Leaking black boxes: Whistleblowing and big tech invisibility

by Philip Di Salvo

Abstract
In a time when socially impactful technology plays a central part in a variety of political and societal dynamics and processes, new forms of secrecy have emerged. The “black box” metaphor is used to define socio-technical systems that operate in non-transparent and prone-to-abuse ways. Frequently, Big Tech companies, their platforms, services and practices have been described as such, especially for their secretive nature and lack of transparency. Whistleblowers and leaks have contributed extensively and at various levels to the understanding of these systems, providing otherwise unaccessible information for public debate. Based on the discussion of a series of recent instances and a review of the available literature, this paper discusses the peculiarities of whistleblowing from Big Tech companies, and how the practice is helping to shed light on various and new technological black boxes and secrecy, while also expanding the scope of whistleblowing itself.

Contents

Big Tech, obscurity, black boxes and leaks
Whistleblowing, secrecy and technology
“The Facebook Loophole” and Sophie Zhang
Google’s “Project Dragonfly” and Jack Poulson
Facebook content moderation anonymous leaks
Unbundling Big Tech whistleblowing
Discussion and conclusions

Big Tech, obscurity, black boxes and leaks

Recently, whistleblowing cases involving Big Tech companies such as Facebook and Google, among others, have multiplied. Overall, Big Tech whistleblowers have the chance to reveal details about how information and communication technologies (ICTs), digital platforms, business decisions of technological companies and their “dysfunctional im-pacts” may affect society at large [1]. This paper argues that Big Tech whistleblowers represent a unicum, do carry some peculiar characteristics and are expanding the scope and role of whistleblowing and leaks in contemporary society. As it will be further discussed, Big Tech whistleblowers’ actions go beyond the exposure of business wrongdoings and corporate malpractices and are in some regards capable of revealing some of the inner functionings of the datafied society itself, by
exposing how some of its core socio-technical systems operate in invisible ways and the secrecy surrounding them. The role of these whistleblowers is even more emphasized in light of the almost hegemonic success of Big Tech companies in the digital realm (Moore and Tambini, 2018), where digital infrastructures are owned by private for profit conglomerates which are in the position of shaping the nature and power distributions in the contemporary public sphere and society. Additionally, whistleblowing has also represented, in various occasions, the only effective strategy to shed light on controversial and obscure elements of Big Tech companies that would have otherwise remained secret or invisible to the public, thanks to various “blackboxing” strategies (Latour, 1999). Or, to put in more direct terms, whistleblowing has been one of the most powerful tools for enabling scrutiny and accountability into Big Tech companies’ otherwise non-transparent actions and practices. In 2018, for instance, the Cambridge Analytica (CA) case started thanks to the revelations that former CA employee Christopher Wylie made to journalists in the U.K. and U.S. (Cadwalladr and Graham-Harrison, 2018). The CA activities and business have not surprisingly been defined as a “black box” itself (Hu, 2020) and the initial whistleblowing led to a huge debate and scandalization (Di Salvo, 2019), to a U.S. Federal Trade Commission (FTC) investigation and to a record-breaking US$5 billion penalty imposed on Facebook (Hu, 2020), confirming once more the crucial role of leaks and whistleblowing when it comes to scrutinize technological black boxes. The CA leak, like the others that will be discussed in this paper, have to be seen in the context of what Bridle (2018) has described as the “new dark age”. The obscurity that Bridle relates to digital technologies and their power encompasses technical details and their effects on society making both of them sometimes “hard to see”, as Crawford and Whittaker (2016) have argued in relation to artificial intelligence (AI) systems. In Bridle’s (2018) view, again, this obscurity has also being reinforced through an abundance of almost magical and sublime metaphors, such as the one of “the cloud” to indicate the ubiquity, portability and accessibility of online services, making them even more “hard to see” (Crawford and Whittaker, 2016) and controversial, when it comes to power relations [2]. The “black box” metaphor has become increasingly popular in technology-related issues. Bruno Latour (1999) has defined the process of “blackboxing” as the way “technical work is made invisible by its own success” [3] to the point that technology carries now a “fundamental opacity” [4]. In Frank Pasquale’s view, the “black box” term carries instead a dual meaning: it can refer to a recording device with wide surveillance capacities and to a “system whose workings are mysterious” [5]. In particular, technological black boxes are usually defined as such because they let only their inputs and outputs be visible, while to “tell how one become the other” is often difficult [6]. Again in Pasquale’s view, Big Data technologies — including those used for reputation and search, as well as in financial services and algorithms at large — sit among the “black boxes” of the greatest efficiency (Pasquale, 2015): Big Tech companies and their algorithms in particular fit perfectly within the metaphor, as they operate under the veil of at least two layers of secrecy: a technical one and a legal one. According to Pasquale, this leads to having “interlocking technical and legal prohibitions” that “prevent anyone outside such a company from understanding fundamental facts about it” [7]. Big Tech conglomerates have put themselves in the unprecedented position of having access to the vastest amount of data and personal information about their users without revealing much about the mechanisms that govern their extraction capabilities or how certain outcomes are produced by their technologies. In the case of Google, for instance, the company’s data extraction capabilities have always been protected by secrecy and obfuscation in order to hide its “exclusive capabilities in unilateral surveillance of online behaviour” [8]. Additionally, and following Latour’s (1999) definition of “blackboxing” again, it is also the success and efficiency of Google’s technologies that have contributed to their “invisibility”.

With the inner mechanisms of platforms kept as well guarded and controlled corporate secrets and their operations and uses becoming increasingly secretive and opaque, “our knowledge of their inner workings is all too often partial and incomplete” (Brevini and Pasquale, 2020), sharpening information imbalances between Big Tech companies and their users and resulting in a even more pervasive accumulation of power. The invisibility of Big Tech companies doesn’t only impact on the functioning of their technologies, but on society at large, as the companies’ platforms have become the “equivalent of a public square” used for political discourse [9]. These companies, while operating as for-profit entities, do perform de facto “civic functions” that none of their predecessors in the private sphere has never been in the position of performing [10]. More precisely, platforms operated by Big Tech companies have also increasingly embodied the role of infrastructures where citizenship takes place, to the point that citizenship in the digital realm is being
Leaking black boxes: Whistleblowing and big tech invisibility

“privately and commercially mediated”, a dynamic that transferred an enormous and unprecedented power as well as crucial social roles to those private companies [11].

Ulises Ali Mejias (2013) has described this reliance on private platforms for social life as one of the most evident symptoms of the ongoing “privatization of social life” [12] that brings individuals to “increasingly live their lives through these services” [13]. Moreover, the online spaces provided by Big Tech companies do “exemplify the Habermasian vision of diverse groups engaging in practical discourse more than any other realm today” [14], making them the actual spaces where the functions of the public sphere are now taking place. As such, company policies and decisions, even when taken for commercial reasons, end up having “enormous consequences for freedom of speech, association, and access to information” [15], among others. Thus, when technologies made available by Big Tech companies become black boxes, their invisibility and secrecy impacts on society at large and its dynamics, not only on technology per se and their performances.

Technological “black boxes” are sometimes discussed in light of how to “open” them in order to provide higher levels of transparency and accountability to such critical and controversial systems (Bucher, 2018; Martin, 2019). Strategies for “opening” black boxes through information and media practices have taken different forms in recent years. For instance, specialized investigative news outlets such as ProPublica [16] and The Markup [17] in the U.S. have used different “computational journalism” (Coddington, 2015) practices to investigate controversial algorithms and other software applications in social media, business, justice and law enforcement activities. These journalistic investigations into algorithmic accountability have frequently used reverse engineering methods to “see inside” technological black boxes and expose their inner workings and code (Diakopoulos, 2015). Parallel to journalistic reporting, also whistleblowing has played an important role in providing access to documents, data and other materials that contributed greatly to the opening of a series of technological black boxes. Overall, the growing centrality of whistleblowing and its frequency in recent years have been indicated as “symptomatic of a powerful shift of the boundaries between the private and the public” which has made practices of public institutions and companies with structural and crucial public roles often “forcefully laid bare of the public gaze” [18]. In this sense, whistleblowing can be seen as a symptom of an increase reliance on secrecy in various realms of contemporary society that makes leaking information more and more frequent.

This article looks at Big Tech whistleblowing through the analysis of three recent examples, analyzed as “descriptive case studies” (Tobin, 2010) by means of a document analysis [19] of news stories dedicated to the various selected cases. The case studies are the following: the revelations around political manipulation and fake engagement strategies on Facebook, provided by the company’s former data scientist turned whistleblower Sophie Zhang — also known as “The Facebook Loophole” investigation (2020); revelations about Google’s “Project Dragonfly” made by the former research scientist Jack Poulson turned whistleblower (2018); the various anonymous leaks about content moderation practices at Facebook, published by different news outlets since 2017 and 2021 [20]. These examples have been chosen because they potentially “give maximal information about the specific features and characteristics” (Bleijenbergh, 2010) of Big Tech whistleblowing and are able to offer different insights about leaks can take place from within Big Tech companies. In particular, descriptive case studies can reveal “patterns and connections, in relation to theoretical constructs” (Tobin, 2010): for this article, the case studies will be analyzed in light of Ceva and Bocchiola’s (2019) and Bok’s (1984) frameworks about whistleblowing and secrecy, respectively, and in light of the growing literature around technological black boxes. Before presenting the case studies, the next section will discuss whistleblowing at large and its role in providing accountability in various areas.

Whistleblowing, secrecy and technology

Whistleblowing is a topic with an established tradition of research in organizational studies, business ethics
and law since the 1970s [21]. Despite this body of literature, various scholars agree that the phenomenon still lacks a “universally accepted definition” [22]. Near and Miceli define whistleblowing as: “the disclosure by organization members (former or current) of illegal, immoral and illegitimate practices under the control of their employers to person or organizations who affect action” (Near and Miceli, 1985) [23]. Generally, malpractices exposed by whistleblowers are such because they threaten the public interest [24]. Generally, lines are drawn between “internal” and “external” whistleblowing (Dworkin and Baucus, 1998) in order to differentiate between recipients of leaks who stay inside or outside the affected organizations, respectively. Whistleblowing towards media outlets and journalists is the quintessential “external” whistleblowing case; in “internal” whistleblowing instances, instead, whistleblowers reach our to their employers, superiors or corporate hotlines within their organizations. In both cases, though, whistleblowing is the realm of the “unauthorized disclosure” of information [25] as a form of dissent. Whistleblowing is contrasted to organizational loyalty as it is a breach of fidelity in form of disclosure of information that is supposed to be confidential or even secret [26]. Bok (1984) outlined a typology of secrecy, indicating whistleblowing as a consequence of “strong tensions” caused by concealment of negligence or abuses [27]. As she argued before the success of Big Tech companies and their platforms, “entire new professions are now being trained to use these methods for probing and concealing secrets” [28].

Inside technological companies, particular attention is devoted to the protection of corporate secrets (i.e., inner functionings or source code of algorithms) or to the obfuscation of technologies’ specifications. Several instances have emerged in the past years: Google, for instance, has made the news for having taken explicit and particularly restrictive initiatives to stop employees talking to the press, in order to potentially curb leaks and other unauthorized disclosures (Lecher, 2017; Knowles, 2018). This quest for secrecy has contributed to an overall “blackboxing” of Big Tech and to the creation of what Olesen has defined as a new “invisibility”, where “increasingly complex modes of digital data production and usage generate new grey areas that escape legal jurisdiction and democratic oversight” [29]. As Olesen argues, it is because of this new “invisibility” that whistleblowing has proliferated in the last decade and this applies to Big Tech companies as well: for instance, some whistleblowers potentially could possess direct access and familiarity with those black box technologies and their applications that are not otherwise accessible, visible or easily understandable from outside. Employees-turned-whistleblowers navigate and operate within those grey areas (and those black boxes) and may thus be the only available option for “opening” them (Olesen, 2022).

Moreover, from a more technological perspective, it is also possible to argue that surveillance-powered platforms and technologies and digitalization more in general have also some how facilitated the practice of whistleblowing, as the new socio-technically mediated space is also “an ethico-political space, in which institutional and organisational practices, discourses and truths are (re-)negotiated, reflected and potentially transformed” (Weiskopf, Loacker and Heinrichs, 2019). In the next section, the three case studies will be presented, in order to outline their major characteristics and outcomes, discussing the most defining elements of Big Tech whistleblowing.

---

**“The Facebook Loophole” and Sophie Zhang**

“The Facebook Loophole” case emerged in April 2021 when the *Guardian* published a series of articles focusing on state-backed manipulation of Facebook (Wong, 2021a). The investigation was based on the revelations made by Sophie Zhang, a former Facebook data scientist turned whistleblower (Wong, 2021b). In a 2021 Reddit post [30], Zhang said to have gone public with the *Guardian* in what she defined as a “deep dive” into the issue. Zhang had already made headlines before for her whistleblowing act. In September 2020 *BuzzFeed News* had published a 6,600-word memo authored by Zhang on her last day at Facebook (Silverman, Mac, and Dixit, 2020). The memo was originally posted by Zhang on a Facebook’s internal message board and replicated on an external password-protected Web site prior to Facebook’s removal of the memo (Wong, 2021b). Upon leaving the company, Zhang turned down a US$64,000 severance package from Facebook in order to avoid signing a non-disparagement agreement and to have the opportunity to continue discussing the case (Silverman, Mac and Dixit, 2020). The memo, and the
revelations that Zhang later made to the *Guardian* discussed how Facebook “repeatedly allowed world leaders and politicians to use its platform to deceive the public or harass opponents despite being alerted to evidence of the wrongdoing” (Wong, 2021a). Moreover, the case let emerge how “Facebook has allowed major abuses of its platform in poor, small and non-Western countries in order to prioritize addressing abuses that attract media attention or affect the U.S. and other wealthy countries” (Wong, 2021a). Among other elements, Sophie Zhang illustrated how Facebook allowed Honduras President Juan Orlando Hernández to use various “fake engagement” tactics to reinforce his propaganda attempts on the platform. Zhang, whose role at Facebook was researching and reporting on manipulative practices known as “coordinated inauthentic behavior”, had reported the case in 2019 already, but the company didn’t intervene for 11 months (Wong and Ernst, 2021). Similar instances emerged also from Azerbaijan and other countries, with similar responses by Facebook, according to Zhang. According to the whistleblower, the company’s actions had been different in terms of speed and scale when wealthier and countries had been involved, such as the U.S., Taiwan, South Korea and Poland (Wong, 2021a). Facebook responded to Zhang’s allegations voicing its disagreement with the whistleblower’s characterization of the company’s priorities.

---

**Google’s “Project Dragonfly” and Jack Poulson**

In 2018, Google made headlines over “Project Dragonfly”, a censored version of Google’s search engine tailored for China and censorship requirements of the Beijing government. Details about Project Dragonfly emerged initially through an article published by *The Intercept* (Gallagher, 2018a). *The Intercept* original reporting was based on revelations by a whistleblower who spoke on the condition of anonymity. According to *The Intercept*, Google was planning to launch its censored search engine in 2018 through an Android mobile app capable of filtering Web sites blocked by the Chinese Great Firewall and blacklisting sensitive queries (Gallagher, 2018a). According to news reports about “Project Dragonfly”, the app could also link online searches to phone numbers, raising serious surveillance and tracking concerns for the Chinese public (Gallagher, 2018c). Following the publication of these details, Jack Poulson, a former Google employee in research and machine intelligence, resigned from the company in protest and published a letter detailing the rationale behind his decision, raising ethical concerns about “Project Dragonfly” (Gallagher, 2018b). Poulson’s resignation followed a series of other internal actions taken by Poulson and other employees (Poulson, 2018). Before his exit from Google and public actions, Poulson had raised his concerns about “Project Dragonfly” internally and had sparked other internal protests by Google employees, four of whom resigned together with Poulson. In a *New York Times* editorial, Poulson (2019) defined himself as a “conscientious objector” and later founded Tech Inquiry, a non-profit dedicated to assist whistleblowers in technological companies and generally to ethical coding. “Project Dragonfly” was abandoned by Google in 2019 (Hern, 2019).

---

**Facebook content moderation anonymous leaks**

A series of leaks have revealed some of Facebook’s policies and strategies about content moderation, censorship and decision-making processes related to content removal. Contrarily to the previous two case studies, these leaks have been provided to various news outlets anonymously and no details have emerged about the identities of the sources. In 2017, for instance, the *Guardian* published the results of an investigation named “The Facebook Files” into blueprints that Facebook has used to moderate sensitive issues such as violence, hate speech, terrorism, pornography, racism and self-harm (Hopkins, 2017). The investigation also featured articles about handling of Holocaust denial, anti-migrants content and hate speech and shed light on the dramatically poor working conditions of Facebook human
Leaking black boxes: Whistleblowing and big tech invisibility

moderators. Overall, the Guardian had access to over 100 internal training manuals, spreadsheets and flowcharts (Hopkins, 2017). The same year, ProPublica published a similar investigation — based this time on a “trove of internal documents” — detailing the “secret guidelines that Facebook’s censors use to distinguish between hate speech and legitimate political expression” (Angwin and Grassegger, 2017). In 2018, instead, Motherboard obtained access to various training materials for Facebook moderators detailing how the company handles revenge porn and sextortion cases (Cox, 2018). Again in 2018, the New York Times “was provided with more than 1,400 pages from the rulebooks” that content moderators have to follow in order to moderate political speech across the world (Fisher, 2018). In 2019, it was BuzzFeed News obtained “hundreds of pages of leaked documents about Facebook’s moderation policy” distributed to content moderation contractors Cognizant andAccenture employees (Notopoulos, 2019). Finally, in 2021, the Guardian had access to documents detailing how “the company controls its mainly outsourced moderators’ work down to the smallest detail” (Hern, 2021). Despite same differences in focus and span, all these leaks have similar characteristics: they shed light on how content moderation takes place on Facebook, what are the guidelines that the company provides to thousands of contractors with details on incoherences, biases and poorly transparent attitudes by the company towards content moderation, censorship and removal. In some cases they also shed light on the poor working conditions of moderators. Overall, the provided details about “the new gatekeepers” of the Internet culture and decision-making [34].

Unbundling Big Tech whistleblowing

The selected case studies encompass different formats and possibilities of whistleblowing from Big Tech companies and are representative of various Big Tech leaks. As such, all three cases can be interpreted in light of the elements of whistleblowing theory (Ceva and Bocchiola, 2019; Bok, 1984, 1981). First, it is possible to analyze whistleblowing cases according to some of their defining elements: such as their involved “actions”, “objects”, “agents”, “loci”, “addressees” and “motivations” [35]. These elements can be used to unbundle these examples according to how the whistleblowing acts took place; by whom; in which organizations and about which aspects and activities; to whom are the recipients of the acts and, finally, arguing about whistleblowers’ motivations and aims. This way, whistleblowing from Big Tech companies can be analyzed in full, allowing its peculiarities emerge. Secondly, it is possible to discuss the nature and characteristics of these whistleblowing examples according to their connections with the kind of the secrecy that they break (Bok, 1981; 1984). Combined, the application of these theoretical backgrounds can work as a way to critically discuss how Big Tech companies and their “black boxes” are related to new peculiar forms of secrecy, their meanings for the “datafied society” at large, while also attempting to outline Big Tech whistleblowing as a peculiar branch of whistleblowing.

“Actions”, “agents” and “addressees”

In Bok’s theory of whistleblowing [36], a clear “dissent” nature of the act is always underlined. Dissent has to be intended as a form of “public disagreement” which can take various forms. In “The Facebook Loophole” and Google’s “Project Dragonfly” cases, for instance, whistleblowers took typical whistleblowing “actions” to raise their concerns. As emerged by news reports, two whistleblowers initially made their complaints within the affected companies, by raising their concerns to their superiors and employers. Later, they used their exits from Facebook and Google as signalling events, also providing more details as well as discussing their rationales through the publication of memos and open letters. In the case of Sophie Zhang, it was the memo that originally gained journalistic coverage once public, while in Jack Poulson’s case it was the resignation itself and his letter (that followed previous anonymous leaks about the project) that made the news. In both cases, whistleblowing has been used as a “protesting” strategy to be conducted through other forms of mobilization, such as a public resignation and other direct actions in the workplace (De Maria, 2008). Despite the “public” traits of these whistleblowing acts, they did not involve external recipients in the first stage. Thus, both Zhang and Poulson can be defined as internal “agents”, being employees of the affected companies. As for the “addresses” of their whistleblowing acts,
cooperation with the media, for both Zhang and Poulsen, occurred in a second phase and news outlets were not the first parts involved in the process. This has been different, instead, for the “Facebook content moderation anonymous leaks” series that were addressed directly to external news organizations. These instances are clear examples of “external whistleblowing”, since the anonymous individuals behind the leaks all reached out to news outlets and journalists as primary choices. In the case of these leaks, the “actions” taken by the whistleblowers took the form of public complaints, made directly to the media and consequently to the public. Although details about the identities of these anonymous whistleblowers are not available, it is still possible to argue that these leaks came from other internal “agents”, either employees or contractors working on content moderation for Facebook.

“Loci”, “aims” and “objects”

In Ceva and Bocchiola’s (2020, 2019) theory of whistleblowing, the “locus” is the place where whistleblowers acquire information about facts that they report and the location where those facts occur. All three whistleblowing cases discussed here are related to private Big Tech companies, private “loci” in nature. The “objects” of these whistleblowing acts are specific operations, activities and projects of these companies. They relate to perspective projects, such as Google’s “Project Dragonfly”; corporate internal routines and practices, such as those exposed in the “Facebook Loophole” investigation in regards to political manipulation, or corporate day-to-day policies and decision-making, like those on Facebook’s content moderation guidelines. Bok’s theory of whistleblowing is a fruitful lens of analysis, as it clearly theorizes whistleblowing around the notion of “accusation” [37]. Big Tech whistleblowers acted by holding their companies and superiors accountable for certain actions or policies, raising concerns about initiatives that, in their views, appeared as unethical. All of these whistleblowing “objects” do not relate to explicit wrongdoings or illegal activities, but rather to corporate “uses of power” [38] or to various “grey areas” of invisibility (Olesen, 2022). In particular, all these whistleblowing acts appear to have “aims” related to encouraging “corrective action” [39].

Secrecy

Whistleblowing is an act eminently connected to secrecy, as it is an action that occurs to bypass various forms of secrecy put in place by organizations and judged as illegitimate by whistleblowers who decide to take actions. The intrinsic element of “disclosure” of information involved in whistleblowing [40] underlines the connections between blowing the whistle and breaching secrecy. Bok’s (1984) typology of secrets is structured around four kinds of secrecy: “confidentiality”, “trade” and “corporate” secrets, secrets “for research” and “secrets of the state”. Whereas Big Tech whistleblowers reveal de facto corporate secrets when they expose inner mechanisms and policies that are kept away from public scrutiny, it is possible to argue that Big Tech whistleblowing goes beyond traditional views on corporate secrecy and mostly in relation to the public roles that Big tech companies play in contemporary society (Moore and Tambini, 2018). Consequently, although Big Tech whistleblowers can certainly be seen as corporate insiders, they also raise concerns of a broader public and civic value and strictly connected to the power dynamics and actors of contemporary society. Big Tech whistleblowers do not expose corporate secrets solely for the sake of damaging businesses or denouncing malpractices or illicit acts. Contrarily, whistleblowers decide to act because the secrets they aim to expose may “debilitate judgment and choice” and “shut out criticism and feedback” [41] in various area of the “quasi-public sphere” (York, 2010) where their employers play a crucial role in providing societal infrastructures. Overall, Big Tech whistleblowers raise concerns about
practices and projects that, while belonging to private companies, directly relate to inner mechanisms of contemporary society at large.

Discussion and conclusions

As in the three case studies discussed in this paper, Big Tech whistleblowing has gained important media presence and in some cases has been at the core of interventions by public institutions to regulate technological companies, as in the case of Cambridge Analytica (CA). This analysis shows that whistleblowing involving Big Tech companies is mostly constructed around classic elements of whistleblowing. It has, so far, taken the shapes of whistleblowing as traditionally intended, especially when in terms of practices and motivations. Yet, there are some new elements, especially relative to the role of secrecy in contemporary society and the power that platforms can exercise socially [42]. Big Tech whistleblowing, while coming from private companies, carries extremely profound “public” implications. The publicness of Big Tech companies is extremely more impactful compared to corporations operating in different fields, well beyond the perimeter of their business actions. As Muldoon (2022) argued “global platform are in the business of world building” as they wish to create “the entire environment within which we live.” Big Tech companies, through their services hold “platform power”, which is based on “attraction, on networks, policies, practices, and technologies that incentivize billions of end users and millions of third-party partners — including publishers — to join” [43].

Yet, this “publicness” and world building efforts are constructed around new layers of “invisibility” which, given the connections with some fundamental societal realms, is eminently controversial or prone to abuse and are amplified and accelerated “under conditions of growing technological and productive complexity” [44]. As a result, companies holding “platform power” “operate opaquely and asymmetrically as black boxes where outsiders can only see a limited subset of both input and output” [45]. To fully understand the power and role that these technologies hold in contemporary society, we need to look at them beyond purely technological specifications, examining their capacity of governing crucial aspects of social life and by considering them as the “secret sauces” of social life [46].

The power of social media companies is frequently difficult to describe, as their technologies appear “magical” if not even “sublime”. These “sublime” views on technology are generally deterministic in nature and hyped by commercial motivations and, without needed scrutiny, they can become myths or unchallengeable beliefs, masking or removing critical views on technology and normalizing it potential negative outcomes [47]. In this sense, sublime and hyped views on technologies frequently pushed by companies themselves can be considered as obfuscation strategies themselves, aiming at masking technological complexity and its controversies. These over-simplistic and deterministic views also tell something about the nature of the power that Big Tech possess, especially algorithmic power. Moreover, it is also telling how these forms of technological powers tend to operate in opaque ways. Overall, Big Tech companies perhaps are not simply mere economic and business entities, as they can affect politics and culture in profound ways [48]. As Moore and Tambini write, “these companies are therefore different from what has come before” and “they are also bigger” [49]. Non-technical companies may hold immense powers and influence over society, but their influence may still be less overreaching than those of Big Tech. As Ghosh and Srinivasan note, challenges posed by the power and role of Big Tech companies are wider and more profound, based on “profound asymmetries of information and power” [50]. Facebook, Google and other technological giants are frequently described as non-transparent, unaccountable and even reckless in how they operate and, sometimes, abuse the immense power that they have. This is clearly visible, for instance, in journalistic coverage of antitrust issues related to Silicon Valley corporations (Baer and Chin, 2021) and in coverage of algorithms and their malfunctions and controversies in data processing and exploitation (Brake, 2017; Carlson, 2017; Barassi, et al., 2022). In the wake of some recent controversies involving Facebook and its data gathering practices and a series of related malfunctions and abuses, questions whether social media platforms operate with the desirable responsibility and accountability have
Leaking black boxes: Whistleblowing and big tech invisibility

become more pressing. In other cases, corporate policies, decisions and initiatives have been described as weakly accountable, confusing and lacking “real transparency”, for example, in the moderation of content (Windwehr and York, 2020). As MacKinnon noted, “Internet and mobile telecommunications companies (whose functions and relationships are increasingly intertwined) create computer code that functions as a kind of law, in that it shapes what people can do and sometimes directly censors what they can see” [51]. The secrecy surrounding these technologies goes beyond direct corporate and trade consequences, with their capacity to make society operate in secrecy or behind layers of opaqueness. This secrecy refers directly to the structure of what Feenberg calls the “technosystem”, a set of “technical practices aimed at control of the environment”, interpreted and structured as an ensemble of “sociotechnical rational functions” [52]. In this context, the problem with technical power happens to be black boxes and their “invisibility” [53].

In face of such a growing and sometimes shady obscurity, whistleblowing has become an increasingly central and powerful resort to shed light on certain socio-technical phenomena. That it is whistleblowing to take this role is interesting per se; it signals the existence of a structural transparency and accountability problem with technological “black boxes”. The fact that whistleblowers offer potential solutions to this lack of visibility also provides insights into the kind of secrecy that Big Tech companies hold. “Black boxes” do not simply indicate a lack of knowledge or information, they rather suggest the presence of a “knowable known unknown”, something that with the “right resources” could be knowable [54]. Something, it could be argued, that should not be unknown but is such until its visibility vanishes. In this sense, Big Tech whistleblowers have more points of contact with other whistleblowers coming from the public sector, such as Edward Snowden and Chelsea Manning, rather than with classic corporate whistleblowers. For instance, Edward Snowden, who shed light on the extension of the U.S. state surveillance apparatus, also blew the whistle about a “knowable known unknown” of incomparable societal implications. Big Tech whistleblowers operate in similar “grey areas” of ambiguity and uncertainty [55] that affect directly the inner dynamics of contemporary society and power. The fact that these whistleblowers speak out about corporate environments and not public bodies (such as Snowden) is yet another peculiarity of the value of their whistleblowing and, at the same time, a reminder of the inner controversies of the ongoing “privatization of social life” [56].

In sum, Big Tech whistleblowing is a telling phenomenon about the nature and some current controversies of the “datafied society” and its most powerful actors. An even more recent confirmation of this crucial role of whistleblowing has emerged in late 2021, with the revelations made by another Facebook whistleblower, Frances Haugen (Horwitz, 2021). Haugen, a former Facebook manager, contributed to the discussion of a series of flaws in how Facebook handles crucial ethical issues and some of its own most controversial practices by disclosing internal documents and insights to both the Wall Street Journal and the U.S. Securities and Exchange Commission. Haugen’s whistleblowing about algorithm and platform design decisions at Facebook may mark a “notable shift” in the attention and focus of policy-makers in the U.S. and Europe (Hao, 2021). At the moment, whistleblowing also appears to be one of the most efficient strategies to “open” black boxes and shed light to avoid a “new dark age” (Bridle, 2018). All of the whistleblowers examined in this paper have taken action denouncing value conflicts with their companies, a proposition that is inspiring to other whistleblowers, especially in the field of AI (Ryan, et al., 2022). It is possible to argue that whistleblowing — and the public pressure it generates — could become one of the most crucial forces towards a less obscure “datafied society” and a more fair technological environment.

About the author

Philip Di Salvo is a post-doctoral researcher at the University of St. Gallen’s Institute for Media and Communication Management, Switzerland. This paper originated during his visiting fellowship (2021–2022) at the London School of Economics and Political Science (LSE) Department of Media and Communications, funded by the Swiss National Science Foundation (SNSF) (grant: P2TIP1_191492). E-mail: philip [dot] disalvo [at] unisg [dot] ch
Notes

2. Bridle, 2018, pp. 7–8; Mosco, 2014.
5. Pasquale, 2015, p. 3.
6. Ibid.
17. See, for instance, the “Citizen Browser” project by The Markup: https://themarkup.org/series/citizen-browser.
18. Olesen, 2022, p. 296.
20. These cases are difficult to map and identify, especially as they may be published in different languages than those commanded by the author. For the sake of clarity, only cases published in English have been included. Those in other languages, like the Italian one published by Valigia Blu in 2017 have been consequently excluded: https://www.valigiablu.it/facebook-moderazione-fonte/.
23. As quoted in Savage, 2016, p. 5.
Leaking black boxes: Whistleblowing and big tech invisibility


27. Bok, 1984, p. 211.


30. https://www.reddit.com/r/IAmA/comments/mqw86u/i_am_sophie_zhang_whistleblower_at_fb_i_worked_to/.


35. Ceva and Bocchiola, 2019, pp. 17–45.


38. Ceva and Bocchiola, 2019, p. 32.

39. Ibid.


42. Nielsen and Ganter, 2022, pp. 157–188.


44. Olesen, 2022, p. 299.


48. Pasquale, 2020, p. 89.


References


Leaking black boxes: Whistleblowing and big tech invisibility


R. Gallagher, 2018b. “Senior Google scientist resigns over ‘forfeiture of our values’ in China,” The
Leaking black boxes: Whistleblowing and big tech invisibility


Leaking black boxes: Whistleblowing and big tech invisibility


---

**Editorial history**

Received 1 June 2022; revised 5 October 2022; accepted 8 November 2022.
Leaking black boxes: Whistleblowing and big tech invisibility
by Philip Di Salvo.
First Monday, volume 27, number 12 (December 2022).
doi: https://dx.doi.org/10.5210/fm.v27i12.12670