ETHICAL AND RESPONSIBLE USE OF STUDENT DATA IN SOCIAL MEDIA MONITORING SYSTEMS

A Case Study of Wired magazine article entitled 'Schools Are Mining Students' Social Media Posts for Signs of Trouble But should they?'

Group 8: Anisah Bacchus, Christopher 'Kit' Bransby, Ian Paul Grant, Loai Alnouri, Lokesh Kumar Sharma

Queen Mary University of London

Case overview

In response to school shootings and violence, several schools in the USA have enlisted the help of technology companies offering programs to monitor student's social media accounts. In response, this case study will investigate:

- How the value of education can be undermined by poor transparency practices.
- If social media monitoring can infringe on privacy rights.
- How AI technologies can have discriminatory effects on school children.
- The importance of public oversight mechanisms, and how technologies can be misused without it.

Transparency

The students at Lakeview's schools were not made fully aware of the district's use of Firestorm's service, resulting in poor transparency.

Shade and Singh (2016) argue that students may distrust their authority figures due to the ongoing surveillance in schools, which may exacerbate existing health issues such as anxiety, as many youths consider social media a safe space to express their thoughts.

Furthermore, monitoring undermines the values of education as the school administration is focusing on the behavior of pupils outside of school premises, as opposed to school progress.

Accountability and Public Oversight

These monitoring programs run by schools lack Accountability and Public Oversight. The questions to ask here is 'Who is responsible to ensure the compliance of GDPR or other Frameworks? What happens to the data after the student graduates?'

Other notable issue is, many of these social media algorithms can be classified as 'Blackbox', which refers to a system where the quantitative process that leads to an output cannot be explained. Now in most cases schools' administrations are not qualified enough to understand the inner working of these programs and they are unable to raise issues on the reports.

Violation of privacy rights

Simonite (2018) states that only public social media posts are monitored by the algorithm. Just because the posts are public, does that make social media monitoring ethical?

Social media is a space in which thoughts can be expressed freely and non-consensual surveillance of this space, be it during or outside of school time, is a violation of privacy rights.

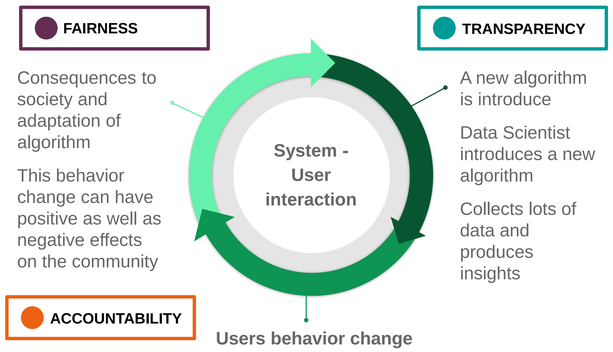
In Philadelphia, 1800 pupils were unknowingly monitored through webcams via school issued laptops (Jessica Shepherd 2010), while in another college the social media accounts of friends and families were monitored to keep the behavior of student athletes in check (Shade and Singh 2016).

Where do we draw the line? In this instance, the line between public posts and privacy was crossed.

Ineffective systems

The effectiveness of this system is dependent on the data it uses. Social media data is useful for development of beneficial systems; however, it is often noisy with idiosyncrasies. As such ground truths may be vague, subjective or simply unavailable (Liu et al. 2016), It is likely that technology based on such data will require a domain specialist such as a school counsellor in its interpretation. This is central to the argument of accountability (Zawacki-Richter et al. 2019).

Any system imposed on an environment may also incur a behavioral change in users potentially rendering models invalid (Shade and Singh, 2016).



Users change their behavior. The nature of the dataset changes. Results change accordingly

Fairness: Algorithmic Bias

Professor Desmond Patton claims that technologies 'may be more likely to misfire on language used by black youths' (Simonite, 2018).

Firestorm and other social media monitoring companies will not be including protected attributes such as race in their datasets. However, algorithms can predict protected attributes from seemingly innocuous ones (Hardt 2016).

- Human data also samples the inherent biases of the real world, therefore a predictive algorithm using this data can potentially reinforce discrimination against minorities.
- COMPAS, a criminal justice algorithm used in Florida, mislabeled African-American defendants as "high risk" at nearly twice the rate it mislabeled white defendants (Angwin et al 2016).

Self Assessment Score Chart

Overarching Principles	Score					
	0	1	2	3	4	5
Transparency		◉				
Accountability	◉					
Fairness		◉				

	Specific Action	Recommendation No.
1.2	Understand unintended consequences of your project	1, 2, 3, 4, 5, 6, 9
1.3	Human rights considerations	1, 2, 3, 4, 5, 6, 9
2.2	Ensure diversity within your team	7, 8
3.6	Ensure the project's compliance with the Equality Act 2010	1, 2, 3, 4, 5, 10, 11
4.3	Bias in data	4, 5, 6, 9

Recommendations

1. Schools should inform students that social media monitoring is happening.
2. Data should be available upon request, with an option to opt out and presented in a way that the students can easily understand.
3. Insights from data should be clarified in close dialogue with students (Hoel and Chen 2018).
4. A Counterfactual Fairness System which evaluates if a decision is fair if the person came from a different demographic background (Kusner et al 2017).
5. AI Fairness 360 data science package to be integrated into the algorithm pipeline (Bellamy et al. 2019).
6. Regularly update software with new 'slang' to mitigate historic bias.
7. Re-evaluation into the structure of management and review into eliminating bias in the HR process to comply with the Equality Act 2010 (Chow, 2020).
8. Introduce unconscious bias training for employees, and management must promote a company culture where every voice is welcome and respected. (Chow, 2020)
9. Open feedback systems fueled by community representatives / domain specialists to be integrated to combat bias.
10. Legal Advisor consultation to ensure the algorithms compliance with Data Protection act 2018 and Equality act 2010.
11. Destroy social media data within one year if a pupil is no longer enrolled in the school system. (Bill Text - AB-1442 Pupil records: social media., 2014).

Conclusion

- Technologies have the potential to reduce harm and ensure safety among students, but it is of the utmost importance that companies follow strict guidelines to ensure fairness, transparency and accountability.
- The solutions will always be a balancing act between conflicting ideas of safety vs privacy or collective vs individual interests
- This case study offers a path forward to strike the right balance for ethical and responsible use of data.

References

Angwin, J. et al 2016. 'There's software used across the country to predict future criminals. And it's biased against blacks.' [online] Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed 23 Mar. 2021] | Bellamy, R. 2019, 'AI Fairness 360: An Extensible Toolkit for Detecting and Mitigating Algorithmic Bias', Ibm Journal of Research and Development PP (99):1-1 | Chow, C., 2020. '15 Ways to Improve Diversity and Inclusion in the Workplace.' [online]. Available at: https://socialchorus.com/blog/15-ways-to-improve-diversity-and-inclusion-in-the-workplace/ [Accessed 23 Mar. 2021] | Kushner, M. et al 2017, 'Counterfactual Fairness', in I. Guyon et al (eds), Advances in Neural Information Processing Systems, Curran Associates, Inc., Volume 30 | Leginfo.legislature.ca.gov. 2014. Bill Text - AB-1442 Pupil records: social media.. [online] Available at: <https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201320140AB1442>. [Accessed 23 Mar. 2021] | Liu, H., et al. 2016.' The good, the bad, and the ugly: uncovering novel research opportunities in social media mining'. International Journal of Data Science and Analytics, 1(3-4), pp.137-143 | Shade, L. and Singh, R., 2016. "Honestly, We're Not Spying on Kids": School Surveillance of Young People's Social Media. Social Media + Society, 2(4), pp.1-12 | Shepherd, J., 2010. 'Schools 'break law' to spy on pupils' [online] the Guardian. Available at: