Proposal Information

Proposal FR-S23-01

Title of Project: Protecting Student Privacy: Studying the Data Handling Practices of Online Education Platforms

Anticipated Number of Participants: female: 100 male: 100 **Approximate ages:** 5-18

Submission Date: 2023-02-22 **Anticipated Start Date:** August 1, 2023

Review type: full

Investigator(s):

1. Marshini Chetty
2. Assistant Professor
3. University of Chicago
4. USA
5. Yan Shvartzshnaider
6. Assistant Professor
7. York University
8. CA
9. Danny Huang
10. Assistant Professor
11. New York University
12. USA
13. Jake Chanenson
14. Graduate Student
15. University of Chicago
16. USA

Faculty Supervisor(s):

1. Noah Apthorpe
2. Assistant Professor

Contact information for Principal Investigator:

Name: Noah Apthorpe

Phone: 315-228-7948

Email: napthorpe@colgate.edu

Proposal Text

Purpose of investigation and procedures: The rapid adoption of online education platforms by K-12 school districts during the Covid-19 pandemic has resulted in a lack of transparency about online surveillance of underage students. Online education platforms operate as part of a complex digital ecosystem of services that collect, share, and trade vast amounts of user information. While recent efforts have identified specific privacy concerns involving learning management systems used by colleges and universities, the prevalence of third-party data collection and tracking by online education platforms provided and recommended by K-12 school districts remains unexamined.

This project, a collaboration between researchers at Colgate University, the University of Chicago, York University (Toronto), and New York University, aims to learn the following details about online education platforms used by K-12 students:
* The identities of third-party services embedded in education platform websites that may receive student information.
* The prevalence of third-party services embedded in education platform websites that are known to engage in user tracking (e.g. for advertising or other user profiling activities).
* Whether online education platforms actively share student information (names, addresses, etc.) with third-party services embedded on their sites or located elsewhere on the Internet.

While this study focuses solely on online education platforms, the data collection process will require instrumentation of the actual web browsers used by K-12 students. This will allow us to examine the behavior of platforms that require paid accounts and district-wide licenses. This will also provide the necessary external validity to argue for expanded privacy protections and meaningful changes in platform behavior should privacy violations be detected. Importantly, personally identifiable information about individual students will not be collected as a part of this study.

IT personnel at participating school districts will install a custom Chrome browser extension (developed by the researchers) on district-owned laptops provided to students. Either IT personnel or individual students will then set up the extension by entering the district name and student's grade level (which we collect to ensure data quality) as well as "sensitive" student information (e.g. student name, address) that online education platforms should not share with third-parties. We do not collect this sensitive information – it is stored locally on the student's laptop for use in detecting data sharing by online platforms.

This extension will not collect any further information until a student browses to an online education platform, identified by URL domain using a whitelist of services that are the focus of the study. While the student uses the education platform's website, the extension will record metadata (URL, timestamp, and packet headers) about HTTP requests to and from the education platform and other online services. The extension will send this information, along with a pseudonymous student identifier (for counting the number of reports per student and per district), to the researchers. These metadata will enable us to identify relationships between online education platforms and known online tracking services. They will also help us characterize the scope of entities that could receive student data through HTTP-based content requests/embedding.

The extension will also locally monitor the communications to and from the education platform and other online services in an effort to detect the exfiltration of the "sensitive" student information entered by the student or the district during setup. If detected, the

extension will send a report to the researchers containing metadata about the communication, a pseudonymous student identifier, and the type of information that was detected (e.g., “name” or “email address”). The report will not contain the sensitive information itself, and we will not ever know what a student was doing on any particular educational technology website. All reports sent from the extension to the researchers will be transmitted using industry-standard encryption protocols (TLS). All data will be stored on password-protected servers at New York University that are already used by the researchers for storing data from a prior IRB-approved study [1].

By aggregating reports collected from many students across several participating school districts, we will be able to characterize the privacy landscape of the online education platforms. This will serve as compelling evidence that student data is being shared by online education platforms in ways opaque to school districts, students, and parents. The results will also inform participating school districts about privacy risks associated with the platforms they provide or recommend, and conduct data-driven advocacy for changes to these platforms’ privacy practices.

Anticipated risk and potential benefits to participants: Participating school districts may benefit from an aggregated summary of the privacy risks associated with various online education platforms used by their students and may choose to disassociate with services that engage in excessive student tracking or data collection.

The proposed study method poses no risks to student participants other than that posed by everyday use of the Internet on school-provided laptops. For this type of study, it is reasonable for an IRB to have concerns about student web browsing privacy vis-a-vis the researchers or collaborating school districts. However, the design of our browser extension prevents these concerns as described in the following section.

Steps taken to protect the participants: Students

Our browser extension will only monitor Internet traffic on whitelisted online education platforms (if a student visits any other website, we will not record their activity in any form). All identifiable information about individual students (e.g., actual names and email addresses) used by the extension will be stored on students’ local browsers only; the researchers will have no access to this information. Storing this information locally poses no additional risk to the students, as the information is limited to the extension and is not accessible by any websites that the student may visit. The student’s browser (Chrome) will have access to this information, but the browser already has full access to the student’s Google account with their name, email address, etc. Encrypting this information in local storage is not feasible, as encryption keys would need to be managed off-device, which would significantly raise the usability barrier of the extension.

Research reports about the behavior of online education platforms may include the type of information shared by the platforms (e.g. “name” or “email address”) and metadata about these platforms’ communications with other online services, but they will not include any identifiable information about individual students. All reports will be indexed using pseudonymous student identifiers that are linked to specific school districts and student grade levels, but not to actual student identities. This will allow us to report the distribution of data across districts and grade levels, which will be required for peer review.

Districts

We will not publish the identities of the school districts participating in this research. We may publish the following anonymous information about participating districts in order to demonstrate the representativeness of our data:
* State in which the district is located
* Size of the district (e.g., “large,” “medium,” or “small”)
* Number of schools in the district, reported as ranges (e.g. “1-3”)
* Binary indicator of district wealth (i.e. “wealthy” or “non-wealthy”)

Manner of obtaining participants: We will collaborate with K-12 school districts to obtain participants for this project. Multiple members of the research team have existing relationships with school districts near their universities. We will approach these districts for an initial “pilot” run of the study before inviting additional districts for wider-scale deployment.

We are waiting for IRB approval before contacting these pilot districts about this project. Once we have determined whether these districts are willing to participate, we will update the IRB with the contact information of the relevant individuals at each participating district.

Participating in this project will not require students or teachers to engage in any actions outside of their normal educational practices. This project also does not involve the collection of any information beyond what school districts can already collect by virtue of their ownership of school-provided laptops.

We have prepared a consent form to be signed by a relevant administrator at participating districts. If participating districts have their own research approval processes, we will insist that those are completed as well. We have also prepared consent forms for the parents of participating students and assent forms for students themselves.

References: [1] Huang, D. Y., Apthorpe, N., Li, F., Acar, G., & Feamster, N. (2020). Iot inspector: Crowdsourcing labeled network traffic from smart home devices at scale. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(2), 1-21.

[2] Chanenson, J., Sloane, B., Morrill, A., Chee, J., Rajan, N., Huang, D., and Chetty, M. Uncovering Privacy and Security Challenges In K-12 Schools. CHI. 2023. Forthcoming.

Consent Form: Informed Consent (docx file)

Proposal Document: Consent Form for parents (docx file)

Proposal Document: Assent Form for children (docx file)

Approval: Approval Doc (pdf file)

Approved on: 2023-03-08
Expires on: 2024-03-08

Download proposal and documents

If you experience any difficulties with the IRB online system, please contact Timothy Collett <tcollett@colgate.edu>