Carson Irving

Jeff Franklin

CPR-E 234

1 March 2024

The Ethical Blueprint of Twitter

In 2006 the goldrush of social media was just starting to form. With facebook starting two years before and exploding in popularity, few, if any, could foresee the exponential growth that would ensue. As a creator of Twitter I find myself on the horizon of the booming industry, I am in the position to shape the future of the industry. Amidst the shiny new area of innovation and freedom to create, the ethical implications of my creations loom over. With not many others to look to for inspiration, being a pioneer in the social media realm, I need to predict and analyze which parts of the platform I am creating have the potential to stir ethical controversies. Though I do not have many to look off of with the exact same struggles, there are foundations for me to build off of. These include previously created ethical theories, Utilitarianism, Deontology, Virtue ethics, and Contractarianism. Additionally I have access to the Ethical OS Checklist (EOC), which serves as a comprehensive guide to navigating ethical risks in the creation process. The EOC stands as one of the most applicable resources at my disposal. As I embark on this journey as a pioneer in the social media industry, I recognize the responsibility to integrate ethical considerations into the foundation of Twitter. I hope this recognition will ensure Twitter not only fosters meaningful discourse, but also upholds principles of integrity and user empowerment. By setting a base of ethical considerations into the platform, I aim to create a digital space that promotes responsible and healthy engagement, respects user privacy, and contributes positively to society as a whole.

The first step into considering ethics for a creation is to create a comprehensive outline of where the areas of the highest ethical risks lie. Looking at Twitter in its simplest form can help identify risk areas. At its core, Twitter facilitates users sharing short messages and media with their audience, while also getting recommendations of content they may like in their feed. Even at this raw conceptual stage there are many vectors where ethical dilemmas are possible. However, by looking at the Ethical OS Checklist (EOC) I can narrow down the vectors to find which are most important to address. Upon scanning the eight risk zones on the EOC, four stand out to me: risk zone one, risk zone four, risk zone six, and risk zone eight. Risk zone one highlights the ethical standards of truth, disinformation, and propaganda. To address this risk zone, we first need to answer what kind of information users will expect to find on Twitter and what the platform represents. I would hope Twitter to be taken seriously, where users anticipate encountering truthful and interesting dialog. To uphold these expectations Twitter must implement measures that will combat misinformation and propaganda, but this is a complex issue to tackle. Who is to say that something was not satire vs misinformation, and you can’t possibly moderate every single post that is made on the service. I think this rules out the approach to delete the content, or ban the accounts responsible for the offending users, as I don’t think that having an automated system deciding what is satire or not is practical, while having a massive moderation team focused on misinformation and propaganda would also be a huge financial burden. There is also the issue of bots being used to drive traffic to posts to promote them. This could be used to maliciously get ads or spam into users' feeds, while driving down the quality of the conversations on the platform. I think these issues are some of the biggest the platform faces, though there are solutions. I think while the bots and the misinformation are similar in the way that they drive down the quality of the discourse, though I think the best ways to solve them are different. To address the possibility of misinformation and propaganda there needs to be a way for the users to label posts as such when they come across it. This addresses the issue of moderation as users would be combating the problem themselves, and are motivated to do so as to create a higher quality of discourse. There would need to be a platform in place for a user to provide factual, cited information that can prove a post is misinformation. If enough users reported the post as misinformation, and agree with the credible information provided with the suggested platform, it would display said credible counter argument under the post warning of the uncredible nature. This would create a community vetting process of sorts, and would help provide the platform with the content and conversations users expect. This would represent an approach grounded in virtue ethics by empowering the users to combat misinformation themselves. It promotes honesty and responsibility among other users, encouraging them to take an active role in maintaining the quality of communication on the platform. This approach also displays lesser utilitarian qualities by leveraging the collective of the user base's knowledge to maximize overall utility and credibility of the platform's dialogue. Though this solution creates another vector that bot accounts could be abused on. If a truthful post was a target by bots, it could be flagged as potentially uncredible and at a quick glance seem less reliable. This ties in to the issue with bot accounts listed above, and again hurts the overall reputation of the platform. The issue of bots is combated with cybersecurity, and automation. To mitigate bot accounts I would log IP addresses to look for account creation and use patterns, and analyze how they interact with the service to algorithmically find accounts that are likely bots. I would also use captchas, and redirect suspected bot traffic for analysis to better improve my understanding and hone the detection algorithm. Bot accounts will most likely never be absent from the platform, but having a great understanding in cybersecurity and reacting to new changes from botnets quickly will prevent as much as possible. Having a moderation team will be important as well for accounts that pass through the detection filters, but get reported by users. Though it is key that you do not rely on the moderation team for the majority of cases, as scaling the team to match the amount of bots is not a cost effective, or practical way to deal with the problem. Therefore it is imperative to employ cybersecurity tactics such as creating a bot detection algorithm and utilizing the other methods mentioned above to ensure the integrity and credibility of the platform. These approaches display a more utilitarian lens as it aims to maximize the overall utility of the platform for users by mitigating spam and minimizing the influence the bots have.

Next I would like to transition to zone four of the EOC. Zone four of the ethics checklist focuses on machine ethics and algorithmic bias. I think this is another very important hurdle that needs to be addressed as a bias in the algorithm that pushes content to users could be damaging to the platform. The main way to address this problem is to make sure your development team is as diverse as possible. With multiple developers that have different viewpoints and opinions it is harder for a machine learning bias to appear, if everyone in the development team shares similar opinions a bias is sure to form. Another step you could take to accomplish more diversification is to allow the algorithm to be open source. Taking the algorithm open source will allow not only developers, but users to make approved contributions. This could also be seen as another example of utilitarianism, as it promotes collaboration and will maximize the benefit and utility of the algorithm for all users. From a contractarianism perspective this could be seen as upholding the social contract between the platform and its users. By providing transparency and opportunity for outside improvement to the algorithm the platform is honoring its obligation to provide the best possible experience to its users. By diversifying the developers through the hiring process, and doing so even further by allowing user additions to the algorithm it provides the best chance at avoiding machine learning bias. I think that this diversity of thought should be spread to as many branches of the platform as possible, this will help to avoid any other unnecessary bias. This will ensure as many viewpoints are explored as possible, leading to an overall better platform.

Addressing machine ethics and algorithmic bias is essential for maintaining user trust and ensuring fairness on the platform. However, another ethical consideration that must be made is on the topic of data control and monetization. Data control and monetization is the sixth risk zone on the Ethical OS Checklist. This risk zone is where cybersecurity is the most apparent, and has the most significant impact. The handling of user data prevents not only ethical challenges but cybersecurity risks as well. Protecting users' sensitive data is paramount in maintaining user trust. Moreover, a monetization strategy which utilizes user data must be transparent and consenting to preserve said trust. There is room to test other monetization strategies such as voluntary subscriptions instead of user data, but for guaranteed monetization user data is the best option. To protect user data I would use proper encryption, test employees on information security regularly, and implement stringent access controls to limit unauthorized access to sensitive user data. Regular security audits and assessments would be conducted to identify any vulnerabilities in the systems. Furthermore, I would establish clear policies and procedures for data sharing, retention, and disposal. This is another area where transparency in the process of data protection would be important for users. By sharing all of the measures taken to prevent data breaches in detail it would boost user confidence and trust. To effectively use a monetization strategy of selling user data while conserving user trust you must prioritize that users have consent and control over what data is sold. You also must be completely transparent on who you are selling their data to. While this won't maximize profits, it will keep user trust high while generating a consistent form of monetization. This aligns with a deontological ethical perspective which emphasizes the adhering to moral principles no matter the consequences. While you could maximize profits by selling as much user data as possible with less transparency, it is important to maintain user trust and do what is morally right. Furthermore, as a global platform there are many data protection laws that need to be followed. For instance the data protection directive in the European Union provides stringent safeguards for users information that is required to be adhered to while operating in the Union. Additionally all global data protection laws will be upheld universally, extending beyond the country or union of origin. This approach not only ensures compliance with the highest standards of data protection mandated by law, but also fosters user trust. Users can be confident that Twitter will follow strict and consistent data protection measures regardless of their country of residence.

When examining the Ethical OS Checklist within the context of the Twitter platform, zone eight emerged as particularly significant. Zone eight entails confronting hateful and criminal actors on the platform. In my view, it is crucial for platforms to take proactive measures to prevent criminal behavior. To deter this activity from Twitter, I would ensure the moderation team is well trained to identify and report any evidence of criminal behavior to the correct law enforcement authorities. However, I have a contrasting view on reporting users that criticize their government in regions where it is prohibited by law. I would not instruct the moderation teams to report them, and I would support their anonymity as much as possible. I would establish a firm unwavering guideline aligned with the laws of the United States, this ensures consistency and clarity on my approach to managing the platform’s content. Additionally I would introduce a streamlined process of muting or blocking other users, enabling users to easily remove undesired content from their feeds. It is much better to create easy to understand guidelines, and empower users to cultivate their own feed than to have overbearing rules that some do not agree with. With this system I am confident users will be able to morph their experience to their liking and tailor their own experiences. Though it does not entirely replace them, the user-run ecosystem will reduce the amount of employees needed for every day moderation. The efforts mentioned will prioritize both user safety and freedom of expression. This closely aligns with the consequentialism view on ethics, these aforementioned moderation strategies provide a platform which measures what is right and wrong based on the legal standards of the United States. A consequentialist stance towards criminal and hate-speech moderation promotes a healthy online environment where users can engage at their own discretions.

In conclusion, as a creator of Twitter, I am aware of the ethical implications that a social media platform can provide. In navigating this new landscape I have relied on established ethical theories such as utilitarianism, deontology, virtue ethics, and contractarianism. I have also used the Ethical OS Checklist as a framework to base my decisions around. By analyzing Risk Zones specified in the checklist such as, truth, disinformation, and propaganda, machine ethics and algorithmic bias, data control, and monetization, and confronting hateful and criminal actors, I have strived to uphold what I think is right, while allowing for the input of others to better my initial decision making. Addressing misinformation and propaganda, combating bot accounts through cybersecurity tactics, promoting diversity in algorithm development, and ensuring transparent data control and monetization have been key focal points. While adherence to data protection laws universally, and beyond their country of origin will build trust in Twitter's users. Furthermore, a balanced approach has been taken to moderation that encourages freedom of expression while also keeping users safe. By empowering users to tailor their own experiences while adhering to the United States legal standards, I aim to foster a healthy online environment where users can converse responsibly. Additionally it is essential to recognize that technology will evolve over time, and ethical standards will evolve alongside it. Twitter must remain diligent and open to change to provide the best user experience. By embracing transparency, accountability, diversity, and user empowerment, I hope to build a platform that reflects the values of its users, and contributes to society in a positive way. While these initial efforts are crucial, ethical excellence is ongoing. I am committed to assessing and refining the initial framework I have laid out to meet the ever changing needs and to match the expectations of Twitter's users. I am determined to foster open communication with users to ensure that Twitter evolves in alignment with users expectations and the best ethical practices. This will enable Twitter's users to maintain their confidence, and potentially propel Twitter to success as a platform.