

# **How Well Are Educational Technology Vendors Following Privacy Regulations?**

**Completed paper**

**Mark J. Keith**

Brigham Young University  
mark\_keith@byu.edu

**Justin Giboney**

Brigham Young University  
Justin\_giboney@byu.edu

**Lisa LeVasseur**

Internet Safety Labs  
lisa.levasseur@internetsafetylabs.org

**Bryce Simpson**

Internet Safety Labs  
bryce.simpson@internetsafetylabs.org

## **ABSTRACT**

Educational technology (EdTech) research often mentions the importance of privacy considerations but has performed little empirical investigation relative to other disciplines like healthcare and business, where recent studies indicate that extensive data collection and third-party data sharing and selling are increasing. Indeed, massive online data aggregators have emerged which purchase the data based on online behaviors of website visitors and app users together to create Internet-wide profiles of all people, including children under 13—the age of significant legal protections according to the General Data Privacy Regulation (GDPR) and the Children’s Online Privacy Protection Act (COPPA). One reason for the lack of privacy research in EdTech may be the technical difficulty of assessing vendor data collection and sharing practices. We contribute by auditing the behaviors of 1,292 of the most common EdTech apps to determine the current state of vendor privacy practices. Our findings reveal that extensive data collection and sharing is occurring, which may violate relevant laws like COPPA and GDPR.

## ***Keywords***

Information privacy, EdTech, network traffic, cultural and social implications.

## INTRODUCTION

The industry for instructional or educational technology (EdTech) has grown exponentially to a global market size of approximately 123.4B USD in 2022 and is expected to reach 348.41B by 2030 (GrandViewResearch.com, 2023). This growth is happening despite significant information privacy risks that have long been identified as a relevant issue in EdTech research (Hussain, 1979; Kumi-Yeboah et al., 2023; Reidenberg & Schaub, 2018). Many studies have acknowledged the relevance of information privacy to students, parents, and teachers (Chen, 2022; Madni et al., 2022; Walker et al., 2023), and much of the prior theorizing about information privacy is based on the perspective of those who have to select app vendors and disclose information (Bélanger & James, 2020; Dinev & Hart, 2006). However, there is little theorizing or empirical validation about the information collection and sharing practices of EdTech vendors (Jibb et al., 2022).

Studies of these sorts have been remarkably informative in other disciplines. A recent study found that 19 of the 24 top-rated health-related apps from the Android app store shared user data with 55 unique entities (Grundy et al., 2019). The study's researchers concluded that data-sharing practices were “far from transparent” compared with the actual disclosures made by providers about their data collection and sharing practices (Grundy et al., 2019, p. 1). More relevantly, Pimienta et al. (2023) performed a small-scale study that tested 25 highly-rated apps from various categories and found that all 25 sent user data to 165 unique hosts. However, no studies we know of have specifically tested apps designed as EdTech, where regulation plays a more significant role in protecting children. This was confirmed by a thorough literature review (Jibb et al., 2022) that research investigating apps designed for children only examined the data sent from apps but did not empirically validate it by examining Internet traffic.

Importantly, regulations such as the Children’s Online Privacy Protect Act (COPPA) and the General Data Privacy Regulation (GDPR) provide extensive protections and prohibit third-party data sharing for users under the age of 13 (or 13-16, depending on the state for GDPR) (COPPA, 1998; GDPR, 2016). Children have been identified as a “vulnerable” privacy population (McDonald et al., 2020) because they have limited capacity to make their own EdTech adoption decisions; instead, they rely on teachers, administrators, or other decision-makers to consider their information privacy interests for them (Anonymous, 2024).

With the explosive growth of online data aggregators (Maras & Wandt, 2019), it is possible, and perhaps likely, that these practices of collecting and sharing data have only increased since those prior-mentioned studies of network traffic analysis were performed. The time is ripe for an audit of data collection and sharing practices of EdTech apps. Our research questions are:

*RQ1: What are the data collection practices of the most popular EdTech apps?*

*RQ2: What are the data-sharing practices of the most popular EdTech apps?*

*RQ3: How do these practices (dis)agree with relevant regulations?*

To answer these questions, we begin by reviewing the relevant regulations that shape the data collection and sharing practices of EdTech vendors. Then, we provide a theoretical explanation of how the current business and legal environment shapes the privacy practices of EdTech vendors. This provides a framework to place the actual data practices in context. Next, we review the criteria used for selecting EdTech apps to audit. Then, we explain the network traffic auditing process used to generate our results. After reviewing the results in detail, we identify implications for research, practice, regulators, and those with legal expertise.

## RELEVANT REGULATIONS, CONTRACTS, AND AGREEMENTS

### National and International Regulations

Perhaps the most recent pioneering piece of information privacy regulation relevant to children is the General Data Protection Regulation (GDPR 2016). It established key terms and concepts that have now become commonplace. For example, *personal data* refers to any information relating to an identified or identifiable natural person (a.k.a. “data subject”). It is important to note that not all data is equally risky or valuable to those who want to use it. Therefore, GDPR differentiates between personally identifiable information (PII), sensitive data, and non-identifiable data. *PII* includes information that directly identifies an individual like their name, identification numbers, location data, identifiers like IP addresses, cookies, and more.

*Data subject* refers to the individual whose personal data is being processed. *Data controllers* are the entities (individuals or organizations) that determine the purposes and means of processing personal data. *Data processors* are the entities that handle personal data on behalf of the data controller. *Processing* is any operation or set of operations performed on personal data, such as collection, recording, storage, use, and deletion. One of the most relevant GDPR terms to the rest of this discussion is *consent*, which is a key basis for lawful data processing under GDPR. For children under the age of 16 (or a lower age set by individual member states, not below 13), consent must be obtained from a parent or guardian.

GDPR states that information provided to children about data processing must be in clear, plain language that they can easily understand. Organizations must take reasonable steps to verify that a parent or guardian has given consent for children under the age of consent. Furthermore, data controllers must implement appropriate technical and organizational measures to protect children's data by default, ensuring that only necessary data is processed. GDPR recognizes children as

vulnerable individuals deserving of special protection, particularly regarding marketing and creating user profiles.

This regulation states that websites, apps, and online services targeting children must implement age-verification mechanisms and obtain verifiable parental consent. Schools and educational platforms processing children's data must secure parental consent when necessary. Direct marketing and profiling activities involving children require careful consideration and robust safeguards to protect their privacy and interests.

The Children's Online Privacy Protection Act (COPPA 1998) is a federal law in the United States that applies specifically to children under 13 years of age and their data privacy while using online services. It requires that EdTech providers (in this context) obtain verifiable parental consent before collecting personal information from children. It outlines what must be included in privacy policies, including how operators will use the collected data. It also allows parents to review and delete information that has been collected about their children. COPPA is enforced by the Federal Trade Commission (FTC). COPPA is relevant to our discussion because it strictly prohibits the following activities without parental consent:

- 1) *Targeted advertising*: The use of personal information from children to target specific ads based on that information.
- 2) *Behavioral advertising*: The use of persistent identifiers (e.g., cookies, IP addresses, and other unique user identifiers) that allow children to be tracked over time and across different websites or online services to profile them and target them with ads.
- 3) *Disclosure to third parties*: The personal information of children under 13 cannot be shared with third parties.
- 4) *Geolocation information*: Using geolocation information for targeted advertising

Like GDPR, verifiable parental consent must be obtained for any of the above four data collections to occur. While general advertising is allowed, targeted and behavioral advertising requires parental consent.

While there are other national laws, such as the Personal Information Protection and Electronic Documents Act (PIPEDA) from Canada, China's Cybersecurity Law (CSL), and the Cross-Border Privacy Rules (CBPR), we will limit this paper to US laws as this is where data collection took place.

## **US State Regulations**

Several US states have implemented their own additional regulations, including California with the California Consumer Privacy Act (CCPA 2018) and California Privacy Rights Act (CPRA 2020), Delaware with the Delaware Online Privacy and Protection Act (DOPPA 2016), Utah with the Utah Consumer Privacy Act (UCPA 2023), Nevada with the Nevada Privacy of Information Collected on the Internet from Consumers Act (NPICICA 2008), and Illinois on the Illinois Biometric Information Privacy Act (BIPA 2008).

These state-level regulations cannot contradict federal regulations but can add additional restrictions. One example is that UCPA requires that privacy policies used by EdTech vendors must clearly state how they are following the requirements of federal laws and makes vendor privacy policies enforceable contracts.

## **EdTech App-Level Agreements, Policies, Assurances, and Contracts**

In addition to federal and state regulations, individual contracts can be entered into between EdTech service providers (a.k.a. data controllers and processors) and school districts, providers and students/parents, and providers and third parties who act as auditors or organizers. These contracts cannot contradict state or federal regulations but can add additional requirements.

Contracts are commonly used by local educational agencies (LEA) (i.e., school districts) to establish a set of rules for the usage and management of student data. One example in the US that

is growing in popularity comes from the Student Data Privacy Consortium (SDPC). The SDPC provides a data governance framework for EdTech (Hillman 2023; Zimmerle 2021) that includes a standard contract for LEAs and vendors to use to specify data elements that EdTech providers may collect (Hillman 2023). A key stipulation often found in these contracts is that data sharing is not allowed with any third party other than “subprocessors” who are “performing services on behalf of the Provider” (SDPC 2022). Therefore, even if high-authority state and federal regulations permit certain data sharing, these contracts (and assumably many others) prohibit data sharing that isn’t directly required for the use of the app. SDPC contracts are currently in use in more than half of the states.

In addition to specific app-level contracts, there are third-party agreements and verifications to which EdTech app vendors have claimed to comply. These third parties provide a level of validation or verification of EdTech providers that can inform adoption decisions. We provide two prominent examples but acknowledge there are several more detailed in prior literature (InternetSafetyLabs.org 2023).

First, the Student Privacy Pledge (StudentPrivacyPledge.org 2020) was introduced by the Future of Privacy Forum (FPF) and The Software & Information Industry Association (SIIA) as a formalized commitment by EdTech providers to follow existing federal regulations regarding the collection and handling of student data. EdTech providers can “sign on” to the pledge by completing an application that requires providers to report the portions of their privacy policy that apply to each portion of the pledge found here: <https://studentprivacypledge.org/privacy-pledge-2-0/>. Their website (<https://studentprivacypledge.org/>) states, “The Pledge is not intended as a comprehensive privacy policy nor to be inclusive of all requirements to achieve compliance with all applicable federal or state laws.”

Second, iKeepSafe.org provides a slightly more robust third-party verification of EdTech provider student data privacy practices. Like The Student Privacy Pledge, providers can apply to iKeepSafe.org. However, it is a “certification” process as opposed to a “pledge” or “promise.” While iKeepSafe.org does not provide significant details on their certification approval process, they do claim to use “a series of proxy and web traffic analysis tools to complete the technical assessment, depending on the environment, to reveal the third parties receiving data from the product” (iKeepSafe.org 2024). Afterward, they state, “Based on the findings of the manual and technical assessment, our privacy assessors will work with you to resolve any emergent privacy or security gaps to bring your product into compliance.” Although we can’t verify any of these claims nor how iKeepSafe is completing its technical assessment, this process is certainly more objective than The Student Privacy Pledge.

In summary, neither form of third-party validation is contractually binding with legal protections, nor do they guarantee what EdTech vendors are/aren’t doing with their users' data. COPPA, GDPR, and many state regulations do require EdTech apps to include privacy policies for children under 13. Privacy policies are agreements between a vendor and the end user and must state how the vendor is meeting regulations. Many state-level regulations like UCPA and CCPA offer additional stipulations to these policies to make them more enforceable as contracts.

## **Notable Violations of Regulation**

There are many examples of firms that do not comply with privacy laws or their own privacy policies. Meta (€1.2 billion, €405 million, €390 million), Amazon (€746 million), TikTok (€345 million), and WhatsApp (€225 million) have been hit with the largest GDPR fines thus far (DPM 2023). Also, between 2021 and 2023, the Federal Trade Commission (FTC) of the US brought 97 privacy cases against large organizations (FTC 2024b). For example, Amazon Alexa violated



COPPA by retaining children's voice recordings. Epic Games was penalized \$275 million for COPPA violation of children's data being improperly managed in their popular game Fortnite. EdTech vendor Edmodo was found using children's personal information for advertising in violation of COPPA, which led to a proposal to turn off advertising by default, limit push notifications, and restrict surveillance in schools (FTC 2024a).

This begs the question of what effect regulation truly has on EdTech provider data collection and sharing behaviors and why. We turn to relevant theories that can inform firm privacy practices to answer this.

## **WHY DO VENDORS MEET/VIOULATE PRIVACY REGULATIONS?**

There are several theories that can help explain an EdTech vendor's or service provider's data collection and sharing practices. Trust based theories (e.g., Culnan and Armstrong 1999; Mayer et al. 1995; McKnight et al. 2002) assume that organizations will optimize their data collection and sharing practices to engender consumer trust through transparent and ethical data practices because that is essential to business success. Based on a similar assumption that firms desire to build trust with the consumer, other theories have helped to shape the formation of privacy policies based on the assumption that companies have ethical or moral obligations to protect user information to ensure fairness and respect in the data handling process (Floridi 2006; Solove 2005).

However, given the recent related findings from other disciplines reviewed above (Grundy et al. 2019; Jibb et al. 2022; Pimienta et al. 2023) concerning the actual data collection and sharing practices which result in large amounts of data being sent to large online data aggregators (Maras and Wandt 2019), we believe it is more appropriate to assume that online service providers will take a more utilitarian approach to maximize their own profit and minimize their own costs of

ensuring and maintaining consumer privacy (Cyert and March 2015). To this end, EdTech vendors will maximize their utility by withholding or obfuscating—via complicated “legalese” in their privacy policies (Becher 2019)—their data collection and sharing intentions in order to maximize their profits from selling data while ignoring their customers' privacy risks.

## **Principal-Agent Theory**

Principal-agent theory, or agency theory (Eisenhardt 1989; Fama and Jensen 1983; Jensen and Meckling 2019; Ross 1973), may explain online service providers' current data collection and sharing practices. It examines the “agency” relationship between principals (those who delegate the work) and agents (those who perform work on behalf of the principals). It addresses issues arising under incomplete and asymmetric information conditions when the principal cannot fully monitor the agent's actions. In the EdTech context, this phenomenon is further complicated by the fact that the principal could be either a) the student or parent who must disclose their personal information to the EdTech vendor or b) the teacher, administrator, or other decision-maker who selects the EdTech vendor. The “agency relationship” becomes a three-way relationship between two principals—the decision maker (teacher) and the risk-bearer (student)—and one agent, the EdTech vendor. This relationship is formalized through contracts (Lane and Kivisto 2008).

The concept of “asymmetric information” comes into play because the principal does not know how the agent intends to behave. This leads to a “moral hazard” where the agent takes actions that are not in the principal's best interests. In addition, “adverse selection” occurs when the principal selects an agent that is not ideal because they are not able to accurately decipher whether the agent will act in their best interests. In EdTech, this applies when teachers, administrators, or other decision-makers (i.e., the principals) select EdTech vendors without knowing their true intentions regarding the students' data. A moral hazard occurs when the EdTech vendor collects more data

than they truly need to provide their services and shares or sells that data to large data aggregators, thus making using that vendor an example of adverse selection.

According to agency theory, a moral hazard is mitigated through “incentive alignment,” where the principal offers incentives that align the interests of the agent with their own. However, it is not realistically possible for individual teachers or students to enact incentives for the EdTech vendors. Instead, in the EdTech context, moral hazard is reduced through contracts, legislation, penalties, and enforcement (Lane and Kivisto 2008). To this end, we integrate other theories of the firm that relate to these mechanisms.

## **Institutional Theory**

Nissenbaum (2004) argues that information privacy is about appropriate information flows in specific contexts. She uses this theory to propose that firms develop privacy policies that ensure information sharing aligns with the social norms and expectations of the context. From the consumer side of the information transaction, Bélanger and James (2020) also use the concept of social norms—as introduced by Laufer and Wolfe (1977)—in their theory of multigroup information privacy (TMIP) to explain how individual- and group-level information disclosure norms are formed which influence consumer information disclosure. These social and societal norms concepts are critical to our discussion because they are significantly shaped by relevant regulations (Bélanger and James 2020). However, how exactly does regulation affect EdTech vendor's data collection and sharing practices? Theoretically, EdTech vendors would be obligated to act according to relevant laws and their own privacy policies. But, as stated above, recent evidence indicates that vendors—not just a few, but the vast majority—in other industries have found ways to side-step regulation in their data privacy practices (Grundy et al. 2019).

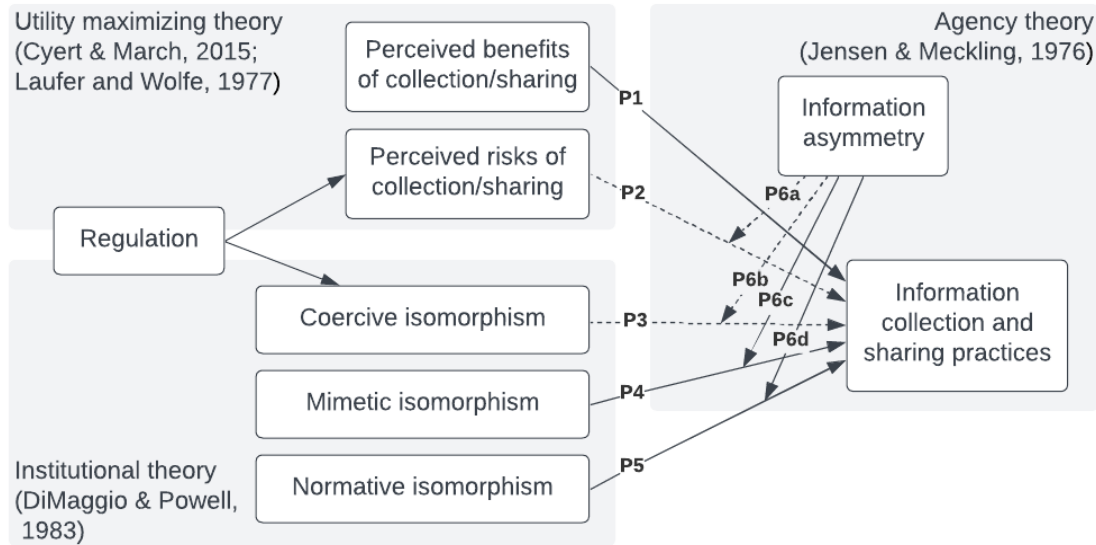
Institutional theory (DiMaggio and Powell 1983) may provide the best explanation for what we are seeing in practice. Institutional theory examines how institutions—defined as established laws, practices, and societal norms—shape the behavior and practices of organizations, emphasizing the influence of broader social and cultural norms. It assumes that organizations are embedded in social contexts that profoundly influence their behaviors and structures. As a result, they are not isolated entities but interdependent with other organizations in their market.

The key concept of institutional theory is isomorphism, which refers to the process by which organizations in a particular field or industry become increasingly similar to one another over time (Deephouse 1996). Three general forms of isomorphism are relevant to this theory. *Coercive isomorphism* results from the formal and informal pressures exerted on organizations by other organizations. This process is enhanced to the degree that one organization is dependent on another exerting pressure. It is also encouraged by societal norms. Coercive isomorphism occurs in most industries, including education worldwide (Anafinova 2020; Marini 2021). It also influences corporate social responsibility (Martínez-Ferrero and García-Sánchez 2017; Othman et al. 2011). *Mimetic isomorphism* occurs when organizations imitate the practices of other organizations that are perceived as more legitimate or successful, especially under conditions of uncertainty (Tingling and Parent 2002). This has been exhibited when successful firms adopt a new technology that others then copy as a “best practice” (Currie 2012; Haveman 1993). Finally, *normative isomorphism* is driven by professionalization and arises from the shared norms and values within a particular professional community. It can result when a community has a similar educational background or comes from professional networks of managers and employees (Mizruchi and Fein 1999).

Of these three types, coercive isomorphism would appear to be the most relevant as it has already been used to explain regulatory pressures and industry standards as a method of aligning the interests of firms with their social responsibilities (Frumkin and Galaskiewicz 2004; Roszkowska-Menkes and Aluchna 2017). Indeed, this is the argument that Chan and Greenaway (2005) laid out concerning how institution theory explains company privacy practices. However, we are beginning to see an isomorphism of not to following legal obligations, but to shirk these responsibilities *en masse*. What may have only started out as a handful of companies sharing and selling customer data has now become possibly 79 (Grundy et al. 2019) to 100 (Pimienta et al. 2023) percent of mobile app firms. Furthermore, it is the technology industry “giants” who do much of the data aggregation and hoarding (Albuquerque et al. 2023) which sets an example for all others. Therefore, we propose that the effect of regulatory coercive isomorphism is lessening in the information economy while the effects of mimetic and possibly normative isomorphism are increasing. This results in greater data collection and sharing practices as firms strive to follow other peer and aspirational firms. In other words, they adopt the position of, “If everyone [or Microsoft, Google, Amazon, Meta] is doing it, then we should do it.”

Are most online service/app providers truly just miming each other’s profit maximization with no thought to the long-term implications for consumer privacy? There is likely more to this phenomenon than profit alone. Institutional theory explains that firms desire legitimacy over efficiency (DiMaggio and Powell 1983). This means that conforming to institutional norms and values is crucial for their survival and success, even if it means adopting practices that may not be optimal in the long run. In the case of EdTech and other online apps and services, many companies outsource their software development process (Uvik 2024). Implementing third party data tracking software in applications for the use of monitoring performance and usage is a common practice in

mobile (Paci et al. 2023) and web applications (Carlsson et al. 2023). If an EdTech vendor wants to conform to the norms of the industry, they will likely follow the same practice of sharing data with large online aggregators who build out Internet-wide profiles of users (Maras and Wandt 2019). Figure 1 visualizes the theory we have been discussing.



**Figure 1. Theoretical Model of EdTech Vendor Privacy Behavior**

In summary, we propose that two primary forces are driving the data collection and sharing behaviors seen across various online apps/services markets, which we test in the EdTech market in this research. First, the need for utility/benefit maximization in a market drives firms to collect more data and share/sell that data for profit (**P1**) (Cyert and March 2015) while the risks—in the form of regulatory penalties (e.g., DPM 2023) and brand devaluation (Acquisti et al. 2006; Martin et al. 2017)—reduce information collection and sharing (**P2**). This is the same assumption used in information disclosure theory from the consumer perspective (Laufer and Wolfe 1977) that drives them to withhold as much data as possible while striving for maximal benefits.

Second, firms have a desire to achieve legitimacy and conformity in their markets (isomorphism), which they sometimes value over utility maximization (DiMaggio and Powell 1983). As prior

research and theory would suggest (Chan and Greenaway 2005), coercive isomorphism, driven by relevant regulations, policies, and agreements, will limit data collection and sharing practices (**P3**). However, unlike prior research, we argue that the current landscape of over-collecting and over-sharing, which is the norm among online providers (Grundy et al. 2019; Jibb et al. 2022; Pimienta et al. 2023) will only increase data collection and sharing practices via mimetic (**P4**) and normative isomorphism (**P5**).

Regulation is a key indicator of the perceived risk of collecting and sharing user data and the degree of coercive isomorphism. But since this relationship is already established in prior theory (DiMaggio and Powell 1983; Laufer and Wolfe 1977), we don't explicitly propose it here. However, we identify that is more salient in the EdTech context than most because of the significant federal and state regulations reviewed above.

The key moderating factor in our theory is the degree of information asymmetry between the students, parents, teachers, administrators, and other decision-makers and the EdTech vendors. If vendors can successfully hide their intentions either through complicated privacy policies that will not be read (Becher 2019) or outright violation of their contractual or legal agreements (e.g., DPM 2023), then they will be less concerned about the risks of their data collection and sharing practices because it is less likely that the principals (i.e., students and teachers) will identify that moral hazard (**P6a**). If information asymmetry is reduced or eliminated, for example, via network traffic audits revealing their data collection and sharing practices (re: Grundy et al. 2019; Pimienta et al. 2023), then it will strengthen the effect of coercive isomorphism based on regulation and contracts because the agent (EdTech vendors) cannot hide their behaviors from the principals; thus, eliminating the moral hazard (**P6b**). On the other hand, if information asymmetry can be maximized by the EdTech vendors, then they will not have to worry about backlash from the

principals, making them more easily influenced by their peer and aspirational firms (**P6c**) and their personal networks (**P6d**).

In summary, what makes the EdTech market unique are the specific regulations (GDPR, COPPA, and state regulations) that apply specifically to data collection and sharing practices. While regulation is relevant in all markets, EdTech is unique because the decision maker (teacher, administrator, and others) is separate from the risk bearer (student or parent). Special requirements have been put in place to protect children who are a vulnerable population (McDonald et al. 2020). EdTech vendors are aware of this dynamic, realizing that it is in their best interests to maintain the highest possible information asymmetry between them and the adoption decision-makers at LEAs. Thus, agency theory is a critical part of the model in Figure 1 designed for the EdTech context as opposed to a traditional consumer-focused vendor who knows their end user is also the decision maker. In the traditional consumer context, information asymmetry is still relevant but much harder to maintain because the consumer is both the risk bearer and the decision-maker (Anonymous 2024). Therefore, app vendors must make great efforts to assure them of their privacy practices so that their perceived risks do not outweigh their perceived benefits (Dinev and Hart 2006). In the EdTech context, vendors do not even mention privacy because it is not relevant to the decision-makers personally, who bear no risk. This is an advantage to the EdTech vendors because any mention of privacy—even when it is being assured through statements and third-party seals—has the effect of heightening individuals perceived privacy risk (Keith et al. 2018; Masters et al. 2022).

## **METHODOLOGY**

Our methodology for data collection is not typical of a theory-building paper. We collected data for quantitative analysis, not to test our proposed variance theory explicitly, but to (1) validate the



findings from related research that online service providers are collecting and sharing data on a large scale (Grundy et al. 2019; Pimienta et al. 2023) and to (2) determine whether this has also become the predominant behavior among EdTech vendors. This is a primary assumption of our theoretical model proposed in Figure 1. Therefore, establishing this assumption provides a basis to test our theory more explicitly in future research.

## Procedures

To test the above assumption, we performed network traffic analysis, which has been established in prior research (Carlsson et al. 2022; Grundy et al. 2019; Joshi and Hadi 2015; Pimienta et al. 2023; Taylor et al. 2017) as an objectively valid method of identifying the personal and sensitive information sent from both mobile and web-based applications as well as operating systems (Liu et al. 2021). It includes using software that captures the data packets sent across the Internet from an individual mobile app or website. The complete details have been outlined elsewhere (Continella et al. 2017). Each app was installed on a standard image to control the testing environment. We registered an account on each app as a child under the age of 13. While logged in (if possible), we performed all the app's primary functions, including visiting each page, performing common tasks and activities, and modifying the user profile.

We aimed to identify *who* and *how many* recipients of the EdTech users were receiving data from each app. This is accomplished by tracking the Internet protocol (IP) address of the transmission control protocol / Internet protocol (TCP/IP) packets sent from each app. Each IP address uniquely identifies a location on the Internet, which can be compared to a known database of domains (e.g., microsoft.com) and sub-domains (ads.microsoft.com), which then identify the company that uses the address. As reviewed above, data may be legally sent to “data subprocessors” who aid the app with its functionality. However, one way we can add value is by checking domain and sub-domain

names against a database of known “identity resolution” or “customer data platform” (CDP) companies. These companies take website or app usage data and, using sophisticated machine learning techniques, connect that data to existing usage behavior to create a consolidated customer profile. This purpose would be in clear violation of COPPA and many state privacy laws for children under 13. Additionally, we can compare domains against databases of known data aggregators (e.g., Microsoft, Facebook, Google, Twitter (X), Amazon) who compile large amounts of online data for various purposes and resell or share the aggregated data. The analysis resulted in an overview of how well/poorly the EdTech market may be following the abovementioned legislation.

## **Sample Description**

In total, we captured network traffic from 1,292 EdTech apps from the United States, making this the largest single study of this kind that we could find in existing research and the first in the EdTech context. These apps were selected based on a variety of criteria. These apps varied based on a variety of factors, including popularity in the app store (based on the number of downloads), those that include advertising and behavioral advertising versus those that do not, apps of different age categories and minimum ages, those that include privacy labels versus do not, and those who’s privacy policies excluded children under 13 versus those that do not. In addition, some were identified based on prior research (InternetSafetyLabs.org 2023) which explored the EdTech listed in a nationwide search of LEA websites. The reasoning for these various factors was to ensure that our results could show the different practices of those who followed the regulations reviewed above versus those that do not.

Table 1 summarizes the sample by category. Some apps were available and categorized in the Android App Store, while others were identified differently. Therefore, the most common category

is “Other” (32 percent) for those not in the Android App Store. Table 2 summarizes how popular the apps were in the Android App Store. Ten percent of the apps had between 0-1,000 downloads, while 5 percent had over 10 million. Tables 3 and 4 indicate how many apps included advertising and behavioral advertising—which is a highly regulated feature of apps designed for children. Eighteen percent included advertising, and 10 percent had behavioral advertising.

Category	Count (%)
Other	417 (32%)
Community Engagement Platform	385 (29%)
Library Management Software	87 (6%)
Game	83 (6%)
Safety Platform	60 (4%)
News	47 (3%)
School Management Software	36 (2%)
Productivity	25 (1%)
Study Tools	25 (1%)
Classroom Messaging Software	20 (1%)
Digital Learning Platform	19 (1%)
Sports	19 (1%)
Reference	19 (1%)
Student Information System	17 (1%)
Learning Management System	11 (0.8%)
School Transportation Software	11 (0.8%)
Virtual Classroom Software	8 (0.6%)
Music	2 (0.1%)
Single Sign On	1 (0.07%)

**Table 1. App Categories**

No of Downloads	Count (%)
0-1,000	139 (10%)
1,000-5,000	96 (7%)
5,000-10,000	36 (2%)
10,000-50,000	77 (5%)
50,000-100,000	26 (2%)
100,000-500,000	60 (4%)
500,000-1M	21 (1%)
1M-5M	50 (3%)
5M-10M	21 (1%)
10M+	77 (5%)
Unknown	689 (53%)

**Table 2. Number of Downloads**

Contains Advertising	Count (%)
No	1,055 (81%)
Yes	237 (18%)

**Table 3. Apps with Advertising**

Contains Behavioral Advertising	Count (%)
No	1,162 (89%)
Yes	130 (10%)

**Table 4. Apps with Behavioral Advertising**

Age Category	Count (%)
Specific Minimum Age	683 (52%)
Everyone	579 (44%)
Teen	30 (2%)

**Table 5. Age Category of Apps**

Contains Privacy Label	Count (%)
No	738 (57%)
Yes	554 (42%)

**Table 6. Contains Privacy Label**

Excludes Children Under 13	Count (%)
No	983 (76%)
Yes	309 (23%)

**Table 7. Privacy Policy Excludes <13**

Minimum Age	Count (%)
Unknown	609 (47%)
4	544 (42%)
12	72 (5%)
17	45 (3%)
10	16 (1%)
9	6 (0.4%)

**Table 8. Minimum Age**

Table 5 indicates how many apps have a specified minimum age. This is important because regulation like COPPA has greater restrictions for users under 13 years old. Table 6 indicates how many apps include a privacy label (57 percent do not). Table 7 indicates how many of the apps' privacy policies excluded children under 13 (76 percent do not). This is important because those who do not exclude children under 13 should not be sharing data or using behavioral advertising per COPPA. Finally, Table 8 indicates the minimum age of the app's intended users.

## RESULTS

Tables 9 and 10 summarize the results of the network traffic analysis described above broken down by app popularity (in terms of downloads) and EdTech app category.

The second column of Table 9 provides the number of apps tested in each download category. The third column shows the average number of unique recipients detected in the network traffic. However, the total at the bottom of this column is the summed product of the number of apps in each category multiplied by the average number of unique recipients. In other words, the 1,292 apps we tested sent data to 19,610 additional recipients. The fourth column indicates the average

number of these 19,610 recipients who are ID resolution, or CDP, companies. Finally, the last column indicates how many of the recipients are data brokers/aggregators (e.g., Microsoft, Facebook, Google, Twitter (X), and Amazon).

No of Downloads	No (%)	Average No of Unique Recipients in NW Traffic	Average No of ID Res / CDP Companies	Average No of Data Brokers Recipients
0-1,000	139 (10%)	18	4	6
1,000-5,000	96 (7%)	22	6	8
5,000-10,000	36 (2%)	14	5	9
10,000-50,000	77 (5%)	18	5	11
50,000-100,000	26 (2%)	11	3	5
100,000-500,000	60 (4%)	11	3	6
500,000-1M	21 (1%)	10	2	1
1M-5M	50 (3%)	12	3	5
5M-10M	21 (1%)	8	2	1
10M+	77 (5%)	11	2	3
Unknown	689 (53%)	15	4	10
<b>Total, sum-prod:</b>	<b>1292</b>	<b>19,610</b>	<b>5,099</b>	<b>10,676</b>

**Table 9. Summary of Data Sharing by EdTech App Popularity**

Table 10 provides the same information broken down by app category. News, sports, and community engagement platforms are the most egregious data sharers/sellers. Virtual classroom software, music, digital learning platforms, single sign-on services, and school transportation software appear to be the best actors on average. However, it is important to note that there were great differences even within each category. Many apps in a particular category did not share data, while some shared with dozens of recipients.

Category	Average No of Unique Recipients in NW Traffic	Average No of ID Res / CDP Companies	Average No of Data Brokers Recipients
Classroom Messaging Software	7	2	3
Community Engagement Platform	24	5	8
Digital Learning Platform	5	1	0
Game	10	2	4
Learning Management System	5	1	0
Library Management Software	11	2	2
Music	12	0	0
News	44	8	9
Other	11	2	4
Productivity	9	1	0
Reference	12	2	4
Safety Platform	6	1	0
School Management Software	7	2	1
School Transportation Software	6	0	0
Single Sign On	5	0	0
Sports	26	5	9
Student Information System	6	4	8
Study Tools	8	1	0
Virtual Classroom Software	3	1	0

**Table 10. Summary of Data Sharing by EdTech App Category**

## DISCUSSION

Our findings reveal extensive data sharing occurring in EdTech vendor’s software. In addition, most data recipients are CDP companies or data brokers and clearly not valid “subprocessors” as allowed in most contracts and privacy policies. In fact, most of the apps tested were caught sharing data with multiple, potentially illegal, recipients.

There are both theoretical and practical implications of these results. First, our findings underscore our theoretical model in Figure 1, suggesting that extensive isomorphism is occurring in the EdTech market around potentially illegal data practices. Companies are mimicking each other because they find it sufficiently plausible to maintain enough information asymmetry i.e., by

hiding their data collection and sharing practices) between them and the teachers, students, and parents who use their apps. While we did not explicitly test our model with any statistical model, our results imply that isomorphism occurs in that there are more than a handful of bad actors. Indeed, the majority are following the same practice to maintain and advance their position in the EdTech market. Future research should more concretely test our model by collecting data about EdTech vendors' intentions and reasoning behind their data collection and sharing practices, although we admit doing so would likely be problematic because it would require a certain amount of self-incrimination on the part of vendors.

The primary practical implication of our findings is that EdTech decision-makers and stakeholders, including teachers, administrators, students, and parents, must find some way to make themselves aware of the privacy risks of EdTech apps *before* they make an adoption decision. Thus far, it has been quite difficult to do this because of the technical expertise required to perform network traffic analysis. However, some non-profit organizations, like Internet Safety Labs, perform these types of analyses and make their data publicly available (e.g., <https://appmicroscope.org/>) specifically for this purpose. Yet, thus far, to the best of our knowledge, no academic research, including case studies or surveys, has indicated that these types of checks are part of any known process (Zimmerle 2021).

The implication for EdTech vendors who are practicing “gray” or potentially illegal data collection and sharing is that information asymmetry will not last forever. As network traffic analysis becomes more commonplace, all practices will become more transparent. Given that research on network traffic analysis is increasing, these vendors should be forward-thinking and ensure that their practices align with all relevant regulations. Simply being third-party certified as a COPPA-

compliant app (e.g., COPPA Safe Harbor certification (FTC 2000)) does not protect an EdTech vendor from FTC prosecution if they are caught in violation.

The implication for lawmakers is that regardless of any legislation put in place, vendors will follow any practice they want if information asymmetry can be maintained. As an anecdote, during our network traffic analysis process, we notified two EdTech vendors of what appeared to be violations of COPPA in their data collection and sharing practices. Their two responses nicely demonstrate routes vendors might take to achieve compliance. The first vendor quickly acknowledged their “mistake” and changed their practice without hesitation. The second vendor initially ignored our notification and made no move until we published some preliminary findings with their name and results. At that point, the vendor contacted us and indicated that their app was never intended for children (although it has been used in many schools for years). Instead, they have a different version of their app intended for children that does not collect or share data illegally. Although the authors have years of experience in EdTech privacy and network traffic analysis, we had never heard of or seen any schools using this “safe” version of their app. At the same time, this vendor updated their privacy policy to specifically state that their primary app is not intended for children. However, it is very unlikely that any school currently using their app will notice this change or stop using it. In summary, some vendors will find ways to subvert the law with deceptive practices like this. We recommend that regulators consider these types of “workarounds” in future legislation.

## **CONCLUSION**

In conclusion, we have demonstrated that pervasive data collection and sharing practices exist in the EdTech marketplace. We have used principal-agent theory and institutional theory in the context of privacy to build a model of motivations for EdTech vendors to compromise their users'



privacy. These practices appear illegal. However, we caution that this research is not an indictment of illegality, and these practices should be reviewed in greater detail by those with legal expertise and who are qualified to interpret the law regarding children's data privacy. In addition, we have provided a theoretical model to explain EdTech vendor behavior, which can be tested more fully in future research.

## REFERENCES

- Acquisti, A., Friedman, A., and Telang, R. 2006. "Is There a Cost to Privacy Breaches? An Event Study," *ICIS 2006 proceedings*), p. 94.
- Albuquerque, U.P., Cantalice, A.S., Oliveira, E.S., de Moura, J.M.B., Dos Santos, R.K.S., da Silva, R.H., Brito-Júnior, V.M., and Ferreira-Júnior, W.S. 2023. "Exploring Large Digital Bodies for the Study of Human Behavior," *Evolutionary Psychological Science* (9:3), pp. 385-394.
- Anafinova, S. 2020. "The Role of Rankings in Higher Education Policy: Coercive and Normative Isomorphism in Kazakhstani Higher Education," *International journal of educational development* (78), p. 102246.
- Anonymous. 2024. "[Authors and Title Withhold to Maintain Double-Blind Review]," *Proceedings of [name of conference withheld]*.
- Becher, S. 2019. "Are Online Agreements Readable?,").
- Bélanger, F., and James, T.L. 2020. "A Theory of Multilevel Information Privacy Management for the Digital Era," *Information systems research* (31:2), pp. 510-536.
- BIPA. 2008. "Illinois Biometric Information Privacy Act (Bipa), 740 Illinois Compiled Statutes 14/1 Et Seq.", from <https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004>
- Carlsson, R., Heino, T., Koivunen, L., Rauti, S., and Leppänen, V. 2022. "Where Does Your Data Go? Comparing Network Traffic and Privacy Policies of Public Sector Mobile Applications," *World Conference on Information Systems and Technologies*: Springer, pp. 214-225.
- Carlsson, R., Puhtila, P., and Rauti, S. 2023. "Towards an Automatic Tool for Detecting Third-Party Data Leaks on Websites," *Proceedings http://ceur-ws.org ISSN* (1613), p. 0073.
- CCPA. 2018. "California Consumer Privacy Act (Ccpa), California Civil Code §§ 1798.100 Et Seq.", from [https://leginfo.ca.gov/faces/codes\\_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5](https://leginfo.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5)
- Chan, Y.E., and Greenaway, K.E. 2005. "Theoretical Explanations for Firms' Information Privacy Behaviors," *Journal of the Association for Information Systems* (6:6), p. 7.
- Continella, A., Fratantonio, Y., Lindorfer, M., Puccetti, A., Zand, A., Kruegel, C., and Vigna, G. 2017. "Obfuscation-Resilient Privacy Leak Detection for Mobile Apps through Differential Analysis," *NDSS*, pp. 10-14722.
- COPPA. 1998. "Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501-6506.."
- CPRA. 2020. "California Privacy Rights Act (Cpra), California Civil Code §§ 1798.100 Et Seq.", from [https://leginfo.ca.gov/faces/codes\\_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5](https://leginfo.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5)
- Culnan, M.J., and Armstrong, P.K. 1999. "Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation," *Organization Science* (10:1), Jan-Feb, pp. 104-115.
- Currie, W.L. 2012. "Institutional Isomorphism and Change: The National Programme for It-10 Years On," *Journal of Information Technology* (27:3), pp. 236-248.
- Cyert, R., and March, J. 2015. "Behavioral Theory of the Firm," in *Organizational Behavior* 2. Routledge, pp. 60-77.
- Deephhouse, D.L. 1996. "Does Isomorphism Legitimate?," *Academy of management journal* (39:4), pp. 1024-1039.
- DiMaggio, P.J., and Powell, W.W. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields," *American sociological review* (48:2), pp. 147-160.

- Dinev, T., and Hart, P. 2006. "An Extended Privacy Calculus Model for E-Commerce Transactions," *Information Systems Research* (17:1), pp. 61-80.
- DOPPA. 2016. "Delaware Online Privacy and Protection Act, Delaware Code Title 6, Chapter 12c." from <https://delcode.delaware.gov/title6/c012c/index.html>
- DPM. 2023. "20 Biggest Gdpr Fines So Far".
- Eisenhardt, K.M. 1989. "Agency Theory: An Assessment and Review," *Academy of management review* (14:1), pp. 57-74.
- Fama, E.F., and Jensen, M.C. 1983. "Separation of Ownership and Control," *The journal of law and Economics* (26:2), pp. 301-325.
- Floridi, L. 2006. "Four Challenges for a Theory of Informational Privacy," *Ethics and Information technology* (8), pp. 109-119.
- Frumkin, P., and Galaskiewicz, J. 2004. "Institutional Isomorphism and Public Sector Organizations," *Journal of public administration research and theory* (14:3), pp. 283-307.
- FTC. 2000. "Coppa Safe Harbor Program."
- FTC. 2024a. "Ftc Proposes Strengthening Children's Privacy Rule to Further Limit Companies' Ability to Monetize Children's Data."
- FTC. 2024b. "Ftc Releases 2023 Privacy and Data Security Update."
- GDPR. 2016. "General Data Protection Regulation."
- Grundy, Q., Chiu, K., Held, F., Continella, A., Bero, L., and Holz, R. 2019. "Data Sharing Practices of Medicines Related Apps and the Mobile Ecosystem: Traffic, Content, and Network Analysis," *BMJ* (364).
- Haveman, H.A. 1993. "Follow the Leader: Mimetic Isomorphism and Entry into New Markets," *Administrative science quarterly*, pp. 593-627.
- Hillman, V. 2023. "Bringing in the Technological, Ethical, Educational and Social-Structural for a New Education Data Governance," *Learning, Media and Technology* (48:1), pp. 122-137.
- iKeepSafe.org. 2024. "The Certification Process; Understand the Process and What Vendors Get from the Subscription." Retrieved June 3rd, 2024, from <https://ikeepsafe.org/our-process/>
- InternetSafetyLabs.org. 2023. "2022 Us K12 Edtech Benchmark."
- Jensen, M.C., and Meckling, W.H. 1919. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," in *Corporate Governance*. Gower, pp. 77-132.
- Jibb, L., Amoako, E., Heisey, M., Ren, L., and Grundy, Q. 2022. "Data Handling Practices and Commercial Features of Apps Related to Children: A Scoping Review of Content Analyses," *Archives of disease in childhood* (107:7), pp. 665-673.
- Joshi, M., and Hadi, T.H. 2015. "A Review of Network Traffic Analysis and Prediction Techniques," *arXiv preprint arXiv:1507.05722*.
- Keith, M.J., Frederickson, J.T., Reeves, K.S., and Babb, J. 2018. "Optimizing Privacy Policy Videos to Mitigate the Privacy Policy Paradox,").
- Lane, J.E., and Kivisto, J.A. 2008. "Interests, Information, and Incentives in Higher Education: Principal-Agent Theory and Its Potential Applications to the Study of Higher Education Governance," *Higher education*, pp. 141-179.
- Laufer, R.S., and Wolfe, M. 1977. "Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory," *Journal of Social Issues* (33:3), pp. 22-42.
- Liu, H., Patras, P., and Leith, D.J. 2021. "Android Mobile Os Snooping by Samsung, Xiaomi, Huawei and Realme Handsets," *techreport*, (Oct).
- Maras, M.-H., and Wandt, A.S. 2019. "Enabling Mass Surveillance: Data Aggregation in the Age of Big Data and the Internet of Things," *Journal of Cyber Policy* (4:2), pp. 160-177.
- Marini, G. 2021. "Coercive and Mimetic Isomorphism as Outcomes of Authority Reconfigurations in French and Spanish Academic Career Systems," *Policy Reviews in Higher Education* (5:1), pp. 89-108.
- Martin, K.D., Borah, A., and Palmatier, R.W. 2017. "Data Privacy: Effects on Customer and Firm Performance," *Journal of marketing* (81:1), pp. 36-58.
- Martínez-Ferrero, J., and García-Sánchez, I.-M. 2017. "Coercive, Normative and Mimetic Isomorphism as Determinants of the Voluntary Assurance of Sustainability Reports," *International Business Review* (26:1), pp. 102-118.
- Masters, T.M., Keith, M., Hess, R., and Jenkins, J.L. 2022. "Do Privacy Assurances Work? A Study of Truthfulness in Healthcare History Data Collection," *Plos one* (17:11), p. e0276442.

- Mayer, R.C., Davis, J.H., and Schoorman, F.D. 1995. "An Integrative Model of Organizational Trust," *Academy of management review* (20:3), pp. 709-734.
- McDonald, N., Badillo-Urquiola, K., Ames, M.G., Dell, N., Keneski, E., Sleeper, M., and Wisniewski, P.J. 2020. "Privacy and Power: Acknowledging the Importance of Privacy Research and Design for Vulnerable Populations," *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1-8.
- McKnight, D.H., Choudhury, V., and Kacmar, C. 2002. "Developing and Validating Trust Measures for E-Commerce: An Integrative Typology," *Information Systems Research* (13:3), Sep, pp. 334-359.
- Mizruchi, M.S., and Fein, L.C. 1999. "The Social Construction of Organizational Knowledge: A Study of the Uses of Coercive, Mimetic, and Normative Isomorphism," *Administrative science quarterly* (44:4), pp. 653-683.
- Nissenbaum, H. 2004. "Privacy as Contextual Integrity," *Wash. L. Rev.* (79), p. 119.
- NPICICA. 2008. "Nevada Privacy of Information Collected on the Internet from Consumers Act (Npicica), Nevada Revised Statutes Chapter 603a." from <https://www.leg.state.nv.us/nrs/nrs-603a.html>
- Othman, S., Darus, F., and Arshad, R. 2011. "The Influence of Coercive Isomorphism on Corporate Social Responsibility Reporting and Reputation," *Social Responsibility Journal* (7:1), pp. 119-135.
- Paci, F., Pizzoli, J., and Zannone, N. 2023. "A Comprehensive Study on Third-Party User Tracking in Mobile Applications," *Proceedings of the 18th International Conference on Availability, Reliability and Security*, pp. 1-8.
- Pimienta, J., Brandt, J., Bethe, T., Holz, R., Continella, A., Jibb, L., and Grundy, Q. 2023. "Mobile Apps and Children's Privacy: A Traffic Analysis of Data Sharing Practices among Children's Mobile Ios Apps," *Archives of disease in childhood* (108:11), pp. 943-945.
- Ross, S.A. 1973. "The Economic Theory of Agency: The Principal's Problem," *The American economic review* (63:2), pp. 134-139.
- Roszkowska-Menkes, M., and Aluchna, M. 2017. "Institutional Isomorphism and Corporate Social Responsibility: Towards a Conceptual Model," *Journal of Positive Management* (8:2), pp. 3-16.
- SDPC. 2022. "Standard Student Data Privacy Agreement: Ndpa Standard Version 1.0/ with Exhibit E."
- Solove, D.J. 2005. "A Taxonomy of Privacy," *U. Pa. L. Rev.* (154), p. 477.
- StudentPrivacyPledge.org. 2020. "K-12 School Service Provider Pledge to Safeguard Student Privacy 2020." Retrieved June 3rd, 2024, from <https://studentprivacypledge.org/privacy-pledge-2-0/>
- Taylor, V.F., Spolaor, R., Conti, M., and Martinovic, I. 2017. "Robust Smartphone App Identification Via Encrypted Network Traffic Analysis," *IEEE Transactions on Information Forensics and Security* (13:1), pp. 63-78.
- Tingling, P., and Parent, M. 2002. "Mimetic Isomorphism and Technologyevaluation: Does Imitation Transcendjudgment?," *Journal of the Association for Information Systems* (3:1), p. 5.
- UCPA. 2023. "Utah Consumer Privacy Act (Ucpa), Utah Code §§ 13-61-101 Et Seq.", from <https://dcp.utah.gov/ucpa/>
- Uvik. 2024. "Understanding Software Outsourcing in 2024: A Comprehensive Guide for Businesses."
- Zimmerle, J.C. 2021. "Safe, Sound, and Private: Promoting Data Protection for Students," *Computers in the Schools* (38:1), pp. 1-18.