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.

[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.

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. 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")