



The Bonfire of Safeguards and the Potential Whistleblower Fightback at Twitter

Stephen Holden

Ph.D. Candidate at Manchester Law School

Senior Contributor of the [Corporate Social Responsibility and Business Ethics Blog](https://corporatesocialresponsibilityblog.com)

Suggested citation (OSCOLA): Stephen Holden, 'The Bonfire of Safeguards and the Potential Whistleblower Fightback at Twitter' (*The Corporate Social Responsibility and Business Ethics Blog*, 16 February 2022) retrieved at: <https://corporatesocialresponsibilityblog.com/2023/02/16/twitter-bonfire>

Following the acquisition of Twitter by Elon Musk for \$44 billion, the company has undergone a system of radical changes including losing over half of its staff, the eradication of teams responsible for human rights and accessibility compliance, and the detection and prevention of hostile information operations, and shifted the huge burden of international regulatory compliance to engineers without intimate knowledge of such subjects. In the shadow of such circumstances, this article asks, where there exists a rogue actor unconstrained by economic, reputational, or industry impacts, what, if anything can be done to prevent clearly identifiable risks of harm, and what can we truly expect of whistleblowers?



Following the takeover of social media company Twitter by billionaire Elon Musk, it would not be unfair to say the company has been thrown into disarray. The litany of questionable decisions made by Musk since acquiring a \$44 billion controlling stake in the company on 28th October is long and includes, but is not limited to, [firing the company's top executives](#) and those with vast amounts of experience of the practical running of the company (not to mention the relationships, soft authority, and goodwill built up over time with regulators, oversight bodies, stakeholders, and advertisers which will potentially walk through the door with them), laying off approximately [half of the existing 7,500 staff](#) via e-mail, and the almost complete lack of understanding that the company has a globalized workforce with a number of employees outside of the US with the associated regulatory misalignment in employment standards. For example, within the UK if a company is seeking to make 100 or more redundancies then employees are entitled to a period of collective consultation for 45 days prior to any decision being made. There are reports that British Twitter staff have had their [access to the companies' systems removed](#) and been placed on gardening leave, with a [subsequent notice](#) for staff to select a representative to engage in consultation. By pre-selecting those employees and placing them on gardening leave prior to any consultation period, it provides an argument that those individuals earmarked for redundancy have already been selected and any subsequent consultation is effectively meaningless window dressing. In addition, when attempting to force resignations if remaining staff do not agree to [significant changes to employment](#) terms without consultation, Musk has failed to take into account the lack of enforceability of such provisions outside of the US. Of those remaining staff within the US, an [additional 1,200](#) decided to leave the troubled company.

Musk radically altered the use of the recognized 'blue tick' as a means of verification that accounts are authentic and instead allowed any account to purchase a verification symbol, resulting in verified (but fraudulent and inauthentic) accounts impersonating, amongst others, previous US President George W. Bush who declared ['I miss killing Iraqis'](#) to which a verified but imposter Tony Blair responded, 'Same tbh', a verified but fake [Nintendo](#) account posting sexual innuendos and Mario making inappropriate hand gestures, and the impersonation of a number of corporate accounts such as insulin producer [Eli Lilly](#) who informed customers that 'we are excited to announce insulin is free now' before having to publicly disavow the account and inform people they would still be charged for the essential life-saving drug. This resulted in a [4.37% drop in Eli Lilly's stock value](#).

Outside of the internal machinations of the company, Musk's 'move fast and break things' approach typical of Silicon Valley has resulted in users mass migrating to other platforms such as Mastodon, an exodus of celebrities and high-profile users, and ["a massive drop in revenue"](#) from advertisers.



Prior to the anointing of Musk as God-King of the site, Twitter had not been without controversy and issues, by example, in May 2022 the [FTC fined Twitter \\$150million](#) for repeatedly breaching privacy commitments and misleading users as to how their data would be used. This followed previous fines such as [€450,000 by Irish authorities](#) for similar breaches. It has come under fire for a failure to regulate misinformation and disinformation related to a range of subjects from the COVID-19 pandemic, to vaccinations (Rosenberg, Syed, Rezaie 2020), and been the platform of choice to undermine the veracity and legitimacy of democratic elections (Chen, Deb, Ferrara 2022), resulting in the eventual removal from the platform of former US President Donald Trump for his part in the January 6th insurrection. It is worth noting that Musk has now reinstated Trump's account, in addition to other users banned for breaching terms of service. It has been used to spread hate speech, amplify conspiracy theories, foster political extremism, and spread discord (Jackson, Gorman, Nakatsuka 2021).

And it is from here that my concerns are borne. The biggest challenges currently facing users of the platform are not simply the reduced numbers of coding-based staff, the changes to the 'verification' policy, or the disregard for the advertising business model which poses an existential threat to the continued existence of the platform, but rather, that all of these issues are resultant of the laissez-faire approach taken by Musk to regulatory compliance and the importance placed upon wider duties to both individual and societal protections from harm.

The actions of Musk thus far are significantly concerning from the perspective of a bonfire of systems designed to provide regulatory alignment, protect the interests of the vulnerable, prevent exploitation, and a complete and utter disregard for the interests of those individuals and markets who are not necessarily making a profit for the site.

Amongst those first fired by Musk included the [entirety of the Human Rights team](#), whereby exposing systemic vulnerabilities within the company's failure to conform to the [UN Guiding Principles on Business and Human Rights](#) which requires them to not only avoid infringing on human rights but to take active steps to address the impacts that stem from their operations – including the use of disinformation in the invasion by Russia of Ukraine predicated on baseless claims of Kyiv being overrun by Nazis. It is unclear how the platform now intends to abide by its commitment to conduct human rights due diligence that includes identifying, preventing, ceasing, mitigating, remediating, and accounting for potential and/or actual adverse impacts on human rights. Human rights due diligence is not a one-time activity, but an ongoing process, which should enable companies to periodically evaluate new risks as they emerge (Ruggie 2020).

In addition, one of Musk's first acts was to fire the [Accessibility Experience Team](#) whose job it was to promote and ensure that features were able to be used by those with various disabilities, and to expand the platform to those with accessibility needs whereby reducing barriers between those with additional needs and other sectors of society.



Responsibilities to ensure reasonable provisions are made for users to access digital services are not simply a moral or even commercial imperative, but within the United Kingdom there arguably exists a legal duty under s.20(6) of the Equality Act 2010 (EQA) which requires information service providers to take steps to ensure the service exists in an accessible format. Clarifying, the Equality and Human Rights Commission published a Statutory Code of Practice for "[Services, public functions, and associations](#)" under the EQA. The Code explicitly states, *"Websites provide access to services and goods, and may in themselves constitute a service, for example, where they are delivering information or entertainment to the public. The duty to make reasonable adjustments requires service providers to take positive steps to ensure that disabled people can access services. This goes beyond simply avoiding discrimination. It requires service providers to anticipate the needs of potential disabled customers for reasonable adjustments."*

Of further significant concern, Musk disbanded much of the team responsible for [Machine learning ethics, transparency, and accountability](#) (META) who were responsible for the exploratory work in ethical AI and algorithmic transparency. The team was formed with the explicit task of auditing algorithms to investigate potential unintended harms and biases. In doing so, it is possible that Musk has further exposed the platform to hostile manipulation and hampered detection. As evident from the Cambridge Analytica scandal, social-media companies have adopted an irresponsible business model, which is based not only on the combination of their addictive-by-design nature and the capability of collecting data about the users' behaviors but also on the ability to target individuals with specific messages tailored to their personal predispositions can have significant impacts on the views and perceptions of swathes of society (Grasso 2018). The use of AI is rising quickly, and it is important to underscore it is based on ethical practices and not allowed to run rampant to the detriment of vulnerable groups or minorities within society (Bakir 2020).

Having decimated teams responsible for ensuring regulatory and legislative standards are met, a number of executives key to ensuring regulatory compliance and safe use of the platform resigned in protest, including chief information security officer Lea Kissner, chief privacy officer Damien Kieran and chief compliance officer Marianne Fogarty.

In [a note reportedly posted by an attorney for Twitter's privacy team](#) on the company's Slack channel and later leaked to media organizations, they express that, *"Elon has shown that his only priority with Twitter users is how to monetize them. I do not believe he cares about the human rights activists, the dissidents, our users in un-monetizable regions, and all the other users who have made Twitter the global town square you have all spent so long building, and we all love."*

The loss of critical departments and their leads has resulted in reports that the remaining engineers themselves will now be held responsible for ensuring compliance with regulations. The basic expectation is that technical experts in computing algorithms and



other practical digital engineering functions may be adequate to be responsible for ensuring compliance with both national and international standards as they pertain to such diverse matters as international human rights, privacy, accessibility, and market regulations is manifestly absurd.

When challenged over this approach, the head of legal for Twitter (at the time of writing) [reportedly responded](#), "Elon is willing to take on a huge amount of risk in relation to this company and its users, because 'Elon puts rockets into space, he's not afraid of the FTC.'"

Recognizing the inherent difficulties individuals within the organization face, and the harms that may be evident resultant of the shift in regulatory burden from compliance and legal teams to technical engineers, the note recommends that individuals reach out through whistleblowing channels to report their concerns.

And it is here that those of us with a vested interest in the practice of whistleblowing and unauthorized disclosures watch closely. It is an ask of ridiculous magnitude to expect an engineer who is possibly readily identifiable through the already slashed staffing number to act as the final safeguard and to curb the worst of Musk's instincts, however, this is the position in which we find ourselves.

The harm faced by innocent users from preventable exploitation in the short and medium term is not difficult to see. It is not beyond the realms of possibility that before this comment is posted there will be reports of scammers setting up accounts pretending to be airline or corporate complaint and help departments, paying \$8 for their blue tick verification, and tricking people out of their financial details or enough information so as to conduct identity fraud and spread misery. The widespread use of hostile information activities by actors such as the Russian state able to evade the tags previously placed on them as affiliated news organizations is quickly visible across the horizon.

The academic literature surrounding whistleblowers frequently describes them as the most effective means to prevent and detect wrongdoing by an organization, and how to hold wrongdoers accountable (Miceli, Near 2005). It may be argued that such motives are frequently at the heart of disclosures (Grant 2002). However, the situation at Twitter presents us with a new problem.

It is broadly accepted that corporate executives when covered by the veil of corporate personhood and separate corporate liability, and therefore subject to very limited circumstances of personal legal liability, will act with regard to the continued benefit of the organization, even if they would rather personally take an alternative course of action, as they have a vested interest in the continued existence and success of the company. Executives rely on the success of the corporation for their own continued livelihood, to continue to operate within that industry, to provide for their families, and



to protect their own investments, as well as not wishing to be seen as the reason for a company's failures. However, this may not be the case with Musk.

It must be asked, what if the head of an organization subject to forms of specific regulation simply pays enforcement bodies no heed and cares not if the whole ship is brought down as a consequence? What if, as may be the case here, Musk is simply so wealthy that the risk of losing \$44 billion of his own and investors' capital does not act as a constraint on his actions or reign in his worst impulses and in reality, affects his life in no meaningful way? What if he does not intend to continue to operate within the social media sector so cares not how he is viewed by others or if he is banned from operating within similar companies? It is often said that reputational risk acts as a significant behavioral constraint, but what if Musk is perceived to be simply so successful in other ventures that killing a global platform such as Twitter does not pose any real risk of harm? Musk has not attempted to hide his actions nor taken steps we are aware of to prevent detection of what he is doing, and it would appear that he is fully aware of the potential for harm, fines, and intervention but simply continues unphased.

As those of us with deep interests in compliance mechanisms, corporate governance, and disclosure provisions hold our breath and await the incoming and inevitable storm, it is yet to be seen what, if any, difference potential whistleblowers will or can make in protecting us from Musk's willful negligence. If whistleblowers raise the alarm, but regulators and enforcement bodies are either unwilling or unable to act until harm has occurred, it raises the question, what is the point of the disclosure, and why should they place themselves at risk of retaliation? While the EU Commission warns that Twitter must '[fly by our rules](#)', it remains to be seen what proactive enforcement measures are being taken to ensure compliance before harm is evident.

But one thing is for certain unless there are individuals within the organization willing to stand up to Musk and speak truth to power and make regulators aware of potential significant harms before they arise, then we must reassess what actual and meaningful proactive safeguards society has against the extremely wealthy actors who have no regards for ethical responsibilities and preventable harms.

Select Bibliography

- Bakir, V. (2020). Psychological operations in digital political campaigns: Assessing Cambridge Analytica's psychographic profiling and targeting. *Frontiers in Communication*, 5, 67.
- Chen, E., Deb, A., & Ferrara, E. (2022). # Election2020: The first public Twitter dataset on the 2020 US Presidential election. *Journal of Computational Social Science*, 5(1), 1-18.



- Equality and Human Rights Commission (2011) Equality Act 2010 Code of Practice. Services, public functions, and associations Statutory Code of Practice.
- Grant, C. (2002). Whistleblowers: Saints of secular culture. *Journal of Business Ethics*, 39(4), 391-399.
- Grasso, C. (2018). Cambridge Analytica and Facebook data breach: A corporate social responsibility analysis. *Corporate Social Responsibility and Business Ethics Blog*, March 27, 2018, available at: <https://corporatesocialresponsibilityblog.com/2018/03/27/cambridge-analytica-and-facebook>
- Jackson, S., Gorman, B., & Nakatsuka, M. (2021). QAnon on Twitter: An Overview. *Institute for Data, Democracy & Politics, The George Washington University, Washington, DC*. Accessed June, 21, 2021.
- Miceli, M. P., & Near, J. P. (2005). Standing up or standing by: What predicts blowing the whistle on organizational wrongdoing? In *Research in personnel and human resources management*. Emerald Group Publishing Limited.
- Rosenberg, H., Syed, S., & Rezaie, S. (2020). The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. *Canadian journal of emergency medicine*, 22(4), 418-421.
- Ruggie, J. G. (2020). The social construction of the UN Guiding Principles on Business and Human Rights. In *Research handbook on human rights and business* (pp. 63-86). Edward Elgar Publishing.
- United Nations (2011) Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect Respect and Remedy" Framework. New York: United Nations Office of the High Commissioner for Human Rights.

Disclaimer

The views, opinions, and positions expressed within all posts are those of the author(s) alone and do not represent those of the Corporate Social Responsibility and Business Ethics Blog or its editors. The blog makes no representations as to the accuracy, completeness, and validity of any statements made on this site and will not be liable for any errors, omissions, or representations. The copyright of this content belongs to the author(s) and any liability concerning the infringement of intellectual property rights remains with the author(s).



The [Corporate Social Responsibility and Business Ethics Blog](#) is a scientific forum for analysis and discussion of corporate issues around the world. It also represents an innovative teaching platform, which is intended to facilitate a global interaction of both undergraduate and postgraduate students.

Areas of interest:

- *Bioethics, Healthcare, and Pharmaceuticals*
- *Business and Human Rights*
- *Corporate Crime and Financial Crime*
- *Corporate Governance*
- *Environmental Ethics and Sustainable Development*
- *Ethics and Responsibilities within the Supply Chain*
- *Ethics of Corporate Power and Wealth*
- *Standards of Health, Safety, and Security*
- *Sustainability of the Food Supply Chain*
- *Technology and Corporate Activities*
- *The Establishment of Moral Organisations*

Editor in Chief

Costantino Grasso (Manchester Metropolitan University, UK)

Editorial Board

Dawn M. Carpenter (Georgetown University, USA)

Jacobo Dopico Gómez-Aller (University Carlos III Madrid, Spain)

Jérémie Gilbert (University of Roehampton, United Kingdom)

Karin Buhmann (Copenhagen Business School, Denmark)

Luca D'Ambrosio (Sciences Po, Paris)

Donato Voza (University of Roehampton, United Kingdom)

Solomon Lumba (University of the Philippines, The Philippines)

Senior Contributors

Liemertje Sieders, Stephen Holden

Contributors

Amir Sherdil Rana, Bianca Oprea, Célia Mokhtari, Cleander Yu,

Eden Benat, Kellisha Harley, Michael DeJesus, Ololade

Durodola, Demet Altunay