Undergraduate Thesis Prospectus

Proposing Software to Summaratively Inform Users of Website's Data Policies

(technical research project in Computer Science)

Advocating for Privacy when Personal Information is Currency

(sociotechnical research project)

by

Karim Shoorbajee

November 2, 2020

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.
Karim Shoorbajee

Technical advisor: Aaron Bloomfield, Department of Computer Science

STS advisor: Peter Norton, Department of Engineering and Society

General Research Problem

How can exploitation of online personal information be reduced?

Social media is pervasive worldwide. Most Americans use Facebook and YouTube; most young Americans use Snapchat and Instagram (Smith and Anderson, 2018). Social media companies require account holders to disclose personal information, and they collect use data for ad targeting and to promote content (Johnsons, 2020).Some users, however, object to companies' data collection practices. Consumer advocates, including some prominent former tech insiders, warn that social media companies apply user data to induce compulsive use. For example, according to Tristan Harris, social media companies use data to control users (Thompson, 2017). Harris states, "If I have data, then I know exactly what's going to move [a user's] psychology, and I can persuade your mind in ways that you wouldn't even know were targeted just at you."

A Tool to Summarize a Website's Data Policis.

How can software that informs users on the security and data policies of the websites they use be built?

This is a capstone project in the Department of Computer Science, technically advised by Arron Bloomfield.

I propose a web browser extension that informs users on how user data is leveraged by a website and information on the website's security reputation.

Websites' data/privacy policies are seldom read by users. Hart (2019) reports only 13% of American adults read privacy policies before accepting them. Smith (2014) reports over half of Americans incorrectly believe that privacy policies are assurances of information confidentiality. Facebook was discovered to have stored unencrypted user passwords in its

internal databases (Whitaker, 2019). A website's security policy might influence the information users are willing to share.

Website privacy policies would be crowdsourced from users of the extension. The proposal will include explanation of the tool's architecture, use cases, motivation for building such a tool, and broader impact of the tool.

Users can read privacy/data policies themselves. They seldom do, in part due to their length and complexity (Litman-Navarro, 2019). A Google Chrome extension created by the University of Texas at Austin Center for Identity (Zaeem et al., 2020) uses machine learning to concisely report how a website uses user information. The user must have a web page with a privacy policy open in their browser for it to function. If the user leaves the privacy policy page and returns to it, the tool will have to rerun its machine learning algorithm to summarize the policy information, which can be time consuming.

A non-technical person might not have the time or interest in keeping up to date on information and thus might be less likely to be aware if a website/service they use has a strong security reputation.

I will describe what architectures and technologies would make such a tool possible. I would use databases combined with the application programming interface of web browsers to display a summary of a website's privacy policy to a user. I would explain how crowdsourcing would function to provide concise literature describing the data policies and security reputations of websites.

The proposal will set groundwork for potential future implementation of the tool.

Advocating for Privacy when Personal Information is Currency

How have internet users advocated for their personal privacy against companies that rely on selling user information to advertisers to make revenue (e.g. Facebook and Google)?

Social media companies collect user data and monetize it by using it to target advertisements. Critics of social media and privacy advocates, including some social media users, accuse companies of invading users' privacy and of abusing data to manipulate users. In 2015-18, Facebook granted Cambridge Analytica access to vast user data without users' consent; political campaigns then paid Cambridge Analytica for targeted ads.When the abuse was exposed in 2018, Facebook appeared to be all its critics had alleged. During the 2020 pandemic, as millions have depended on technology to collaborate, opportunities for data collection and abuse have proliferated (ACLU, 2020). The Electronic Privacy Information Center (EPIC) is an advocate for social media users' privacy rights. They state that many social media users "feel that their personal details are being circulated far more widely than they would like," (EPIC, n.d.).

According to Lamdan (2015), librarians, as "champions of privacy and intellectual freedom," bear a professional responsibility to resist social media companies' excesses and to equip users to resist.

In its public relations, Facebook (2020) assures users that "it's important that you have choices when it comes to how your data is used," that "we give you control of your privacy," and that "we design privacy into our products from the outset."

Some digital rights advocates perceive user data as users' property, for which data collectors owe them compensation. According to Jaron Lanier (2019), social media companies should pay their users for the data they monetize, stating users "Should have the moral rights to

every bit of [their] data that exists, because they exist forever" According to the Electronic Privacy Information Center (EPIC), Facebook's privacy policy is misleading and exposes its users to risk (EPIC, 2020). EPIC objects to requirements that users' public profiles must include photos and lists of friends, and must disclose gender. It claims that Facebook "misleads users into believing that their information is safe," while disclosing it "to third-party application developers." To the American Civil Liberties Union (ACLU), digital privacy is a civil right. Chris Conley (2011) on behalf of the ACLU states "We shouldn't have to choose between browsing the Web and keeping Facebook from tracking everything we do online,"

To many insiders in the tech sector, lax data collection practices are not a problem of principle but a practical business hazard. To Manish Dudharejia (2018), a tech entrepreneur, "selling data to a third party can be a fantastic revenue stream," but "if you aren't careful, it could end up alienating your customer base."

References

ACLU. (2020, March 27) American Civil Liberties Union. Those "Free" Remote Learning Apps Have a High Cost: Your Student's Privacy https://www.aclu.org/news/privacy-technology/those-free-remote-learning-apps-have-a-high-cost-your-students-privacy/

Dudharejia, M. (2018, April 11). 5 Takeaways for Entrepreneurs From Facebook's User Privacy Mistakes. Entrepreneur. https://www.entrepreneur.com/article/311252

Conley, C. (2012, Feb. 17). The Social Network is Stalking You. ACLU. https://www.aclu.org/blog/privacy-technology/internet-privacy/social-network-stalking-you?redirect=blog%2Ftechnology-and-liberty%2Fsocial-network-stalking-you

EPIC (2020) Electronic Privacy Information Center. Frequently Asked Questions Regarding EPIC's Facebook Complaint. EPIC. https://epic.org/privacy/socialnet/fbfaq.html

EPIC. (n.d.) Electronic Privacy Information Center. EPIC - Social Networking Privacy. EPIC. https://epic.org/privacy/socialnet/

Facebook (2020). Facebook Privacy Basics. Facebook. https://www.facebook.com/about/basics

Glac, K., Elm, D. R., & Martin, K. (2014). Areas of Privacy in Facebook. Business and Professional Ethics Journal, 33(2), 147-176. doi:10.5840/bpej2014111113. JSTOR

Hart, K. (2019, Feb. 28). Privacy policies are read by an aging few. *Axios* https://www.axios.com/few-people-read-privacy-policies-survey-fec3a29e-2e3a-4767-a05c-2cacdcbaecc8.html

Johnston, M. (2020, Oct. 07). How Facebook Makes Money. Investopedia. https://www.investopedia.com/ask/answers/120114/how-does-facebook-fb-make-money.asp

Kerry, C. (2019, Oct. 25). Why protecting privacy is a losing game today-and how to change the game. Brookings. https://www.brookings.edu/research/why-protecting-privacy-is-a-losing-game-today-and-how-to-change-the-game/

Lamdan, S. S. (2015). Social Media Privacy: A Rallying Cry to Librarians. The Library Quarterly, 85(3), 261-277. doi:10.1086/681610. JSTOR

Lanier, J. (2019, Sept. 23). Jaron Lanier Fixes the Internet. New York Times. https://www.nytimes.com/interactive/2019/09/23/opinion/data-privacy-jaron-lanier.html

Rubinstein, I., & Good, N. (2013). Privacy by Design: A Counterfactual Analysis of Google and Facebook Privacy Incidents. Berkeley Technology Law Journal, 28(2), 1333-1413. JSTOR

Smith, A. (2014, Dec. 04). Half of online Americans don't know what a privacy policy is. Pew. https://www.pewresearch.org/fact-tank/2014/12/04/half-of-americans-dont-know-what-a-privacy-policy-is/

Smith, A., & Anderson, M. (2018, March 1). Social Media Use 2018: Demographics and Statistics. Pew. https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/

Thompson, N. (2017, July 26). Social Media Has Hijacked Our Minds. Click Here to Fight It. Wired. https://www.wired.com/story/our-minds-have-been-hijacked-by-our-phones-tristan-harris-wants-to-rescue-them/

Whittaker, Z. (2019, March 21). Facebook admits it stored 'hundreds of millions' of account passwords in plaintext. TechCrunch. https://techcrunch.com/2019/03/21/facebook-plaintext-passwords/

Zaeem, R. N., Anya, S., Issa, A., Nimergood, J., Rogers, I., Shah, V., . . . Barber, K. (2020). PrivacyCheck's Machine Learning to Digest PrivacyPolicies: Competitor Analysis and Usage Patterns (p. 2, Tech. No. 20-10). Austin, Texas: The University of Texas at Austin.