolution, our society is grappling with fundamental questions about how to harness the transformative potential of digital technologies while also preventing the destructive forces that their abuse can unleash. Central to this debate is the conversation around how to regulate digital platforms and, particularly, social media – i.e., the private companies that operate services like Facebook, YouTube, Twitter, WhatsApp, TikTok, etc.

From a regulatory standpoint, digital platforms have been granted nearly absolute autonomy in setting the rules of their nascent business (Cusumano *et al.*, 2021). This model, known as self-regulation,[1] has in effect proved essential for enabling the unprecedented growth of digital platforms. At the same time, however, left unchecked platforms have been systematically caught abusing and misusing their power (Mozur et al., 2021). As a result of a long trail of scandals, court trials, and unkept promises that eroded public trust in digital platforms, policymakers around the world have faced increasing pressure to establish principles and limits to prevent the spiraling out of control of platform governance.

Regulating digital platforms, however, is no easy task. These companies are drivers of social change, and the services that they offer have become constitutive of the very structure of contemporary society: they provide spaces where people gather to see friends, get informed, discuss both personal and public matters, find jobs, buy products, and simply live their digital lives. Ill-conceived legislation can stifle innovation, restrict freedom of expression, slow down technological innovation and economic growth (Balkin, 2020). Not only platform regulation is so delicate, it is also very complex. In fact, examples of well-inteded but ineffective, if not harmful, tentative regulation are plentiful (Weissmann, 2019). One of the primary reasons why it is so difficult to effectively regulate digital

[1] "Self-regulation refers to any system of regulation in which the regulatory target—either at the individual-firm level or sometimes through an industry association that represents targets— imposes commands and consequences upon itself" (Coglianese and Mendelson, 2010: 6).

platforms is that very little is known of how they operate. Nearly all available information on their corporate culture, procedures and systems comes from leaked documents, whistleblowers and independent investigations. Equally obscure are the implications of the impact of platforms on society. Nobody except from platforms themselves really knows, for example, to what extent their recommendation algorithms amplify divisive or harmful content, or how exactly they handle user data. This is concerning because to harden out democracies to the challenges posed by digital platforms "we need research and legislation: research to understand the threat, and legislation to neutralize it" (Aral, 2020: 310). Today, however, research is drastically limited by the difficulty to access digital platform data.

Over the last decade, scholars, policymakers, civil society and digital platforms have engaged in an open debate about the opportunity of making digital platforms more 'transparent' and accountable to the public. With time, this conversation has grown deeper and more focused, ultimately leading to the introduction of transparency-centered bills in the United States,[2] and to the adoption of relevant legislation in the European Union[3] and United Kingdom.[4] Nevertheless, the debate around platform transparency still lacks important context as to what exactly should be made transparent, how, and to who. As a consequence, calls for greater platform transparency are often vague and superficial. And this is far from ideal because, as showed by the relevant literature, disclosing information indiscriminately, without a thought criteria "doesn't automatically enable people to make more informed choices" (Fung & Graham, 2007: xv) and does not necessarily lead to accountability (Flyverom, 2016). In order to be effective, transparency needs to be targeted: specific information

[2] E.g., the Platform Accountability and Consumer Transparency Act (PACT), S.797 – 117th Cong. (2021).
[3] European Commission (2022), 'Digital Services Act,' accessed 07/01/2022, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en.
[4] UK.Gov (2022), 'Online Safety Bill,' accessed 07/01/2022, https://www.gov.uk/government/publications/online-safety-bill-supporting-documents/online-safety-bill-factsheet.

shall be disclosed to specific subjects based on the value that that information holds for them (Fung & Graham, 2007).

Addressing the need to provide structure to the conversation around platform transparency, this paper proposes an exercise in concept clarification. Its objective is to provide a systematic framework for thinking about platform transparency as a necessary component of the larger conversation around the governance of digital platforms. In particular, based on a qualitative and computational analysis of the discourse around platform transparency as framed by both digital platforms and U.S. policymakers, this paper performs a twofold conceptual operation: first, it reconstructs the historical trajectory of the conversation around platform transparency and contextualizes it within the broader literature on transparency and governance (see Part 1); secondly, it identifies the most prominent features of the conversation analyzed (see Part 2) and locates them on a formal conceptual space represented by a typology matrix (see Part 3).

By so doing, this papers proposes a possible interpretation of the constitutive dimensions of a complex and often ambiguous debate, and offers a formal hermeneutical device to locate specific topics in the thick conceptual space of platform transparency.

Methodology

This research is centered around the analysis of textual data relative to the public discourse of digital platforms and the U.S. Congress around transparency in the context of platform governance.

The analysis relies on both interpretive and computational methods – combining these two types of methodological approaches for extracting meaning from text data, in fact, has already yielded positive results (Wiedemann, 2016).

More specifically, Part 1 of this paper aims at historicizing the emergence of platform transparency in public discourse by tracing its genealogy. To this end, genealogy makes for a generative method in

that it focuses the narrative on the effect that historically and contingently produced discourses exercise on the formation of concepts and subjectivities (Barnett & Duvall, 2005). Moreover, the results of this qualitative inquiry also serve the purpose of aiding in the interpretation of the findings presented in Part 2, coherently with the specific methodology adopted therein.

In fact, Part 2 implements a methodological framework called 'computational grounded theory' (Nelson, 2020). Nelson's method is iterative and includes validating the results of an unsupervised learning model for text analysis with the qualitative exploration of the data. The model chosen for the present study is a structural topic model, or 'stm.' The core assumption of this model is that that documents are the results of a combination of multiple topics, and can be measured based on the prominence of certain topics over others (Roberts *et al.*, 2019). In general, unsupervised methods of text analysis have proven useful in tracking the variations of concepts over time (Rodman, 2021), and structural topic models are particularly effective in detecting latent semantic structures in complex textual data (Grajzl & Murrell, 2019).

Finally, Part 3 expands on the findings of Part 1 and 2 with the aim of constructing a heuristic device to classify qualitatively different types of platform transparency. Typologies, in fact, are particularly helpful to map the fundamental dimensions of complex concepts and, especially when formatted as matrices, they force the researcher to be rigorous in thinking about the boundaries between types, their coherence and the relations among them (Collier, 2012).

Part 1 – A Genealogy of Platform Transparency

The concept of platform transparency is a direct progeny of the more traditional idea of transparency applied to government and corporate affairs, and builds on many of the considerations that animated that earlier debate. Therefore, to capture some of the nuances of the ongoing conversation around

platform transparency, it is necessary to take a step back and look at where that conversation originated in the first place.

1.1 – Transparency as a doctrine of democratic governance: making government and corporations accountable to the public

From a purely theoretical standpoint, the origins of transparency can be traced back to the Enlightenment, and to the thinking of authors like Jean-Jacque Rousseau and Jeremy Bentham. In particular, in the late XVIII Century questions of public accountability and openness/secrecy in government entered the nascent field of modern political analysis through the writings of some of the most influential authors of the time. For example, in his essay proposing a new constitutional regime for Poland, Rousseau argued that "if the rich man wants to shine in his fatherland, let him have no choice but to serve it […] and to aspire […] to posts that only public approbation can bestow on him and that public blame can take away from him" (Rousseau, [1772/1782] 1985: 72); and, in his proposal to reform the poor laws in England, Bentham noted that "secrecy, being an instrument of conspiracy, ought never be the system of regular government" (Bentham, [1790s] 2001: 277).

It was not until the XX Century, however, that the contemporary notion of transparency took shape and was transposed from abstract doctrine into policy. Since its very beginning, the policies of transparency have developed following two parallel axes that reflected the two broad domains of societal organization: the public and the private. While a thorough discussion of the different kinds of transparency doctrines and policy measures that have been developed within these two macro-categories goes beyond the scope of this paper, it is nevertheless useful to highlight at least some of their most salient characteristics. Doing so will help in identifying the common root from which different strains of transparency have evolved.

Transparency applied to the governance of public affairs – s.c., governmental transparency – typically refers to the citizens' right to access information related to the action of federal, state, or local administrations. It dates back to the mid-1900 when the U.S. government adopted the first laws that would later define the so-called 'Freedom of Information' (FOI) legislation.[5] At the core of governmental transparency is the assumption that democracies are institutionalized communities of self-governing citizens. Following this premise is the consequence that democratic governance is sustainable only as far as citizens are aware of what are the rules that govern society, and dispose of all information necessary to assess whether the action of the public administration is consistent with those laws (Roberts, 2008). In this sense, the blooming of governmental transparency has been seen as a response to the authoritarian distortions that led to World War II, and as a way of futureproofing democracies by making public administrators more visible and accountable to the general public.

However, because "democratically important kinds of information may be information about the activities of private and civic organizations rather than governments themselves" (Fung, 2013, p. 188), many noted also that some degree of transparency may be appropriate for particularly influential private organizations too. More specifically, the conversations around corporate transparency first emerged in the 1920s and '30s as a reaction to the growing power of corporations in strategic industries like oil, banking, finance, etc. One of the first advocates of corporate transparency was judge Louis Brandeis who, in a critique against the opaqueness that characterized business operations of banks and financial intermediaries, famously stated that "sunlight is said to be the best of disinfectants; electric light the most efficient policeman" (Brandeis, 1913, p. 92). In effect, not long after Brandeis' premonitory statement, a series of corporate scandals and disruptive events led to the adoption of

[5] Some of the most notable examples of FOI include the Administrative Procedures Act (1946), the Freedom of Information Act (1966), and the Government in the Sunshine Act (1976).

such as the 1929 Market Stock Crush led the United States to adopt the first corporate transparency measures, including the creation of the Securities Exchange Commission (SEC).

Different from FOI legislation, transparency regulation adopted to tame the power of large corporations was motivated by the necessity to protect consumers from the harmful information asymmetries that unregulated markets can produce (Kosack & Fung, 2014). By enabling higher visibility into business operations, corporate transparency aimed at providing consumers with the necessary information to assess the fairness of the products and services they were offered.

After a few decades, in the 60s and 70s, a new set of concerns populated the public agenda. Thanks to the advocacy efforts of consumers associations and activists around the world, companies faced public scrutiny not only for the fairness of their business practices but also for a new set of emerging concerns: e.g., their environmental impact, the tendency to exploit labor in global supply chains, corruption, etc. (Klein, 2000). In the 1990s, the advent of the internet scaled up the capacity of consumer activists to amplify their demands and coordinate internationally. Ultimately, increased pressure from consumers nudged many of the targeted corporations to implement several transparency measures such as agreeing to codes of conduct, expanding the range of information disclosed in their periodic reports, issuing formal declarations of allegiance to standards and principles in line with the new concerns expressed by the public (Waddock, 2004). Interestingly, many such measures were adopted voluntarily. Instead of waiting for new transparency obligation to be imposed upon them by the government, private companies realized that thanks to the new affordances provided by digital technologies they could handle social pressure proactively, and strategically manage the degree to which they were visible to the public (Fylverbom et al., 2016).

While transparency critics have long debated whether the various instances of government and platform transparency can effectively be read as part of a consistent and organic regulatory framework (Hood, 2006), a theoretical underlying ethos seem to inform the development of the conversation

around transparency: as a doctrine of governance, transparency serves the purpose of exposing the inner functioning of organizations that, due to the power that is either recognized to them by the citizenry or acquired through commercial operations, generate externalities that are consequential for large portions of society, if not for society as a whole. By making an organization visible to the public, transparency systems are "designed to allow people to improve the quality of decisions they make in some way, shape or form, and […] enable them to improve their decisions to reduce the risks they face or to protect their interests" (Fung, 2009).

1.2 – From the promise of radical transparency to contemporary data enclosures: Transparency in the Digital Age

Starting from the late 90s/early 00s, the diffusion of the internet and the development of digital technologies led to a dramatic increase in the production and distribution of information. Foreseeing the internet's potential to transform the socio-economic and cultural fabric of society, a new generation of techno-utopians looked at the internet as the enabler of a new, 'radical' form of transparency: "in the network era, openness wins, [and] central control is lost" (Kelly, 1994, p. 116). Following this line of thought, the proponents of radical transparency argued that thanks to the new forms of openness brought about by digital technologies, grassroots communities around the world could drive economic growth and social progress without the intermediation of the state apparatus (Hammond, 2001).

This libertarian view of a decentralized and ground-up organization of society, in effect, inspired much of the technological innovation coming out of Silicon Valley at the turn of the millennium. Among other initiatives, it also spearheaded the intuition that led to the creation of The Face Book, which was set to become the largest social network in history. As Mark Zuckerberg, one of its co-founders, recalled in an interview:

> "Then one day it kind of hit us that we could play a leading role in making this happen and pushing it forward… And what seemed obvious to my group of friends who were just armchair intellectuals talking about this in college—about how transparency coming from people would transform how the world works and how institutions were governed—it was like, 'Hey, maybe other people aren't actually pushing this, and maybe it takes this group of people who grew up thinking these things and having these values to push it forward" (Kirkpatrick, 2010).

The intellectual blueprint that saw the internet as a herald of radical transparency, however, faded away as companies like Facebook and other emerging digital platforms fully realized the market potential of their business model. Originally seen as a proto-public good by many early Silicon Valley innovators, digital information rapidly became a strategic asset around which centered the whole profitability of the "platform economy" (Kenney & Zysman, 2016). Those who could build the most powerful technology to collect and process data held the key to unlocking potentially unlimited wealth. In fact, the growth recorded by major digital platforms set new record heights.[6] The drawback, however, was a race to seize, store and process as much information as possible; a process that led to the commodification of data. Today, most information existing on the internet is not available to the public; instead, it is stored in servers owned by a handful of private companies. Data turned out to be too marketable for being freely shared with the internet community.

Furthermore, while society has become increasingly transparent to the eyes of contemporary data brokers (Harcourt, 2015), platforms have invested great efforts in sharing as little information as possible about their systems and procedures (Flyverbom, 2016). Nondisclosure agreements,

[6] E.g., Amazon jumped from $8.4 billion of revenue in 2004 to $386 billion in 2020 (BusinessOfApps, 2022); Facebook's monthly active users (MAU) skyrocketed from 100 million in 2008 to 2,9 billion at the end of 2021 (Statista, 2022); Instagram gained 882.5 million MAU from 2013 to 2018 (Our World in Data 2019).

proprietary methods, and a general corporate culture of secrecy have been established as industry standards since the early years of 'big tech.'

Secrecy is of course an important and to a degree necessary dimension of organizational design (Costas & Grey, 2016). In the case of digital platforms, for example, secrecy is tantamount to protecting intellectual property rights while experimenting with product design and innovation (Flyverbom, 2019). At the same time, however, secrecy enables power: "[t]o scrutinize others while avoiding scrutiny oneself is one of the most important forms of power" (Pasquale, 2015). And, as shown by the history of governmental and corporate transparency, the more powerful an organization grows, the more it will have to face demands from civil society and policymakers to be accountable to the rest of society.

1.3 – Transparency and digital platforms

Perhaps surprisingly, the first application of transparency to digital platforms was neither the result of regulation nor a response to particularly intense pressures from civil society: it was a way for platforms to address unwelcomed political pressures that could jeopardize their public image as trusted agents (Suzor, 2018). In fact, the value and potential uses of the information collected by digital platforms soon attracted the attention of several governments around the world. They wanted platforms to provide them with access to information about certain users, or to remove specific content from their websites, at times in overt violation of basic digital human rights (United Nations, 2020).

One of the first investigations on the risks connected to secretive government access to user data resulted in the 2007 report of the Department of Justice's Inspector General on the FBI's abuse of National Security Letters to obtain data on millions of Americans from telecommunication and internet companies (Department of Justice, 2007). This and similar scandals put digital platforms in the uncomfortable situation of potentially losing their users' trust. After some attempts to lobby for

reform of privacy regulation in the US,[7] tech companies decided to adopt a more proactive stance, and started to disclose the details of governments' data requests and orders of content removals in periodical reports. Google opened the way in 2010, publishing the first transparency report in the digital platforms industry, and prompting few other companies to follow its example.

These initiatives were well received by transparency activists, and motivated civil society organizations to push more companies to implement transparency reporting practices. Some of the most influential efforts in this sense have been the Electronic Frontier Foundation's 'Who Has Your Back' annual report and the Ranking Digital Rights' 'Corporate Accountability Index', which scored internet companies based on how transparent and socially responsible they proved to be.

The real turning point that cemented transparency reports as an industry standard, however, was the 2013 scandal sparked by Edward Snowden's revelations on the US National Security Agency's PRISM Program. By the end of 2013, most major US internet companies had published their first transparency report – these included Google, Twitter, Microsoft, LinkedIn, Facebook, Apple, Yahoo, etc. (Bankston, Schulman & Woolery, 2017). The diffusion of transparency reports undoubtedly marked a step forward for users, but the scope of these documents was limited to information on actions that platform had to take in response to external compulsory requests: they included little to no information about how platforms chose to implement their own policies (Parsons, 2019).

Things slowly changed after a new wave of public outcry resulting from yet another series of high-profile scandals. Most notably, the realization that the affordances provided by social media companies had been key in enabling Russia's influence operations during and after the 2016 US presidential elections, and the news that Cambridge Analytica's harvested sensitive data of millions of Facebook users to disseminate targeted political advertising, stroke another major blow to the perceived

[7] E.g.: advocacy efforts by the Global Network Initiative (GNI) and the Digital Due Process Coalition (DDP).

legitimacy of platforms' policies and practices (Howard & Kollanyi, 2016; Woolley, 2016; Howard, 2020).

In particular, when the news of Russia's election meddling broke in 2017, it refocused much of the transparency conversation on the vulnerabilities embedded in digital platforms' advertising and content moderation systems. As the notions of 'algorithmic transparency' and 'advertising transparency' became common jargon, platforms faced mounting pressures to disclose information about the principles informing the design of their automated decision systems and products (Mittelstadt, 2016; Crain & Nadler, 2019). As a result, by the end of 2018, Google, Facebook, and Twitter expanded their transparency reports to include information on the enforcement of their terms of use (Crain & Nadler, 2019), provided public access to searchable libraries containing data on political advertising (Leersen et al., 2019) and even granted independent researchers a deeper degree of access to some other categories of data through Application Programing Interfaces (APIs) – i.e., tools designed by platforms for delivering data directly to third parties based on formal queries. They are versatile programs that offer platforms some control over what types of and how much data can be queried, while enhancing the ability of third parties to rapidly obtain large bulks of data. Though originally designed to provide data for app developers and other commercial partners of the platforms, APIs have been essential enablers of the research conducted by of data journalists, watchdogs and academics.

When the Cambridge Analytica scandal broke later in 2018, however, it had an unforeseen regressive impact on transparency standards. Among the measures adopted to limit the risk of further data breaches, in fact, platforms revised their data accessibility protocols for third parties by restricting access to or retiring several of their APIs (Facebook, 2018; Twitter, 2019). While the initiative of locking up access to platform data was presented as a safety-net for the users, its most direct consequence was a drastic limitation of the ability of independent researchers to carry out data-driven

investigations. This represented a major setback in terms of public accountability of digital platforms as independent research was the only available instrument to cross-check the data published by platforms in their voluntary reports.

Furthermore, not only APIs shut down, but platforms adopted a much stricter approach to web scraping practices. Scraping is somewhat complementary to querying APIs, and can be described as the automated extraction of data from the source code of public web pages (Cooley *et al.*, 1997). Compared to API queries, web scraping is technically more complex but also more powerful, as it allows the extraction of any data that exists on a web page regardless of the existence or limitations of an API. Most digital platforms, however, have sought to legally persecute web scrapers, banning the practice in their terms of service, and suing individuals caught in the process (Freelon, 2018). While web scraping is in effect a technique that can give access to sensitive data and, if employed unethically, can lead to violations of user privacy, it remains a fundamental instrument at the service of transparency and accountability. Without web scraping, initiatives like the award-winning 'COVID Tracking Project,' Reveal's investigation of the overlaps in membership of law enforcement and extreme politics Facebook groups (Carless, 2019), The Markup's finding that Google's search results consistently favors its own products (Jeffries & Yin, 2020), or Reuters' discovery of an illicit market for adopted kids that led to two convictions for kidnapping (Twohey, 2013), would not have been possible. And the same is true for academia, where web scraping has enabled field defining findings on issues ranging from Chinese censorship (King et al., 2013) to the localization of armed conflict (Raleigh et al., 2010), from the psychology of personality (Youyou et al., 2015) and learning (Baker & Yacef, 2009), to the diffusion of opioid medications (Moeller & Svensson, 2021). Despite encouraging signals from the public administrations that things may chance in the future,[8] as of today the legitimacy

[8] On Thursday, May 19, 2022 the Department of Justice announced the revision of its policy on the enforcement of the Computer Fraud and Abuse Act based on which "good-faith security research should not be charged. Good faith security research means accessing a computer solely for purposes of good-faith testing, investigation, and/or

of web scraping remains sanctionable under the Computer Fraud and Abuse Act. As such, providing a legal safety-net for ethical scrapers while ensuring adequate privacy protection remains a challenge that needs to be addressed by future transparency legislation.

Most recently, a new political storm hit digital platforms. At the end of 2021, former Facebook employee and whistleblower Frances Haugen handed over to various media outlets a throve of nearly ten thousands internal Facebook documents. In what was arguably the most famous case of whistleblowing on digital platforms, the 'Facebook files' played out as a stark testimony of how little of what platforms do is visible to society and, at the same time, of how pervasive and dangerous their operations can be when left unchecked. In particular, the documents touched on issues such as Facebook's internal research showing that some of its products can cause psychological harm to kids; its perseverance in implementing products like 'downstream MSI' in Myanmar, despite evidence that they contributed to spreading hate speech and violence; the company's management systematic disregard for concerns raised by employees about how the platform is used in developing countries to promote, among other things, conflict, organ selling and human trafficking.[9]

The Facebook files effectively re-ignited widespread anxieties on the societal impact of digital platforms, and conferred a new sense of urgency to the debate around transparency regulation. For the first time, policymakers in the US and in Europe seemed united in arguing that the disregard for sufficient standards of platform transparency could not be left unsanctioned.

correction of a security flaw or vulnerability, where such activity is carried out in a manner designed to avoid any harm to individuals or the public, and where the information derived from the activity is used primarily to promote the security or safety of the class of devices, machines, or online services to which the accessed computer belongs, or those who use such devices, machines, or online services."

[9] For more details and sources refer to 'The Facebook Files: A Wall Street Journal Investigation', accessed 07/02/2022, https://www.wsj.com/articles/the-facebook-files-11631713039.

In the U.S., over the last two years at least nine bills that provide specific transparency obligations for digital platforms have been introduced.[10] While it is difficult to foresee whether any of these proposals will ever be passed into law, the fact that so many transparency measures are being discussed by policymakers shows how lively this policy space is. At the same time, however, the proliferation of different legislative proposals on platform transparency signals the lack of a shared understanding among policymakers about what exactly should be done to make digital platforms more transparent. The measures proposed in the aforementioned bills, in fact, run the gamut from basic transparency requirements related to the terms of service enforced by platforms, to supervised access to sensitive user data for verified researchers, covering a lot of ground in between included additional advertising transparency, independent audits of platforms' recommendation algorithms, 'legal shields' for researchers doing public interest research, and so on. In effect, each bill proposes a specific definition of transparency. Some proposals are more comprehensive, while other are narrow and issue-specific; they prescribe a plethora of different transparency procedures; and aim at addressing various problems of the platform business.

Taken together, the platform transparency measures proposed by these bills reflect a fragmented landscape. This is in part the result of different ways of conceptualizing transparency as part of diverse political agendas, and in part the consequence of trying to operationalize a complex and still relatively new policy area. In any case, the more rarefied the debate, the less likely it is to resolve in concrete action.

[10] These are: the Algorithmic Justice and Online Platform Transparency Act (2021); the Filter Bubble Transparency Act (2021); the Social Media Data Act (2021); the Platform Accountability and Consumer Transparency Act (2021); the Platform Accountability and Transparency Act (2021); the Algorithmic Accountability Act (2022); the Digital Services Oversight and Safety Act (2022); the Digital Platform Commission Act (2022); and the Social Media NUDGE Act (2022).

Part 2 – Defining features of the emerging concept of platform transparency: Discourse analysis and the extraction of semantic patterns

As shown by the brief genealogy traced in Part 1, platform transparency is a heterogeneous and dynamic concept. Born as techno-utopian critique against the concentration of power in the hands of traditional political and financial elites, in less than a decade it has evolved into a policy space aimed at regulating the uncontrolled ascent of the "new governors" (Klonick, 2017) of the internet: digital platforms.

Though the conversation around platform transparency is increasingly gaining focus and sophistication, it still lacks the structure necessary for thinking systematically about how transparency and digital platform interrelate. In public discourse, 'platform transparency' is often employed in a rather vague way: as noted by Tarleton Gillespie, for instance, "calls for greater transparency in the critique of social media are so common as to be nearly vacant" (Gillespie, 2018: 212).

Ambiguity has been a distinguishing trait of transparency since it first entered the policy agenda – and some have argued that such semantic fluidity was in fact among the primary reason of its success (Hood, 2013); yet, the tendency to employ this concept indiscriminately, or in an ill-specified fashion has only accrued in the case of platform transparency. Not only it is still a relatively new concept, thus still in the process of being properly specified, but the very limited general knowledge of how platforms operate, and the large range of political interests that transparency is conducive to accommodate, have led to a situation in which calls for greater transparency in the platform economy risk being made more to win the favor of a certain audiences than to signal concrete efforts towards a clear objective.

With the aim of contributing some clarity to the field, the following paragraphs offer a tentative model for systematizing the emerging concept of platform transparency. In particular, drawing from Nelsons'

computational grounded theory methodological framwork (Nelson, 2020), Par. 2.1 and Par. 2.2 respectively discuss how digital platforms and policymakers in the U.S. understand transparency.

2.1 – Transparency in the public discourse of major digital platforms

This section illustrates the results of a computational text analysis performed on transparency-related newsroom posts by Meta (Facebook's parent company), Google and Twitter.

While a comprehensive overview of platforms' public discourse on transparency should ideally include a larger cohort of actors, Meta, Google, and Twitter have been the most active and outspoken ones in the field of transparency policies. Furthermore, most major leaps forward in platform transparency have been the consequence of public concern for events that directly involved one or more of these three companies. And, provided that policymakers have mostly discussed transparency in relation to such events, it follows that even in public policy the conversation around platform transparency has been largely driven by issues related to the specific practices of Google, Twitter, and Meta. For these reasons, and considering the limited scope of this paper, I chose to focus on these three specific platforms so as to gather a sample that is as representative as possible of my object of analysis.

Before moving further, it shall also be noted that, as these posts are directed to civil society and policymakers – i.e. to individuals external to the organizations – they reflect a notion of transparency filtered through corporate communication strategies. As demonstrated by recent transparency studies, "Transparency [can] take the shape of the purposeful and strategic production of insight and openness in attempts to position organizations as attractive and willing to engage with their employees and stakeholders", and so can communications around it (Flyverbom, 2016: 114). Provided that the scope of this paper is to map how the public discourse around platform transparency has developed, this makes for an adequate data source.

2.1.1 Data Collection and Preprocessing

The data collected includes all transparency-related posts listed on the newsrooms of Meta, Google, and Twitter as of the 15th of July 2022. The total number of documents collected equals to 539, of which 239 have been published by Google, 227 by Meta, and 73 by Twitter.[11]

Each document was identified through a keyword search on the platforms' websites, scraped, and selected as part of the final corpus only if it included at least two explicit references to 'transparency' – newsroom posts are generally short-form and issue-specific; as such, a minimum threshold of two explicit references to transparency has been considered sufficient to ensure the relevance of posts for the analysis.

As is common practice in 'text as data' methods, words considered unlikely to contribute substantive meaning to the analysis have been filtered out from the corpus: these include 'stop' and 'function' words, and words that we would expect to encounter often in the documents at hand like the words 'Facebook,' 'Twitter,' and 'Google' (Grimmer & Stewart, 2013).

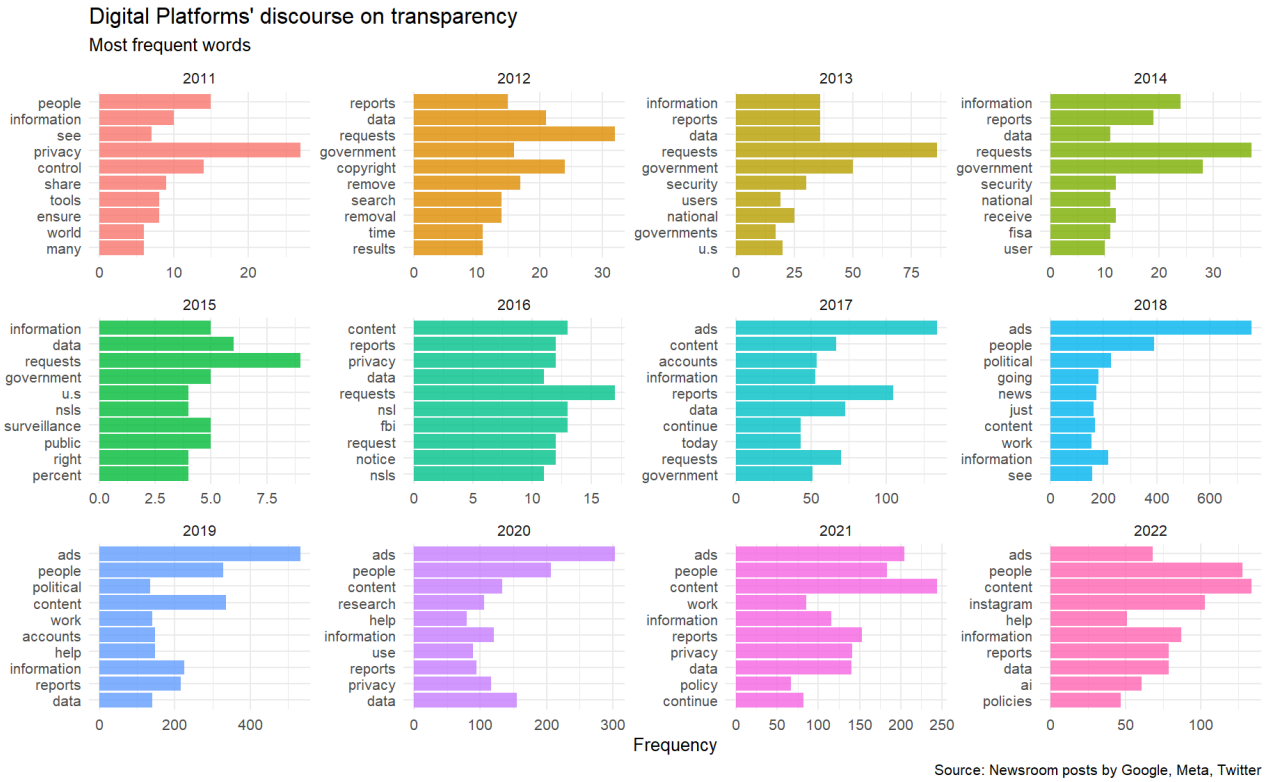
2.1.2 Discussion

To gain a better understanding of how platforms' discourse on transparency has evolved, I first employed a basic lexical selection technique to identify the words used most often in transparency-related communications. By grouping the results by year, it is also possible to gain a high-level view of how platform discourse on transparency has evolved in time.

Figure 1

[11] At the time of collection, Twitter only made available documents published from 2016 to 2022.

As shown by *Figure 1*, platforms first started to discuss transparency in relation to growing concerns for the privacy rights of users. At that time, transparency was often juxtaposed to how much detail a platform like Facebook would provide to its user about how their data were handled. Essentially, at first transparency simply referred to digital platforms' being explicit with users on their data collection and processing procedures.



Already by 2012, however, it is possible to see how platforms framed transparency in relation to a concrete procedures. In particular, the topic of transparency reports starts to gain traction in the conversation of digital platforms. Reports remained at the very core of the transparency discourse for some time, as their prominence only accrued in the aftermath of the Snowden revelations. At that time, platforms used to frame transparency as a symbol of their efforts to keep users as informed as possible on the degree to which public authorities disposed of their information.

Only after the 2016 presidential elections, once it became apparent that some of the affordances provided by social media enabled politically-motivated actors to manipulate the digital information

space, new topics such as advertising and content moderation entered the public debate. The focus on advertising and content was later reinforced by the Cambridge Analytica scandal, and has remained critical to the debate ever since. Once again, platforms justified their transparency efforts as steps taken in order to better safeguard the interests of their users. For example, as can be read in a 2020 post by Google,

> "our policies are designed to promote transparency for our users and help protect against misinformation. This work is critically important when it comes to safeguarding elections as people use apps to research candidates, register to vote, or find a polling place. As mobile apps disseminate voting information and increasingly support voting activity, we're ensuring safety and transparency for app users." (Stewart, 2020)

2020 was also the year in which the U.S. held a new presidential elections cycle, and platforms were particularly wary of the risk that they could once again provide the stage for tentative manipulations of the public opinion. In addition to increasing transparency on advertisement and content moderation for the benefit of users, the issue of enabling independent research on platform data also received substantial attention. In effect, even if 2020 was the first year in which 'research' showed up among the most frequent words used by platforms, a more detailed inspection of the data sample shows that transparency started to be framed in terms of data disclosures for the purposes of third party research in already in 2018. That is, when all three platforms subject to this study first published data with the explicit purpose of providing useful information to independent researchers.[12]

[12] Evidence from the sample of documents analyzed include: Google announcing the disclosure of advertising data in a searchable library "to make this information as accessible and useful as possible to citizens, practitioners, and researchers" (Junius, 2018); Twitter's announcement that "we are releasing all the accounts and related content associated with potential information operations that we have found on our service since 2016. We had previously disclosed these activities, but are now releasing substantially more information about them to enable independent academic research and investigation." (Gadde & Roth, 2016); and Facebook's announcement that 'This summer, we'll launch a public archive showing all ads that ran with a political label. […]So researchers, journalists, watchdog organizations, or individuals who are just curious will be able to see all of these ads in one place. This will offer an unmatched view of paid political messages on the platform. We recognize this is a place to start and will work with outside experts to make it better." (Rosen, 2016).

In order to validate the trends highlighted by means of word frequency analysis, and to verify whether major differences exist between the transparency discourses of different platforms, I performed a second lexical feature selection based on the Term Frequency-Inverse Document Frquency (tf-idf) metric. Substantively, tf-idf is a formula that describes the distinctiveness of a word based on how frequently that word appears in a document or group of documents adjusted for how frequently it appears in the rest of the corpus (Silge and Robinson, 2016: 31). In other words, if a term has a strong presence in a document and is rarely found in other documents, that term will receive a high tf-idf score.

The results (see *Figure 2*) are consistent with the previous findings inasmuch as the words with the highest tf-idf score reflect the same identified features like privacy, reports, and advertising as most distinctive of digital platforms' transparency discourse. Furthermore, although minor differences exist, the results are consistent across all three platforms, signaling coherence in how the concept of platform transparency has developed in the industry.

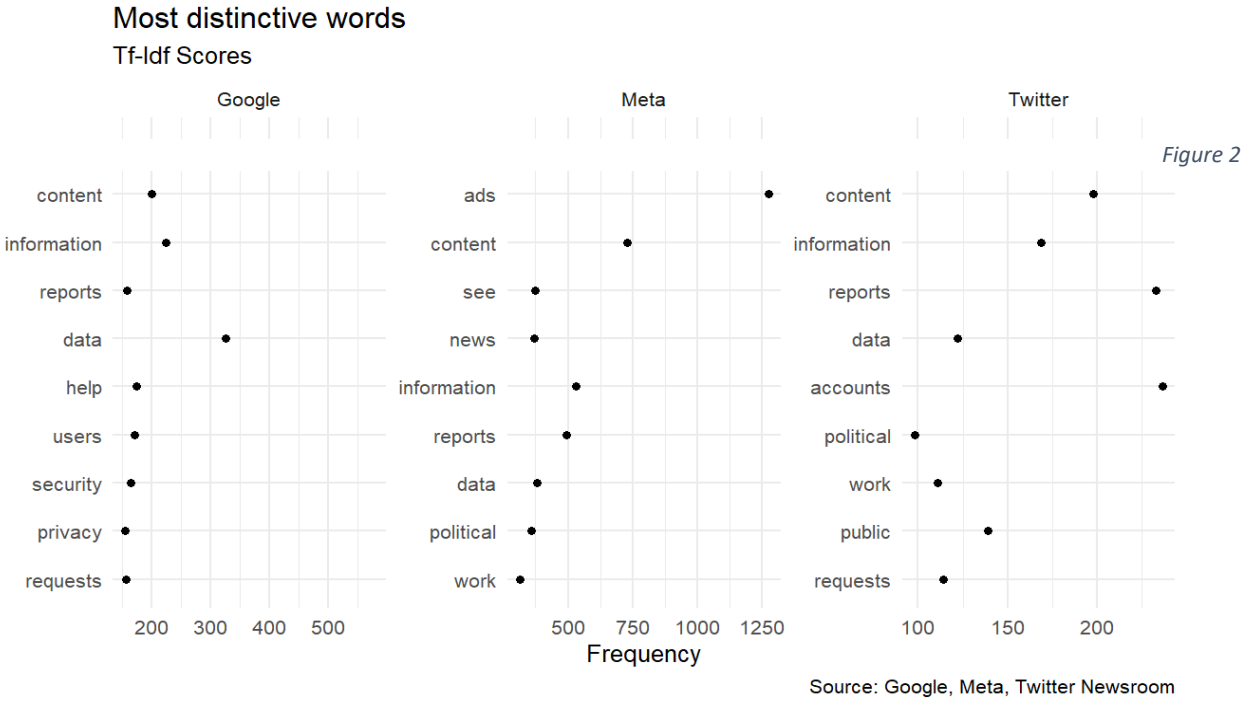


Figure 2

Finally, in the third part of the analysis, I resorted to yet another technique to identify the most evident substantive patterns patterns in the data. The results of the structural topic model are reported in *Figure 3*, and largely confirm the inference based on lexical selection methods. In particular, three main topics have been identified

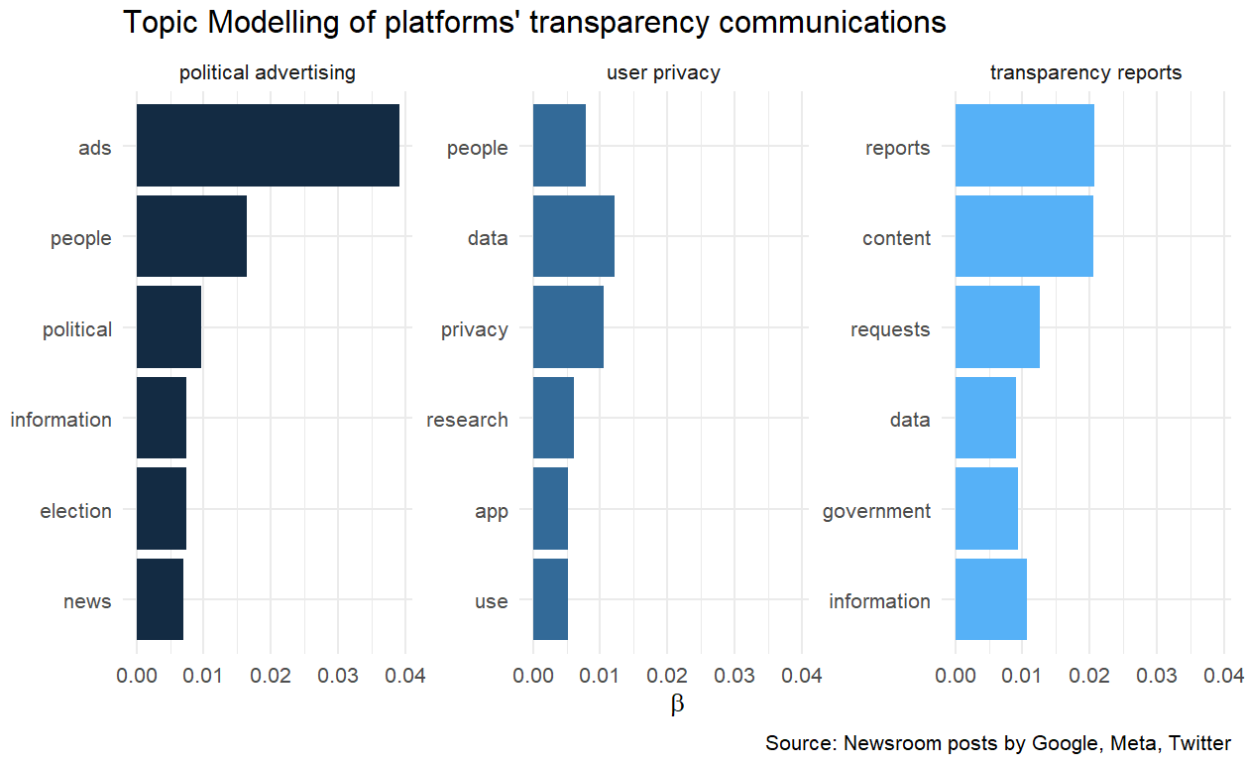


Figure 3

2.2 – Transparency in the public discourse of the U.S. Congres

This section replicates the same analytical steps covered in Section 2.1 on textual data representing transcripts of U.S. Congress Committee sessions. The aim of this section is to trace the development of platform transparency in the institutional discourse of U.S. policymakers, and reflect on the differences and similarities between how policymakers and platforms have shaped the formation of

such concept. To avoid redundancy with the preceding Section, the discussion will be limited to elements of novelty.

2.2.1 Data Collection and Preprocessing

This analysis relies on a corpus obtained by scraping transcripts of U.S. Congress Committee meetings from the Congress' official website. More precisely, a total of 1512 documents have been collected by programmatically scraping all transcripts that mentioned both 'transparency' and 'digital platforms' available as of May 31st of 2022.

Differently from the newsroom posts analyzed in Section 2.1, Congress sessions typically cover a variety of topics. For instance, the words 'transparency' and 'digital platforms' could well be employed in two wholly distinct contexts within the same document – also considering that references to 'transparency' in U.S. politics are particularly frequent in several policy areas – and their mere presence in a session's transcript is no guarantee of its relevance to the present analysis. For this reason, provided that this Section aims at understanding how policymakers have discussed transparency relative to platform governance, the corpus had to be further refined by selecting only the paragraphs of interests. Some of the major cleaning steps followed to eliminate most of the noise from the data included: filtering out all documents that mention transparency and social media less than five times; selecting only the transcripts of committees competent to discuss topics relevant to social media regulation; random sampling and reading of the remaining documents to verify if they are relevant to the field of interest; selection of the specific paragraphs mentioning 'transparency.' As a result of this drastic but necessary cleaning, the corpus' dimension shrunk from 1512 to 32 committee sessions, for a total of 23780 tokens and 4890 features.

Further preprocessing steps have been performed that mirror those already mentioned in Section 2.1.1.

2.2.2 Discussion

Similarly to what we saw in the case of digital platforms, the U.S. Congress first discussed the theme of platform transparency in relation to user privacy (see *Figure 4*): for the first time, in 2011 U.S. legislators heavily criticized digital platforms for not adequately informing users on how their data was treated after collection, and for inducing them to accept vague and obscure terms of use that granted broad discretion to platforms on privacy matters.

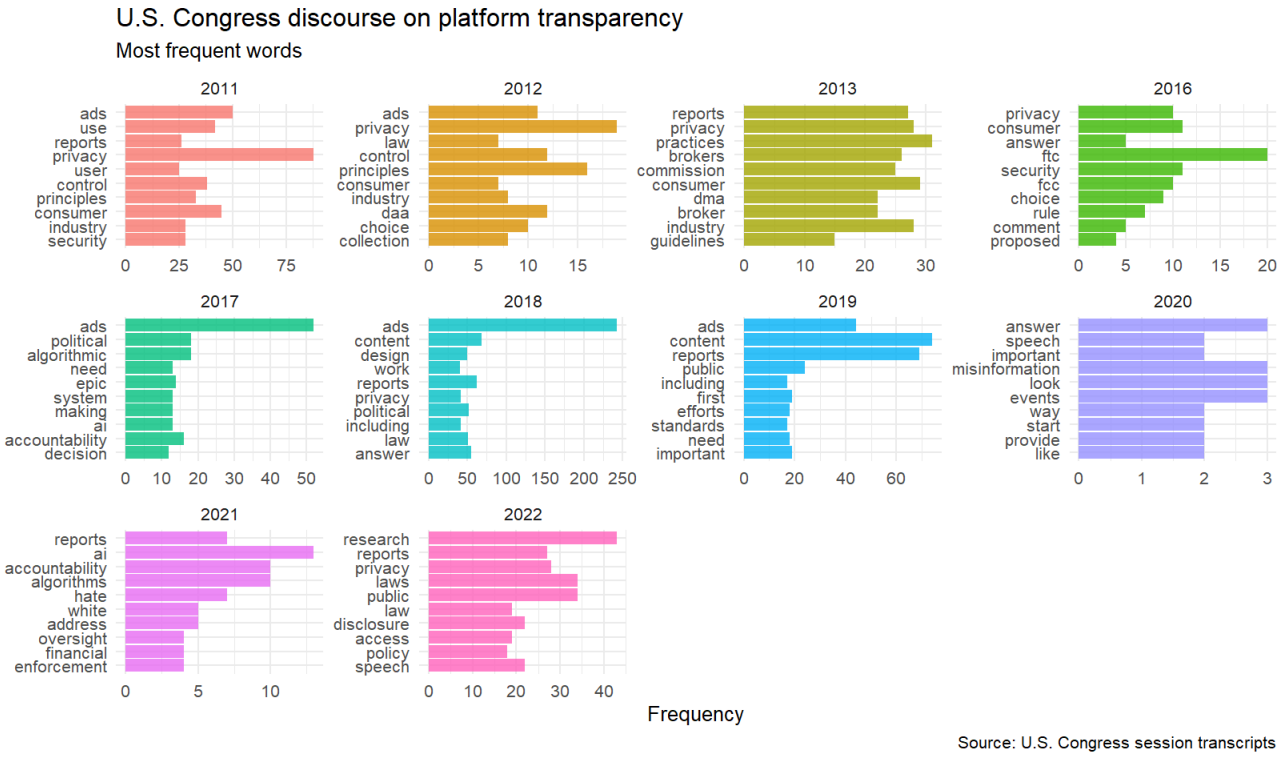


Figure 4

A turning point for the entrance of platform transparency in the public policy agenda was a 2011 Senate Hearing titled 'The State of Online Consumer Privacy' addressing the challenge of balancing a free internet with concerns for the practices of online information collection.[13] Interestingly, already in 2011 Congress predated digital platforms by linking transparency not only to more explicit and

[13] Concerns sparked by the Federal Trade Commission charges against Facebook for "deceiv[ing] consumers by telling them they could keep their information on Facebook private, and then repeatedly allowing it to be shared and made public" (U.S. Federal Trade Commission, 2011), and against Google for "Deceptive privacy practices in [the] rollout of its Buzz social network" (U.S. Federal Trade Commission, 2011[2])

succinct terms of use, but also to providing users with more information on platforms' advertising practices. At this time, however, advertising transparency had not yet been defined as the practice of disclosing information about the actors and transactions behind the adds themselves. When juxtaposed, the concepts of transparency and advertising were used to discuss the opportunity that platforms put users in the condition to understand when they were tracked and, eventually, to opt-out from behavioral advertising.

After 2012, due to the clamor of the Snowden revelations, the issue of data privacy took over the conversation, dominating it until after the 2016 elections. Once again, as it was the case for Google, Twitter and Facebook, when the news of Russia's election meddling and, later on, of the Cambridge Analytica data breach broke, policymakers' conceptual framing of platform transparency evolved so as to systematically encompass reporting on content moderation and political advertising practices. Interestingly, starting in 2017, Congress also debated transparency as it applies to algorithms – a conversation that platforms have been more reluctant to have.

While 2020 has been relatively quite as far as platform transparency is concerned – probably also due to the concentration of most political focus on managing the Covid-19 pandemic – the topic of algorithms and artificial intelligence re-emerged in 2021, the year in which Congress hosted the field-defining testimony of Facebook whistleblower Frances Haugen. One of the cornerstones of Haugen's recommendations for reducing the harm that social media can inflict to open societies was precisely

to subject them to more public scrutiny and, in particular, to independent research carried out on platform data.

Ever since, research has been at the center of Congress' discourse around social media transparency. A trend which is also confirmed by the bills on platform transparency introduced between the end of 2021 and the first half of 2022.

Interestingly, as Figure 5 shows, the 'term frequency-inverse document frequency' score strongly underlines how transparency has been a distinctive focus in specific Congress sessions. Considering that, once again, tf-idf scores measure the frequency with which a term is used in a document relative to the rarity with which it is used in the rest of the corpus, one way of interpreting this result would be to note that there must have been few Congress sessions that explore the topic of research in platform transparency in great detail. Probably the most striking example in this sense is the

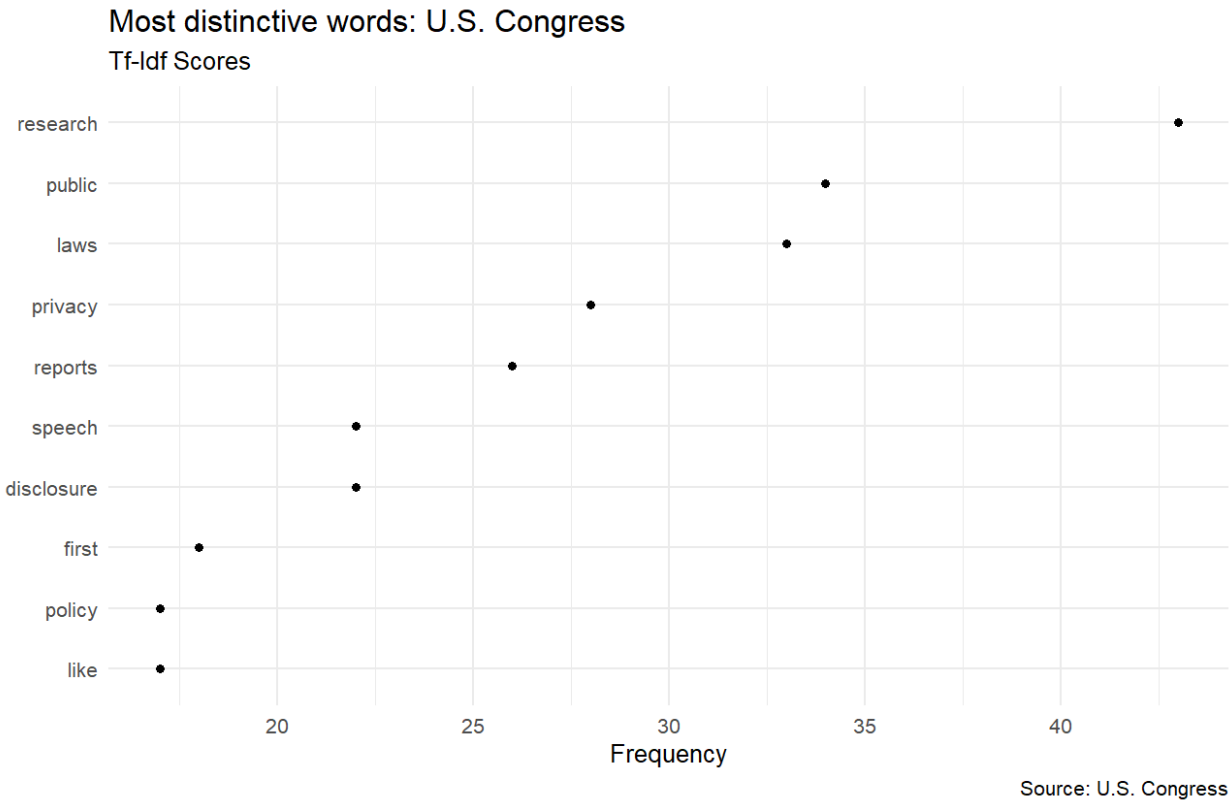


Figure 5

05/04/2022 Senate 'Hearing to examine platform transparency, focusing on understanding the impact

of social media.' In that occasion, testimonies mentioned the words 'research,' 'researcher,' or 'researchers' 198 times.

Finally, the structural topic modelling of U.S. Congress sessions on transparency and digital platforms returned very similar results to the those already described at the end of Section 2.1.2 (see *Figure 6*). It should be kept in mind that, especially with smaller datasets like the one used in this analysis, topic modelling algorithms can only detect the most evident patterns in the data. For this reason, in this case the model does not detect topics related to research or algorithmic transparency, as these are represented much less in the data compared to major features like privacy, advertising and reporting. Nevertheless, for the purposes of this analysis, topic modelling is a valuable instrument to validate the conclusions provided by the two lexical selection analyses.

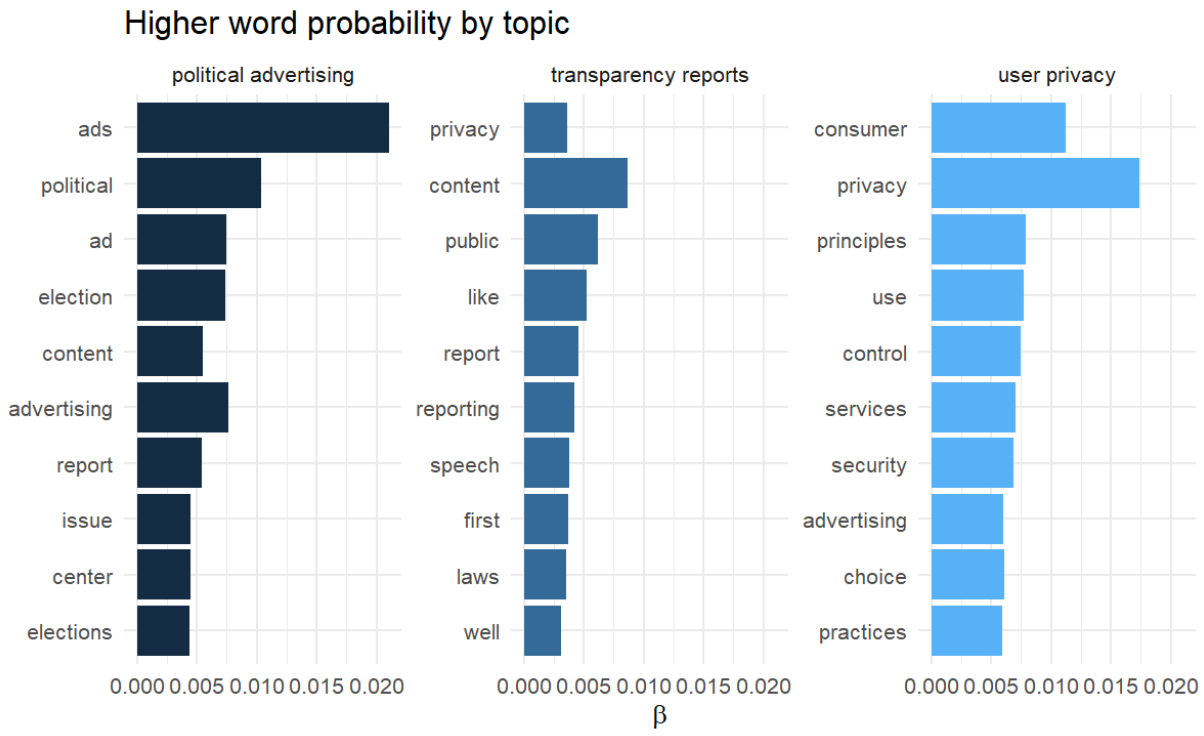


Figure 6

PART 3 – Platform transparency: A conceptual framework

Based on the results of the analysis described in Part 2, the emerging concept of platform transparency can be understood as a composite and stratified entity. Though still in progress, the debate that is forging the concept of platform transparency has evolved following identifiable and consistent patterns. These, in turn, suggest that platform transparency could be understood as a conceptual category, thus encompassing lower-level conceptual types.

A clear understanding of what these types are and of how they relate to each other is a necessary, but not yet fulfilled condition for the debate to unleash its regulatory potential. As already showed by the literature on corporate transparency, in fact, making sure transparency is effective in producing accountability, knowledge and more socially responsible behaviors is less straightforward than it seems. In order for transparency to be useful, it needs to be targeted to specific information and to the actors that can benefit from it: transparency measures are "effective only when they provid[e] facts that people wanted in times, places, and ways that enabled them to act. (Fung, Graham & Weil, 2007). Considering the above, and based on the conviction that a focused debate is a precondition to generate focused regulatory measures, this final section aims at synthesizing the findings of Part 2 to develop a hermeneutical tool capable of contributing some clarity and structure to the debate around platform transparency. More precisely, Part 3 proposed a formal typology of platform transparency.

3.1 – Key dimensions and types of platform transparency

Studying the discourse of digital platforms and U.S. policymakers around platform transparency revealed the centrality the topics of privacy, advertising and content moderation. Furthermore, the analysis also detected a relatively newer set of elements contributing to the formation of the concept of platforms transparency. Among these, independent research and algorithmic transparency emerged

as the most prominent. To summarize, the debate around platform transparency has primarily been centered around:

- measures intended to inform users about potentially harmful government and law enforcement practices (first generation of transparency reports).
- measures intended to clarify how specific platform policies are implemented (second generation of transparency reports).
- measures designed to provide information about the actors involved in political advertising (digital ad transparency).
- measure designed to enable independent research conducted by academics, journalists, factcheckers, and other analysts (providing different types of APIs; establishing closed collaborations between platforms and selected researchers).
- measured intended to gain a better understanding of the design and effects of recommendation algorithms (algorithmic transparency).

Each of these thematic clusters, then, has its own theme-specific debate that encompasses a plethora of challenging questions reflecting opposed interests and ideologies. For instance: should platforms limit disclosures to strictly political advertisement, or should they open up to "universal digital ads transparency" (Edelson et al., 2021)? Are sanctions for mishandling data pursuant to current privacy legislation the primary reason behind the conservative data sharing approach held by platforms, or should these invest more resources to provide privacy-safe disclosures? Should platforms report only on the downstream effects of their content moderation practices, or should they provide access also to the systems and procedures that regulate content moderation? Should researchers be allowed to scrape data from platforms? Can algorithmic transparency create greater value for society than the harm it causes to digital platforms?

Undoubtedly, platform transparency accounts for a vast conceptual space. Despite its heterogeneity, however, the findings presented in Part 2 suggest that each and every topic-specific issue in effect shares the same fundamental concerns: problems of platform transparency can ultimately be reduced to a combination of questions of what and how information is made transparent:

a) what is the quality of the information shared? More specifically, is it potentially violating of the privacy of the individuals to which it refers?
b) how is that information shared? Is there a formal or supervised procedure in place, or can individuals access to certain information independently?

These two sets of concerns effectively provide for the foundational structure underpinning the conceptual space of platform transparency. By combining them two questions in a matrix, it is possible to develop a bi-dimensional typology of the platform transparency discourse as it has evolved until now (see *Table 1*).

PLATFORM TRANSPARENCY

	Non-sensitive data	**Sensitive data**
Controlled procedure	General purpose transparency	Supervised transparency
Independent access	Open-source transparency	Harmful transparency

In a conceptual typology, individual cells are bound to the specified concept by a relationship of hierarchy in kind. As noted by Collier *et al.*,

> "Understanding this hierarchy helps to answer the following question: What establishes the meaning of the cell types, that is, of the concept that corresponds to each cell? The answer is twofold. (1) Each cell type is indeed "a kind of" in relation to the overarching concept around which the typology is organized, and (2) the categories that establish the row and column variables provide the core defining attributes of the cell type." (Collier *et al*, 2012: 222)

Accordingly, this exercise in abstraction has made it possible to identify four kinds of platform transparency.

- general purpose transparency, which includes all transparency measures that refer to non-sensitive data and are implemented through controlled procedures. This, at least in theory, should be the least costly and safest type of platform transparency, and includes: information about how platforms implement their content moderation policies; fully transparency terms of use; aggregate level data about the content hosted by platforms (e.g. Facebook's reports of most viewed content); selected data on advertising transparency; information about governments data requests, content removals pursuant to copyright infringements, content removals pursuant to orders from law enforcement agencies and governments; selected data on how recommendation algorithms work;
- open source transparency refers to the ability of individuals and organizations to retrieve non-sensitive data with relatively little limitations – in an ideal world, the only procedural limitations here would be those necessary to prevent access to sensitive data. Examples of open source transparency include APIs, other data access tools like CrowdTangle, publications of dataset by platforms, and ethical web scraping;

- supervised transparency responds to the necessity of analyzing also sensitive data in order to answer specific questions. This type of transparency is much more costly and difficult to implement. It can be realized through a variety of methods. Some examples include using techniques to introduce noise in the data like. 'differential privacy,' or sharing data with selected individuals and subjecting the use of such data to procedural limits, including if necessary third party supervision, to limit the risk of privacy violations. Mention Social Science One and PATA.
- Finally, transparency can be harmful when access to sensitive information happens in a non-supervised way. Potentially harmful transparency can result, for example, from abusive use of APIs and web scraping, hacking, abuse of government or law enforcement authority, etc. 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.

Generally speaking, defining a conceptual structure is useful because it helps with organizing thinking when concepts are elusive, and limits the risk of stretching the boundaries of those concept. In the context of this analysis, then, using a formal typology to segment the conversation around platform transparency is particularly important considering that, as mentioned earlier, transparency has been proved to lead to accountability only when targeted to the right audience (Fung *et al.*, 2008). In this respect, identifying different types of transparency helps with focusing the conversation around questions such as who are the subject that can be granted access to sensitive data? Who can benefit from open source transparency and how can it be designed in order to contribute as much value as possible to those subjects? What safety measures should be implemented in order to reduce the risk

that dangerous actors access sensitive data? To who does general purpose transparency speak exactly, and what information are these subjects really interested in?

Conclusion

Transparency provides for a promising policy framework to approach some of the complexities inherent to the governance of digital platforms. First, it can enhance the public understanding of how digital platforms operate, how the design of their products affects the ways in which users behave, what data is collected and how it is used, etc. In short, transparency can disclose information necessary to answer questions that are tantamount to calibrate any sensible framework of platform regulation. Second, one of the assumptions that has accompanied debates about all forms of transparency is that making an organization more visible to the public will provide incentives for that organization to act more responsibly. As far as digital platforms are concerned, scaling up their accountability when they lay out business strategies and objectives would have a major impact for many of the concerns flagged by people who are studying this space (e.g., misinformation, polarization, mental health, discrimination, etc.).

Developing effective transparency policies, however, is not at all straightforward. History is riddled with examples of useless, if not harmful transparency initiatives in both the public and the private sector. The mere disclosure of information unaccompanied by a clear idea of who will benefit from that information and of how they will use it to foster change is usually doomed to failure. Despite so, transparency measures that are not sufficiently targeted are as common as they are idle. In fact, transparency is by its very nature a vague concept, prone to be interpreted in a plethora of different ways. Furthermore, it speaks directly to some of the core values of democratic societies: openness, public accountability, social responsibility. As such, it is an easy talking point for politicians and corporations that aim at harvesting popular consensus. The reality of applied transparency, however,

is a different story. Well-crafted transparency policies are highly technical and requires answering difficult questions on how to balance equally important but conflicting interests. For example, any serious legislative effort to design targeted transparency measures cannot prescind from also thinking about how transparency interact with privacy: how should we draw the line between sensitive and non-sensitive data with the goal making platforms more transparent without violating the rights of their users?

Against this backdrop, this paper proposes a possible framework to make sense of the composite nature of platform transparency and simplify the complexity of the relative conversation. By identifying the core dimensions of the debate, it shows that it is possible to locate different instances of platform transparency in an ordered and consistent space. This, in turn, encourages thinking about a specific instance of platform transparency as a combination of different sets of needs and interests, ultimately opening the way for the design of more targeted transparency.

Bibliography

Albu O.B. & M. Flyverbom (2016). 'Organizational Transparency: Conceptualizations, Conditions, and Consequences.' *Business & Society, 58(2): 268–297*.

Aral, S. (2020). *The Hype Machine: How social media disrupts our elections, our economy, and our health – and how we must adapt*. New York: Penguin Random House.

Balkin, J. M. (2020). 'How to regulate (and not regulate) social media.' Knight First Amendment Institute, Occasional Papers, accessed 07/15/2022, https://knightcolumbia.org/content/how-to-regulate-and-not-regulate-social-media.

Bankston, K., R. Schulman & L. Woolery (2017). 'Getting Internet Companies To Do The Right Thing – Case Study #3: Transparency Reporting.' *New America*, accessed 06/08/2022, https://www.newamerica.org/in-depth/getting-internet-companies-do-right-thing/case-study-3-transparency-reporting/.

Barnett, M. & R. Duval (2005). 'Power in International Politics.' *International Organization*, 59(1): 39–75.

Benkler, Y., R. Faris & H. Roberts (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press.

Bentham, J. (1790s). *Writing on the Poor Laws, vol. 1*. Ed. by M. Quinn (2001). Oxford: Clarendon.

Black, J. (1997). 'Transparent Policy Measures.' *Oxford Dictionary of Economics*. Oxford: Oxford University Press.

Black, J. (2001). 'Decentring regulation: Understanding the role of regulation and self-regulation in a "post-regulatory" world.' *Current Legal Problems*, 54(1): 103–146.

Brandeis, L. (1913). *Other Peoples' Money and How Bankers Use It*. New York: F.A. Stokes.

Carless, W. (2019). 'Hundreds of cops are in extremist Facebook groups. Why haven't their departments done anything about it?.' *Reveal*, accessed 07/05/2022, https://revealnews.org/article/hundreds-of-cops-are-in-extremist-facebook-groups-why-havent-their-departments-done-anything-about-it/.

Castells, M. (1996). *The Information age: Economy, Society and Culture. Vol. 1: The Network Society*. Chichester: Wiley-Blackwell

Coglianese, C. & E. Mendelson (2010). 'Meta-Regulation and Self-regulation." In M. Cave, R. Baldwin & M. Lodge (*eds.*). *The Oxford Handbook on Regulation*. Oxford: Oxford University Press.

Collier, D., J. LaPorte & J. Seawright (2012). 'Putting Typologies to Work: Concept Formation, Measurement, and Analytic Rigor.' *Political Research Quarterly*, 65(1): 217–232.

Cooley, R., B. Mobasher & J. Srivastava (1997). 'Web mining: Information and pattern discovery on the World Wide Web.' *Proceedings of the Ninth IEEE International Conference on Tools With Artificial Intelligence*: 558–567.

Costas, J. & C. Grey (2016). *Secrecy at work: The hidden architecture of organizations*. Stanford: Stanford University Press.

Crain, M. & A. Nadler (2019). 'Political Manipulation and Internet Advertising Infrastructure.' *Journal of Information Policy*, 9: 370–410.

Cucciniello, M., G.A. Porumbescu & S. Grimmelikhuijsen (2017). '25 years of transparency research: Evidence and future research.' *Public Administration Review*, 77(1): 32–44.

Cusumano, M. A., A. Gawer & D. B. Yoffie (2021). 'Can self-regulation save digital platforms?.' *Industrial and Corporate Change*, 30(5): 1259–1285.

Edelson, L., J. Chuang, E. F. Fowler, et al. (2021). 'A Standard for Universal Digital Ad Transparency.' *Knight First Amendment Institute Occasional Papers*, accessed 07/07/2022, https://knightcolumbia.org/content/a-standard-for-universal-digital-ad-transparency.

European Commission (2019). 'Code of Practice against disinformation: Commission calls on signatories to intensify their efforts [Press release].' Retrieved 06/15/2022, http://europa.eu/rapid/press-release_IP-19-746_en.htm.

Facebook (2018). 'An update on our plans to restrict data access on facebook.' Accessed 05/22/2022, https://about.fb.com/news/2018/04/restricting-data-access/.

Facebook (2018). 'Hard questions: Update on Cambridge Analytica.' Accessed 05/22/2022, https://about.fb.com/news/2018/03/hard-questions-cambridge-analytica/.

Feltzmann, H., E. Fosch-Villaronga, C. Lutz, et al. (2020). 'Towards Transparency by Design for Artificial Intelligence.' *Science and Engineering Ethics*, 26: 3333–3361.

Floridi, L. (2020). 'Artificial Intelligence as a Public Service: Learning from Amsterdam and Helsinki.' *Philos. Technol.* 33: 541–546.

Flyerbom, M, P.M. Leonardi, C. Stohl, et al. (2016). 'The Management of Visibilities in the Digital Age: Introduction.' *International Journal of Communication*, 10: 98–109.

Flyverbom, M. (2019). *The Digital Prism: Transparency and Managed Visibilities in a Datafied World.* New York: Cambridge University Press.

Fox, J. (2010). 'The Uncertain Relationship Between Transparency and Accountability,' *Development in Practice,* 17(4): 663–71.

Freelon, D. (2018). 'Computational Research in the Post-API Age.' *Political Communication*, 35(4): 665–668.

Frenkel, B. & C. Kang (2021). *An Ugly Truth: Inside Facebook's Battle for Domination.* New York: HarperCollins Publishers.

Fung, A. & S. Kosack (2014). 'Does Transparency Improve Governance.' *Annual Review of Political Science,* 17: 65–87.

Fung, A. (2009). 'Open government is about more than catching agencies making mistakes, it's about helping citizens protect their interests, scholar says.' NextGov, accessed 6/20/2022, https://www.nextgov.com/cxo-briefing/2009/09/defining-transparency/44704/.

Fung, A., M. Graham & D. Weil (2007). *Full Disclosure: The Perils and Promise of Transparency*. Cambridge: Cambridge University Press.

Gadde, V. & Roth (2016). 'Enabling further research of information operations on Twitter.' *Twitter*, accessed 07/06/2022, https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.

Gerring, J. (1999). 'What makes a concept good? A criterial framework for understanding concept formation in the social sciences.' *Polity*, 31(3): 357–393.

Glaser, B. G. & A. L. Strauss (1999). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Transaction.

Goanta, C., T. Bertaglia & A. Iamnitchi (2022). 'The Case for a Legal Compliance API for the Enforcement of theEU's DigitalServices Act on Social Media Platforms.' *2022 ACM Conference on Fairness, Accountability, and Transparency.*

Google (2017). 'Expanded safeguards for advertisers.' Accessed 05/22/2022, https://blog.google/technology/ads/expanded-safeguards-for-advertisers/.

Gorwa, R. (2019). 'What is Platform Governance?.' *Information, Communication & Society,* 22(6): 854–871.

Grajzl, P. & P. Murrell (2019). 'Toward understanding 17th century English culture: A structural topic model of Francis Bacon's ideas.' *Journal of Comparative Economics*, 47(1): 111-135.

Grimmer, J.& B. M. Stewart (2013). 'Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.' *Political Analysis*, 21(3): 267–297.

Hammond, A. L. (2001). 'Digitally empowered development.' *Foreign Affair*s, 80(2): 96–106.

Hansen, H. K & M. Flyverbom (2015). 'The Politics of Transparency and the Calibration of Knowledge in the Digital Age.' *Organization*, 22(6): 872–889.

Harcourt, B. (2015). *Exposed: Desire and Disobedience in the Digital Age.* Cambridge, MA: Harvard University Press.

Heemsbergen, L. (2016). 'From Radical Transparency to Radical Disclosure: Reconfiguring (In)Voluntary Transparency Through the Management of Visibilities.' *International Journal of Communication,* 10: 138–151.

Hood, C. & D. Heald (*eds.*) (2006). *Transparency: The Key to Better Governance?* Oxford: Oxford University Press.

Howard, P. N. & B. Kollanyi. (2016). 'Bots, #strongerin, and #brexit: Computational propaganda during the UK-EU referendum.' *Social Sciences Research Network*, accessed 6/1/2022, https://ssrn.com/abstract=2798311 or http://dx.doi.org/10.2139/ssrn.2798311.

Howard, P. N. & M. M Hussain (2013). *Democracy's Fourth Wave?: Digital Media and the Arab Spring.* Oxford: Oxford University Press.

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

Jeffries, B. & L. Yin (2020). 'Google's top search result? Surprise, it's Google.' *The Markup*, accessed 07/13/2022, https://themarkup.org/google-the-giant/2020/07/28/google-search-results-prioritize-google-products-over-competitors.

Junius, L. (2018). 'Supporting the European Union Parliamentary Elections.' Google, accessed 07/06/2022, https://www.blog.google/around-the-globe/google-europe/supporting-european-union-parliamentary-elections/.

Keller, D. (2022). 'The DSA's Industrial Model for Content Moderation.' *VerfBlog*, accessed 6/3/2022, https://verfassungsblog.de/dsa-industrial-model/, DOI: 10.17176/20220224-121133-0.

Kenney, M. & J. Zysman (2016). 'The rise of the platform economy.' *Science and Technology*, 32(3): 61–69.

King, G., J. Pan & M. E. Roberts (2013). 'How Censorship in China Allows Government Criticism but Silences Collective Expression.' *The American Political Science Review*, 107(2): 326–343.

Kirkpatrick, D. (2010). *The Facebook Effect: The Inside Story of the Company That Is Connecting the World.* New York: Simon & Schuster.

Klein, N. (1999). *No Logo: Taking Aim at the Brand Bullies*. New York: Picador.

Klonick, K. (2017). The New Governors: The people rules, and processes governing online speech.' *Harvard Law Review*, 131: 1598–1670.

Kosack, S. & A. Fung (2014). 'Does Transparency Improve Governance?' *Annual Review of Political Science*, 17:65–87.

Kracauer, S. (1952-1953). The challenge of qualitative content analysis. *The Public Opinion Quarterly*, 16(4): 631–642.

Laidlaw, E. (2008). 'Private power, public interest: An examination of search engine accountability.' *International Journal of Law and Information Technology*, 17(1): 113–145.

Le Pochat, V., L. Edelson, T. Van Goethem, *et al.* (forthcoming). 'An Audit of Facebook's Political Ad Policy Enforcement.' *31st USENIX Security Symposium*: 1–22.

Leathern, R. (2002). 'Expanded transparency and more controls for political ads.' *Facebook*, accessed 07/05/2022, https://about.fb.com/news/2020/01/political-ads/.

Leersen, P., J. Ausloos, B. Zarouali, *et al.* (2019). 'Platform as archives: Promises and pitfalls.' *Internet Policy Review*, 8(4): 1–21.

Lessig, L. (2009) 'Against Transparency: The perils of openness in government.' *The New Republic*, accessed 06/08/2022, https://newrepublic.com/article/70097/against-transparency.

Levy, K. E. C. & D. M. Johns (2016). 'When Open Data is a Trojan Horse: The Weaponization of Transparency in Science and Government.' *Big Data & Society*, 3(1): 1–6.

Lim, E. T. (2002). 'An Analysis of Rhetoric from George Washington to Bill Clinton.' *Presidential Studies Quarterly*, 322: 328–357.

Manning, C., P. Raghavan & H. Schütze (2008). *Introduction to information retrieval*. New York: Cambridge University Press.

Margetts, H. (2011). 'The Internet and Transparency.' *The Political Quarterly*, 82(4): 518–521.

Meijer, A. (2014). 'Transparency,' in *The Oxford Handbook of Public Accountability*, ed. M. Bovens, R.E. Goodin & T. Schillemans. Oxford: Oxford University Press: 507–524.

Mittelstadt, B. (2016). 'Automation, Algorithms, and Politics| Auditing for Transparency in Content Personalization Systems.' *International Journal of Communication*, 10: 4991–5002.

Moeller, K. & B. Svensson (2021). '"Shop Until You Drop": Valuing Fentanyl Analogs on a Swedish Internet Forum.' *Journal of Drug Issues*, 51(1): 181–195.

Monroe, B. L., M. P. Colaresi & K. M. Quinn (2008). 'Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict.' *Political Analysis*, 16(4): 372–403.

Morozov, E. (2011). *The Net Delusion: The Dark Side of Internet Freedom*. New York: Public Affairs.

Nelson, L. K. (2020). 'Computational Grounded Theory: A Methodological Framework.' *Sociological Methods & Research*, 49(1): 3–42.

Parsons, C. A. (2015). 'Do Transparency Reports Matter for Public Policy? Evaluating the Effectiveness of Telecommunications Transparency Reports.' *Social Science Research Network*, https://papers.ssrn.com/abstract=2546032.

Parsons, C. A. (2019). 'The ineffectiveness of voluntarily produced transparency reports.' *Business & Society*, 58(1): 103–131.

Pasquale, F. (2015). *The black box society. The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.

Persily, N. & J. A. Tucker (*eds.*) (2020). *Social Media and Democracy: The State of the Field and Prospects for Reform*. New York: Cambridge University Press.

Petit, N. & D. J. Teece (2021). 'Big Tech, Big Data, and competition policy: Favoring dynamic over static competition.' *TPRC49: The 49th Research Conference on Communication, Information and Internet Policy.*

Power, M. K. (1997). *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Raleigh, C., A. Linke, H. Hegre, et al. (2010). 'Introducing ACLED: An Armed Conflict Location and Event Dataset.' *Journal of Peace Research,* 47(5): 651–660.

Roberts, A. (2008). *Blacked Out: Government Secrecy in the Information Age*. New York: Cambridge University Press.

Roberts, M. E., B. M. Stewart & D. Tingley (2019). 'stm: An R Package for Structural Topic Models.' *Journal of Statistical Software*, 91(2): 1–40.

Rodman, E. (2019). 'A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors.' *Political Analysis*, 28(1): 87–111.

Rosen, G. (2016). 'Hard Questions: What is Facebook Doing to Protect Election Security?' Facebook, accessed 07/06/2022, https://about.fb.com/news/2018/03/hard-questions-election-security/.

Rosen, J. (2013). 'The Delete Squad: Google, Twitter, Facebook and the new global battle over the future of free speech.' The New Republic, accessed 06/08/2022, https://newrepublic.com/article/113045/free-speech-internet-silicon-valley-making-rules.

Rousseau, J. J. [1722/1982] (1985). *The Government of Poland*. Indianapolis: Hackett Publishing Co.

Salgado, R. (2014). 'Shedding some light on Foreign Intelligence Surveillance Act (FISA) requests.' *Google*, accessed 07/01/2022, https://www.blog.google/technology/safety-security/shedding-some-light-on-foreign/.

Shapiro, A. L. (2000). *The control revolution: How the internet is putting individuals in charge and changing the world we know*. New York: Public Affairs.

Silge, J. & D. Robinson (2016). 'Tidytext: Text Mining and Analysis Using Tidy Data Principles in R.' *The Journal of Open Source Software*, 1(3): 37.

Stewart, C. (2020). 'Google Play is helping to safeguard elections.' *Google*, accessed 07/04/2022, https://blog.google/outreach-initiatives/civics/google-play-helping-safeguard-elections/.

Stohl, C., M. Stohl & P. M. Leonardi (2016). 'Managing Opacity: Information Visibility and the Paradox of Transparency in the Digital Age.' *International Journal of Communication* 10:123-137.

Suzor, N. P. (2019). *Lawless: The Secret Rules That Govern Our Digital Lives*. New York: Cambridge University Press.

Suzor, N. P., S. Myers, A. Quodling et al. (2019). 'What do we mean when we talk about transparency? Towards meaningful transparency in commercial content moderation.' *International Journal of Communication,* 13: 1526–1543.

Suzor, N., T. Van Geelen & S. Myers West (2018). 'Evaluating the legitimacy of platform governance: A review of research and a shared research agenda.' *International Communication Gazette,* 80(4): 385–400.

Taddeo, M. & L. Floridi (*eds.*) (2017). *The Responsibilities of Online Service Providers.* Berlin/Heidelberg/New York/London: Springer.

Turilli, M. & L. Floridi. (2009). 'The ethics of information transparency.' *Ethics and Information Technology*, 11(2): 105–112.

Twitter (2018). 'Update on Twitter's review of the 2016 US election.' Accessed 05/22/2022, https://blog.twitter.com/en_us/topics/company/2018/2016-election-update.

Twohey, M. (2013). 'Americans use the Internet to abandon children adopted from overseas.' *Reuters*, accessed 07/01/2022, https://www.reuters.com/investigates/adoption/#article/part1.

U.S. Department of Justice (2007). 'Fact Sheet: Department Of Justice Corrective Actions on the FBI's Use of National Security Letters,' accessed 08/06/2022, https://www.justice.gov/archive/opa/pr/2007/March/07_nsd_168.html.

Underwood, S. (2016). 'Blockchain Beyond Bitcoin.' *Communications of the ACM*, 59(11): 15–17.

United Nations (2020). 'Report of the Secretary-General: Roadmap for Digital Cooperation,' accessed 06/08/2022, https://www.un.org/en/content/digital-cooperation-roadmap/assets/pdf/Roadmap_for_Digital_Cooperation_EN.pdf.

Van Dijck, J., T. Poell & M. De Waal (2018). *The Platform Society. Public values in a connective world.* Oxford: Oxford University Press.

Waddock, S. (2004). 'Creating Corporate Accountability: Foundational Principles to Make Corporate Citizenship Real.' *Journal of Business Ethics*, 50(4): 313–327.

Wagner, B., K. Rozgonyi, M. T. Sekwenz, *et al.* (2020). 'Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act.' *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*: 261–271.

Weigend, A. (2017). *Data for The People: How to Make Out Post-Privacy Economy Work for You*. New York: Basic Books.

Weissmann, S. (2019). 'How Not to Regulate Social Media.' The New Atlantis, 58: 58-64.

Wheeler D & J. Elkington (2001). 'The end of the corporate environmental report? Or the advent of cybernetic sustainability reporting and communication.' *Business Strategy and Environment,* 10(1):1–14.

Wiedemann, G. (2016). *Text Mining for Qualitative Data Analysis in the Social Sciences*. Berlin/Heidelberg/New York/London: Springer.

Woolley, S. (2016). 'Automating power: Social bot interference in global politics.' *First Monday*, 21(4).

Youyou, W., M. Kosinski & D. Stillwell (2015). 'Computer-based personality judgments are more accurate than those made by humans.' *Proceedings of the National Academy of Sciences of the United States of America*, 112: 1036–1040.

Zuckerberg, M. (2011). 'Our commitment to the Facebook community.' *Facebook*, accessed 07/01/2022, https://about.fb.com/news/2011/11/our-commitment-to-the-facebook-community/.